Annika Wennersten

Back to the Woods or Into Ourselves?

Kant, Rousseau and the Search for the Essence of Human Nature
Abstract


This thesis contributes to a field of Kant’s practical philosophy that has received renewed attention, namely his moral anthropology. While it is true that Kant, in some of his best-known writings, literally says that the fundamental ground of morality must be pure and thus entirely free from admixture with anthropological principles, he nevertheless admits that these “subjective conditions” in human nature that “either hinder or help people in fulfilling the laws of the metaphysics of morals” make up the foundation of all applied ethics. In other words, in order to know if and to which extent human beings are susceptible to moral commands, we need to know our abilities as well as our limitations.

Kant wrote several works about these topics and his long-term teaching of anthropology shows that he had a continuing interest in the theory of man. Moreover, it is widely acknowledged that Kant, during the mid-1760s was highly influenced by Rousseau. It is hardly a coincidence that Kant’s first reference to the “unchanging nature of human beings” appeared at the same time as Rousseau proclaimed the need of finding the true nature of man – the unmasked being who has not been damaged by social prejudice. In order to understand man and his moral capacities we need to find his true essence or what really constitutes humanity.

Accordingly, a careful examination of the multifaceted characteristics of human nature is needed in order to understand the very concept of a moral being and to account for his moral progress. I will argue that Kant’s early insights about this need runs like a thread through his entire course of philosophy and that Rousseauian ideas actually affect also his critical ethics. They agree that man is sociable, but also suspicious. He has good predispositions but is likewise susceptible to corruption. My analysis will shed light on man’s eternal balance between conflicting forces and on the means needed for the progress towards the vocation of humankind. This reveals the need of knowing oneself and explains why the question: “what is the human being?” ought to be taken seriously.

Keywords: Kant, Rousseau, Moral Anthropology, Human Nature, Radical Evil, Sociability, Gesinnung, Character, Denkungsart, Propensity, Compassion, Self-Love, Self-Conceit, Ambition, Humility, History

Annika Wennersten, Department of Philosophy, Logic and Metaphysics, Box 627, Uppsala University, SE-75126 Uppsala, Sweden.

© Annika Wennersten 2015

urn:nbn:se:uu:diva-262830 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-262830)
7.4. Becoming Humans by Forming Nature ........................................... 232
  7.4.1. Humane Human Beings ........................................................... 233
7.5. Conclusions: Sensibility and Ethics ........................................... 235

Ending ......................................................................................................... 238
  Human Beings – Human Nature ............................................................ 239
    What Have We Learned from Human History? ................................ 241
    Human Morality: Aims and Abilities ................................................ 244
    Rousseau and Kant ............................................................................. 245
    Throwing the Mask ............................................................................ 247

Bibliography ............................................................................................... 249
  Kant’s Works .......................................................................................... 249
  Rousseau’s Works .................................................................................. 251
  Other Primary Sources .......................................................................... 252
  Secondary Sources ................................................................................ 253
Acknowledgements

Writing a thesis is an extensive project and the more one learns about the topic of one’s research, the more one becomes convinced that one will never be fully satisfied with the result. While the present text could surely have been better in many respects, there comes a time when one has to bring an otherwise never-ending project to an end.

Nevertheless, joining the department of philosophy has definitely changed my views on several matters. Although these years have fortified my confidence and taught me that I can do much more than I had ever believed, I would not have gone through this on my own. Two people to which I owe the greatest gratitude are Lilli Alanen and Jan Österberg who were the ones who encouraged me to do this and made me believe that even an ordinary person like me might have something interesting to say. Jan, who sadly will not have the chance to read my appreciations, was without doubt the main cause of my interest in the history of practical philosophy. He supervised some of my earliest texts and his inspiration opened my mind to a whole new world. The combination of his wide wisdom, modesty, wit and gentleness was of an incredibly rare kind. My conversations with Jan about Kant as well as various matters will always remain among my dearest memories. Lilli has always been very supportive especially by cheering me up at states of dejection.

During the course of writing and studying, I am grateful to the help I have received from my supervisors Pauliina Remes and Marcel Qaurfood, who continuously have taken their time to read my often very sketchy drafts. Especially Pauliina made an incredible work by reading, fixing practical matters and, most of all, pepping me during the last intensive phase. Also Tomas Ekenberg deserves to be mentioned since he literally seems to be interested in everything and he has always something clever to say. He is an excellent commentator and his generous and constructive feedback and encouragement have definitely lessened the burden.

Furthermore I do not dare to think about what this work would have looked like without the many charitable and intelligent comments that I have received from my colleges at the higher seminar in the history of philosophy. Great thanks therefore to Erik Åkerlund, Lorenzo Cassini, Tor Freyr, Peter Myrdal, Olof Pettersson, Halvard Stette, Alexander Stöpfgeshoff and Frans Svensson. Moreover, Professor Klas Roth from the Department of Pedagogy at Stockholm University commented on an earlier draft of the text. George
Masterton proofread the final manuscript and saved the printed edition from several embarrassing misprints. Still, of course, no one but me is responsible for possible remaining errors.

For several semesters I had also the privilege to take part at the lectures held by Professor Howard Sobel. His original reading of Kant resulted in engaging conversations and correspondence, and although we disagreed about several matters regarding Kant’s practical philosophy, I value these discussions high; he was yet another brilliant mind who is now missing.

A work like this effects one’s entire life, and there have been ups and downs during these years. Luckily, I have been surrounded by wonderful colleagues who have been very kind and understanding, and without the non-academic talk and activities with them, this period had not been as agreeable as it now were. Thus I want to thank especially Karin Enflo but also Karl Axelsson, Camilla Flodin, Kamilla Freyr, Kent Hurtig, Victor Moberger, Rysiek Sliwinski and all those people already mentioned as well as many others for making the department a nice place.

For help with various practical matters I am indebted foremost to Rysiek Sliwinski and Paulina Hansen but also to Anna Gustafsson, Lars-Göran Johansson, Sandra Olsson and Folke Tersman.

Moreover I would like to thank Helge Ax:son-Johnsson foundation as well as the Kungliga Humanistiska Vetenskapssamfundet for financial support. The grant from the latter made it possible to participate at the XI International Kant Congress in Pisa 2010.

Finally, I want to thank my family and friends for never-ending support, love, encouragement and patience with my various moods, including irritability, gloominess and constant stress and worry. Especially my parents and my sister have always been there for me. A special acknowledge should be given to my uncle Lennart Wennersten (1928-2015) who painted the image that is featured at the cover. I am pleased that I could use it here and I would like to thank Gusten and Helena for the photo.

Annika Wennersten,
Uppsala, October 2015
Abbreviations

Kant’s Works

Anthr. Anthropologie in Pragmatischer Hinsicht
BB Bemerkungen zu den Beobachtungen über das Gefühl des Schönen und Erhabenen
Beob. Beobachtungen über das Gefühl des Schönen und Erhabenen
Beweis Der einzig mögliche Beweisgrund zu einer Demonstration des Daseins Gottes
EF Zum ewigen Frieden
GMS Grundlegung zur Metaphysik der Sitten
IAG Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht
KpV Kritik der praktischen Vernunft
KrV Kritik der reinen Vernunft
KU Kritik der Urteilskraft
MAM Mutmaßlicher Anfang der Menschengeschichte
MAN Metaphysische Anfangsgründe der Naturwissenschaft
MS Metaphysik der Sitten
Nachricht Nachricht von der Einrichtung seiner Vorlesungen in dem Winterhalbenjahre 1765-1766
NET Von einem neuerdings erhobenen vornehmen Ton in der Philosophie
OP Opus Postumum
PS Untersuchung über die Deutlichkeit der Grundsätze der natürlichen Theologie und der Moral (Preisschrift)
Refl. Reflexionen
Rel. Die Religion innerhalb der Grenzen der bloßen Vernunft
TP Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis
TG Träume einer Geistersehers
VA [...] Vorlesungen über Anthropologie [student]
VM [...] Vorlesungen über Moralphilosophie [student]
VMet [...] Vorlesungen über Metaphysik [student]
VR [...] Vorlesungen über Rationaltheologie [student]
Vorarb.[...] Vorarbeiten [work]
Vorschritte Preisschrift über die Fortschritte der Metaphysik
Rousseau’s Works

1D Discourse on the Arts and Sciences
2D Discourse on the Origin of Inequality
3D Discourse of Political Economy
Conf. Confessions
E Emile or on Education
FM Favre Manuscript (Emile, first Draft)
GM Geneva Manuscript
LB Letters to Beaumont
Ld’A Letter to d’Alembert
LM Letters Written from the Mountain
ML Moral Letters
PF Political Fragments
PN Preface to Narcisse
RJJJ Rousseau Judge of Jean-Jacques: Dialogues
RSW Reveries of the Solitary Walker
SC On the Social Contract

References to Kant’s writings denote pages in the Akademieausgabe (29 vol.) with exception for the Kritik der reinen Vernunft for which I use the standard A/B notation. References are given by abbreviation plus volume and page.

With the exception for Emile, references to Rousseau’s works are to the Collected Writings of Rousseau (CW, 13 volumes) First numbers refer to pages in this edition. For Emile I give references to Allan Bloom’s translation since this is still the most widespread. The second number (after /) refer to pages in the French Oeuvres Complètes (OC) Editions Gallimard/Pléiade (5 volumes 1961-), Roman numbers denote “part” or “section” in respective work.

The List above features the most frequently cited works. For a complete list and detailed information about translations and editions, see the bibliography.
Arthur Melzer begins his book *The Natural Goodness of Man* by drawing attention to the apparent lack of “Rousseauians” in philosophy. While there are uncountable Kantians, Platonists, Aristotelians and at least some Thomists and Hegelians, Marxists and so on, nobody puts the label “Rousseauian” to themselves. In a sense, this is understandable; except for some of his contributions to political philosophy, Jean-Jacques Rousseau’s philosophical achievements are not always acknowledged. To the general public Rousseau is known for his personal life that, unfortunately, often colours how one views his overall thinking. Regrettably, a great deal of the commonplace criticism raised against him is based on *ad hominem* reasoning. For many, he is known as a paranoid romantic dreamer and the man who wrote his masterpiece about the proper education of children while his own were left to nursing homes. It is somewhat ironic that this passionate novelist, artist and composer of plays and music could become such a sharp opponent of enlightenment, fine arts and culture. Everyone has heard about his deep mistrust of social inventions as well as of his admiration of the so called “noble savage.” Due to this, his philosophy is commonly simply reduced to an outdated theory of human nature which has made him an object of intellectual dislike and even mockery. Besides this, his views on women, their intellect and proper education are hardly approved by current norms. It is about time to change this view; Rousseau’s influence on anthropology, moral psychology and ethics should not be ignored or belittled. And perhaps, after all, there was at least one true believing Rousseauian, namely Immanuel Kant.

There are several passages in which Kant declared his fondness of the “citizen of Geneva” who opened his mind and changed his views on several matters. His portrait hung above Kant’s working table and was the sole picture in the house. Rousseau saved God and providence from pessimistic criticism and established the optimism of Leibniz and Pope; because of this, Kant considered Rousseau’s contributions to the theory of human nature and

---

1 Melzer p. iv.
2 In his *Letter to mdm de Francueil* 1751, Rousseau explains that it would be better for his children to be orphans than to have a rogue for a father. (*CW V* p. 551)
3 That Rousseau was well aware of this “paradox” is shown in the *Preface to Narcissus*.
4 That the idea that Rousseau held science and modern culture as bad in itself will be shown to be a misconception during the course of this text. It should also be noted here that Rousseau himself never used the expression “bon sauvage”.

11
morals to be as important as Newton’s contributions to physical science. He was a modest man, or as Kant called him, a “subtle Diogenes” who was content with what nature alone could give him. Rousseau set Kant right and made him honor the ordinary man who he had formerly despised on self-conceited grounds, and this insight became the beginning of his theory of the unconditioned value of humanity. There was just one problem that Kant had to handle, namely that the beauty in Rousseau’s language distracted him so much that he had to read him carefully over and over again; only then could he analyze the works with reason. In short, Kant was one of Rousseau’s greatest admirers.\(^5\)

With these facts in mind, one may be surprised by Kant’s comparatively rare explicit references to Rousseau within his published works. Kant’s familiarity with Rousseau is indeed often expressed in pre-critical reflections and marginal notes, as well as in his shorter – and to the general public – less known texts on history and anthropology. Yet, it is notable that Rousseau’s name is almost absent from Kant’s works on moral philosophy. In fact, Rousseau is neither mentioned in the Groundwork, nor in the second Critique, and his name figures altogether only three times in the “Doctrine of Virtue”, the Religion and the Critique of Judgement and these times only parenthetically. (MS 6:445; Rel.6:20; KU 5:204)

This fact has not deterred commentators from trying to highlight the supposed connection between their ethical theories. Most works on Kant involve some comments, or sometimes a larger section, about the influence of Rousseau; an influence that is said to be “universally known” and “recognized by everyone”.\(^6\) The list of contributions to this area could be made much longer for the scholarly literature on the relationship of Kant and Rousseau is, as Clemens Schwaiger puts it, “endless and hardly manageable”.\(^7\) Among the foremost studies of this relationship one must mention Ernst Cassirer’s classical essays which are cited by everyone touching on this area in recent times.\(^8\) Before him Harald Hoffding (1892), Konrad Dieterich (1878), Klaus Reich (1936), Matthias Menn (1894), Arthur Schilpp (1938) and others had contributed to the topic. Besides this, there are uncountably many contemporary inquiries dedicated to Kant’s so called “Rousseauian turn” during the 1760s which is believed to be the origin of Kant’s ideas in some of his early works; especially the Beobachtungen über das Gefühl der Schönen und Erhabenen (1764), a text that appeared only shortly after the publications of Rousseau’s momentous works Emile and the Contrait Sociale.\(^9\) At that time Kant held that the difference between good and evil could be felt by the hu-

\(^{5}\) E.g., BB 20:30, 43, 44, 58; VM [Collins] 27:249.
\(^{6}\) See e.g., Gurwitsch p. 139.
\(^{7}\) Schwaiger p. 74.n.
\(^{8}\) Cassirer, Das Problem Jean-Jacques Rousseau (1932) and the first essay in the book Rousseau, Kant, Goethe from 1945.
\(^{9}\) Rousseau’s two works are said to have appeared in German translations in 1762.
man heart by means of a sentiment without further need for proof. Even if it is generally accepted that Kant, at the time of his Copernican and critical revolution replaced this early idea with a rationalistic foundation, it is commonly argued that Rousseauian roots are still present in his late ethics. More far-reaching is Richard Velkley’s view, which goes as far as claiming that Kant’s critical philosophy does not only have Rousseauian roots but rather a Rousseauian core, and that Rousseau was the single most important source of Kant’s final theory of the end of reason. The Rousseauian turn is thus not only prior to the later transcendental turn but also the condition of it. In fact, it has been argued that the reading of Rousseau’s works gave Kant all the material he needed for his critical ethics and the Groundwork could thus have been written already in the mid-1760s. But, one might ask, if this is true, why does Kant never openly acknowledge this influence in his critical, or sometimes called mature, ethics? Why does he not cite Rousseau’s lyrical expressions about the endless goodness of compassion and conscience? Why does he repeatedly describe the empirical conditions of man – which Rousseau so explicitly honored – as rude, evil and diametrically opposed to human dignity? The effects of Rousseau’s thinking, so commonly attributed to Kant, must thus be carefully concealed.

However, the wish to draw attention to the philosophical link between Kant and Rousseau is common and certainly understandable. Arthur Schilpp argues that by paying attention to Kant’s entire history of philosophizing one will see that the common characterization of Kant as a “cold Verstandemensch” would turn out to be an “unhistorical caricature”. Accordingly, a reduction of the radically rationalistic view, which – at least according to several commentators – obscures Kant’s multifaceted idea of sensibility, would show that he was more humane than commonly thought. Furthermore, a vindication of the role of reason, and a revelation of some possible “proto-Kantian” features in Rousseau’s system, could perhaps release him from the overly romantic reading that is commonly associated with his philosophy and let us see the real scope of his, indeed, quite sophisticated ideas. Kant obviously saw this sophistication and because of this, no less than John Rawls declared that Kant was “the best interpreter of Rousseau”.

11 Velkley (1989) p. 50 and p. 187 n.21.Velkley is here referring to Schmucker and Henrich (see Henrich p. 406). This claim gains some support from a letter from Kant to Lambert 1765 in which he tells of his plan to write a metaphysical foundation of natural philosophy and a metaphysical foundation of practical philosophy. It is noteworthy that two inquiries on these topics actually appeared in the mid-1780s, namely the Metaphysical Foundation of Natural Science (1786) and the Groundwork of the Metaphysics of Morals (1785). Still, of course, we cannot know if these texts could have appeared already twenty years earlier. Schwaiger calls the hypothesis of Schmucker, Henrich and Velkley an exaggeration. (Schwaiger p. 70)
12 Schilpp p. 168.
13 Rawls p. 200.
Nevertheless, my purpose in this thesis is not primarily to present just another work saying that Kant owed much to Rousseau for there is no doubt that he did; interpretative disagreements concerning this matter concern merely the scope or areas affected by Rousseauian thoughts. While these questions are both interesting and important for Kantian scholarship (and I will continuously refer to them) I will enter the discussion from a somewhat different angle. Rather than actively arguing for the existence of Rousseauian imprints (no matter whether they are to be considered merely roots or core) of Kant’s mature philosophy, I will show that also the bare idea of Kant as a devoted Rousseauian is a useful method for entering an interesting area which, if taken into account, uncovers several perspectives on Kant’s practical philosophy. The area in question is anthropology, that is, the study of human nature. Once one has entered this field one can no longer turn a blind eye on the “impure” sciences since these, although unfit for Kant’s purely metaphysical project, are inevitable for the conception of the factual human being. A retrospective analysis of Kant’s thinking about these areas – a thinking that doubtlessly was triggered by his early acquaintance with Rousseau’s writings – will uncover the complexity of his views about the human being. This, in turn, will give us new insights into Kant’s thoughts on applied ethics. The two philosophers’ thinking, however different they initially might appear, can be seen as highly intertwined and several of the problems that we will meet in Kant can be explained in Rousseauian terms and by paying attention to Rousseau’s way of dealing with similar queries.

My aim in the present text is, therefore, to highlight the significance of disciplines like anthropology, human history and moral education in Kant’s practical philosophy and, in so doing, shed light on the multifaceted topics where Kant and Rousseau intersect. To study Kant in the light of Rousseau is, if not a necessary, then at least an illuminating way of finding the possible sources of Kant’s applied practical philosophy. In fact, Kant’s metaphysical theory of morals is not only better understood in the light of anthropology but it is also only through this perspective that it is possible to comprehend what a pure ethics could be since it is said to be an abstraction from everything that belongs to man as a merely natural, or living, being. Without this extended understanding of Kant’s practical philosophy the common criticism about “empty formalism” presented by Hegel and others might very well be successful. In order to reach the complete picture of the conception of the human being and in that way to find the essence of this unique creature we need, as Kant himself put it, to realize that if there is any science that the human being truly needs, it is that which teaches him what one must be in order to be a human being. (BB 20:45)

---

14 As to the notion of impure with regard to Kant’s disciplines, see Louden (2000).
15 For Hegel’s criticism, see Elements of the Philosophy of Rights §135.
Human Nature as Philosophy

The central enquiry for moral anthropology is to find an answer to the question: what is the human being? The relevance of this question in Kant’s theory is established by the fact that he held it to complete his philosophical system by unifying the three questions he had earlier presented as summarizing the interest of human reason namely,

1. What can I know?
2. What should I do?
3. What may I hope?

These three questions put forward the main problems that are supposed to be answered by metaphysics, ethics and theology respectively. They can all, Kant says, be regarded as either purely speculative, purely practical or, as in the case of the third, both speculative and practical. (KrV A805/B833) In the Lectures on Logic Kant repeats these questions but now under the name of the field of philosophy in the cosmopolitan sense. It is here that Kant adds a fourth question; namely,

4. What is the human being?

This last question is answered by anthropology but, Kant says, the first three can, in fact, also be called anthropological for they all relate to the last. (LL [Jäsche] 9:25)

Anthropology, as an empirical science is clearly distinct from Kant’s pure metaphysics of morals. In the Groundwork, Kant stresses the importance of not letting anything belonging to our nature as sensible beings affect the moral laws themselves, but he nevertheless admits that these “subjective conditions” in human nature that “either hinder or help people in fulfilling the laws of the metaphysics of morals” serve as conditions of all applied ethics.16 In other words, in order to know if and to which extent we are susceptible to moral commands, we need to know our abilities as well as our limitations. In the transcript of Kant’s lectures on ethics we find the following description of the relation of anthropology and ethics:

Anthropology is concerned with subjective practical rules, it observes solely the actual behaviour of man; moral philosophy seeks to bring his good behaviour under rules, namely of what ought to occur. It contains rules for the right use of the will, just as logic contains the rules for the correct use of the understanding. The science of the rules of how men ought to behave is practical philosophy, and the science of the rules of his actual behaviour is anthropology; these two sciences are closely connected and morality cannot exist with-

out anthropology, for one must first know of the agent whether he is also in a position to accomplish what is required from him that he should do. \( (VM [Collins] 27:244) \)

The last sentence indicates the need to know human nature if we are to reach a genuine understanding of morality. Without this knowledge morality would be nothing but speculative, or an idea without practical significance. In spite of Kant’s general occupation with the building of metaphysical systems, there can be no doubt that he had a genuine interest in anthropology. He frequently taught on the topic and his lectures were generally appreciated by his students. In a letter to Marcus Herz in October 1773 in which Kant describes his plan for the lectures that winter semester, he announced that he now wanted to make anthropology a proper philosophical discipline. \( (Brief. 10:145) \) At that time, he had already had his first set of anthropology lectures and there would be many more. During his career as lecturer he taught anthropology every year from 1772 to 1796.\(^{17}\) Even far before that, in his announcement for his forthcoming lectures 1765-66, Kant added to his description of the course on ethics that in order to find the appropriate method for studying human nature, the doctrine of virtue should always begin by considering historically and philosophically what happens before specifying what ought to happen. \( (2:311-2) \) By the term “human being” Kant here referred to the “unchanging nature of human beings, and his distinctive position within the creation” and he thus dismissed all misunderstood ideas that proceed from the “distorted” conceptions that refer merely to “the mutable forms which is conferred upon him by the contingencies of his condition.” \( (2:311-2) \) Indeed, Kant believed that there must be some kind of true basis of our nature that unifies the human species. This idea is further supported by the today so oft-quoted lines from the Bemerkungen zu den Beobachtungen über das Gefühl des Schönen und Erhabenen where Kant associates Rousseau with the most prominent scientist of the time:

Newton saw for the first time order and regularity combined with great simplicity, where before him was found disorder and barely paired multiplicity; and since then comets run in geometrical courses. Rousseau discovered for the first time beneath the multiplicity of forms human beings have taken on their deeply buried nature and the hidden law by the observation of which providence is justified. Before that the objection of Alfonso and Manes still held. After Newton and Rousseau, God is justified, and Pope’s theorem is true. \( (BB 20:58-9) \)

\(^{17}\) A great deal of the thinking of this time was influenced this new scientific way of reasoning. Important works were Ernst Platner’s Anthropologie für Aerzte und Weltweise (1772), Johann Joachim Spalding’s Betrachtungen über die Bestimmung des Menschen (1748) and Johann Gottfried Herder’s Ideen zur Philosophie der Geschichte der Menschheit (1784).
This statement has often been taken to say that Kant regarded Rousseau as the “Newton of the mind” or of the moral world, namely the one who disclosed the order in a discipline that had formerly been thought to be lawless. Although all human beings are different, there is a common ground buried in the depth of their nature from which providence and divine justice could be derived. Although Kant should later reject Pope’s optimistic idea that “whatever is, is right” he was now convinced that anthropology deserved to be studied in a scientific manner.

Association through Anthropology?

In spite of these signs of Rousseau’s influence on Kant, a look in a typical book on the history of philosophy will most likely lead us to believe that they differed significantly in at least some areas, especially regarding their views on human nature. While Rousseau defends the natural goodness and sincerity of human beings, Kant declares with abundant clarity that man is essentially evil by nature. This is the main premise of the Religion where Kant makes a deep investigation into the moral nature of human beings and finds himself forced to conclude that that there is an evil first ground, necessarily present in every human being, that corrupts the very ground of our choice and distorts our way of thinking. Evil is, as Kant puts it, “woven into human nature”. (Rel. p. 30)

All this things being said, it is notable that in spite of this apparent difference in their general views on man and morals, these two philosophers are frequently discussed together with a view to bridging the supposed gulf between them. One prominent example is Allen Wood who says:

> Perhaps Kant’s doctrine that human nature has an inborn propensity to evil also shocked his contemporaries because it seemed diametrically opposed to Rousseau’s (modern, enlightened philanthropic) doctrine that human beings are naturally good. In that case, Kant’s critics were taken in by appearances. For if we look closely, we can see that these doctrines are not only compatible, but they are actually one and the same doctrine. (Wood (1999) p. 291)

---

18 Cf. Pilau who wrote that the reason why anthropology had generally been neglected was due to the belief that there was no need for it since, for the behavior of man, there were no laws. (Vd [Pilau] 25:733).

19 For this optimistic dictum, see Pope, Epistle 1 (last sentence) p. 53. The problems of optimism became actualized by the Lisbon earthquake 1755 and Voltaire did not spare his words when he, in his “Poème sur le désastre de Lisbonne” (and later in Candide), radically criticized the idea of divine providence and the current optimism of Leibniz and Pope. Rousseau’s answer is more to be seen as an expression of his personal faith in the good world order than a well-structured contribution to the problem of theodicy. (see Rousseau’s letter from Aug. 18 1756) As to Rousseau’s optimism and discussion with Voltaire, see Hendel Vol. 1 ch. X. For perspectives on Rousseau’s (as well as Kant’s) thoughts on the problem of evil in relation to the earthquake, see Susan Neiman (1997).
This is a strong claim, and although it is perhaps too strong, it is worthy of examination since it draws attention to the role of anthropology within Kant’s theory of evil in an important way. As we will see, Wood’s further discussion makes clear that he sees Kant’s entire idea of radical evil, and thus also the first and original act of maxim-making, as fundamentally anthropological. This puts him in direct opposition to the majority of Kant scholars who underline the strict metaphysical and transcendental ground of Kant’s theory of the origin of good and evil. What makes this area, and especially the discussion found in the Religion, difficult are Kant’s many distinctions between similar qualities of human nature, especially those that affect choice and action. According to Kant, evil is imbedded in us as a propensity [Hang], while he admits that we at the same time are completely good with respect to our innate predispositions [Anlage]. Still, Kant’s rigorist view excludes the possibility that a human being may be good in one part and evil in another and nor can he be totally neutral in this respect. Accordingly, since man must be either good or evil it is up to our free faculty of choice to navigate between conflicting forces and create the right kind of maxim. Because of the hotchpotch of concepts involved in this process, one must be careful and cautiously inspect every element.

In entering the giant field of discussions that are related to the topic called “moral anthropology”, I have selected themes that have caused much controversy and which thus are highly relevant for any study of Kant’s idea of the moral human being. A considerable part of the more advanced discussion (which will be found in Part II) will be fairly exegetic and I will there present close readings of Kant’s own arguments. At the same time, I will try to make sense of the recent discussion of these difficult matters and, in so doing, present some alternative solutions to the problems involved. Some of the chapters (especially 4, 6-8) may be read as independent contributions to recent discussions of Kant’s anthropology in general but, as the thesis will show, they are all links in a chain that will provide a fuller image of Kant’s applied practical philosophy; a philosophy in which the shadow of Rousseau will continuously be present. I will now present a more detailed overview of the parts and chapters of my thesis.

General Overview of the Thesis

The thesis is divided into two main parts. The first will introduce us to the notion of the human being and his faculties and give us the preliminaries that we need in order to handle the more difficult problems that will be brought up in Part II. In this first part, we will be introduced to the basic features of the human being in general as well as to Kant’s theory of freedom of the will and how motivation works. Recurrent issues will be the origin of good and evil, sociability, self-love, human imperfection and the role of cultivation. There follows a fuller overview of the different chapters.
I. Natural and Moral Beings

The central task for anthropology is, as we have seen, to answer the question what a human being really is. So, what is the human being? That Rousseau offered two main conceptions of man, that is, the man of nature (the savage) and the social man is generally recognized. This distinction, however, needs to be developed if we are to be able to apply it in more advanced contexts. Both these conceptions are portrayed with care and enthusiasm but while especially the Second Discourse may indicate the dual conception so often associated with his philosophy, Rousseau’s discussion in Emile and other works raises the question of whether the savage, strictly speaking, can be said to be a human being.

In dealing with Rousseau’s idea of the human being one must pay attention to the sharp distinction between nature and culture. This is illustrated by his famous declaration that when educating a young person, one must “choose between making a man and a citizen for one cannot make both at the same time”. (E I 39/248) As will become clear, however, in order to see the full picture of Rousseau’s moral anthropology we must add another type of man, namely “the natural man living in society”. This man is Rousseau’s ideal; a being that possesses what Lawrence Cooper’s calls civilized naturalness.

Kant, on the other hand, did not provide a comprehensive picture of the purely natural man. On the contrary, the savage [Wilde] is generally described in merely negative terms as void of everything that actually constitute the proper human being. He is simply an animal-like being; rough, undisciplined and determined by instinct alone. Human beings, Kant says, are distinct from animals by means of their ability to act from principles and this means that they are never directly determined by their physical drives. While clearly distinct from animals, human being is not holy but rather weak and impure. In order to advance towards our highest vocation [Bestimmung] man need to subject himself to several forms of education, namely cultivation [Kultivierung], civilization [Zivilisierung] and finally, moralization [Moralizierung].

In order to account for the forms that human beings can take, Chapter 1 will, therefore, be divided into four main sections that aim to explain the essential aspects of the natural, the social and the moral human being separately. Subsequently, the question of how these conceptions relate to the

---

20 However, it might have been the case that Kant, in 1769, actually wrote a short text called “Moralische Betrachtungen, der Mensch nach der Natur. Ein philosophisches Ideal von J. J. Rousseau” [Moral Thoughts on the Man of Nature – A philosophical ideal of J-J Rousseau]. This text is supposed to have been published in Königsbergenischen Gelehrten und Politischen Zeitungen and is referred to in a letter from Herder to Hamann that year. In this letter, Herder seems unsure about whether Kant actually is the author of the text in question. Nevertheless, there seems to be no general knowledge about this text and, according to Schwaiger, it is nowhere to be found. (Schwaiger pp. 90-1 including footnote no. 259).
ideal human being will be addressed. Through this, we will see that Kant’s and Rousseau’s respective approaches differ in several respects. While Kant at least intended to separate anthropology from morality, and thus treats them as two distinct (though related) disciplines, Rousseau did not and could not do so. As Gauthier points out, Rousseau’s so called “history of man” is primarily a “history of the fall of natural man”. It is a moral history or an explanation of wickedness.\textsuperscript{21} For Rousseau, morality originates in human nature or, more accurately, in man’s original nature. Rousseau’s anthropology is, therefore, essentially a historical explanation of why man has become what he now is and how he, as long as possible, can keep himself in accordance with this nature. Kant, while admittedly hinting at a similar conjectural history, was chiefly concerned about the progress from crudity to moral thinking and how we must proceed in order to restrict nature’s capacity to obstruct this development.

Chapter 2 will introduce the basic functions of the powers of mind which Kant, following Alexander Gottlieb Baumgarten and Christian Wolff, treats under the name empirical psychology [psychologia empirica]. In contrast to rational psychology, which is concerned with the pure will and our cognitive faculties, empirical psychology is the theory of our sensual qualities such as desires, instincts and similar non-rational foundations of action and motivation. For Kant, such explanations obviously fall outside the limits of pure morality and must be treated as material principles, that is, non-moral reasoning. This impurity, however, further strengthens the need for a correct understanding of the human being and its limits, for here we are directly confronted with the powers that are affected by the “subjective conditions” that either hinder or help people in their moral reasoning. Central Kantian concepts that we need to consider here are the will, the faculty of choice and maxims. Since Kant also believed that the human being is essentially a free being, we need to look at his theory of freedom and find out how Kant could persist in the claim that we are always free even though we, as natural beings, are constantly subject to laws of natural causality. In this chapter, I will also briefly explain how I understand Kant’s intentions with regard to pleasure in motivation by re-examining the so often discussed problem of psychological hedonism that seems to be the inescapable consequence of Kant’s Theorem II, a theorem telling us that all actions that are not performed from duty are expressions of self-love or the care for one’s own happiness. (KpV 5:22).

In Chapter 3 I will turn to a notion that is central to the understanding of the moral person, namely character. Because of its significance for Kant’s idea of the moral being, the term is one of the most popular in recent scholarship; especially since it is believed to provide for a more complete picture of Kant’s ethics than the action-centered view that appears to follow from

\textsuperscript{21} Cf. Gauthier p. 4.
the *Groundwork*. By discussing character, Kant seems to shift focus from the mere action to the agent or, more accurately, to the inner constitution of the agent.

Since Kant generally identifies our true moral character as a consistent and stable “way of thinking” [Denkungsart] it seems reasonable to take a closer look at this concept. I will show that an incomplete conception of Denkungsart (or character in general) will leave us in a position from which it is hard to make sense of some of Kant’s very strong claims about this notion. One difficulty that has caused disagreements in the recent debate is Kant’s claim that character (and especially Denkungsart) is extremely rare among human beings. If taken literally, the rarity-argument conflicts with much of what Kant says about maxim-making and human agency in general. Admittedly, there are a number of comprehensive discussions of Kant’s theory of character that admit that the concept of character is complex. Still, one finds that most discussions tend to overlook the extent of this complexity. As we will see, however, there are ways of tackling these problems.

Moreover, a look in Kant’s corpus will show that Denkungsart figures frequently and that it, together with its empirical counterpart (Sinnesart), explain the outlook of the human being. Taken together, these two concepts make up what has generally been identified as Kant’s idea of the “double character” of man, namely the distinction between *intelligible* and *empirical character*. This dual perspective of man is sometimes understood in terms of the two *worlds* to which the human being simultaneously belongs. Alternatively, they are taken to represent two different *standpoints* that we can take in relation to the world, to ourselves and to our actions. However, the polysemous nature of Denkungsart also affects this pair-relation and a closer examination of Kant’s texts shows that we have to deal with at least three Denkungsart-Sinnesart distinctions that operate in somewhat different areas. These will be spelled out, providing us with a fuller conception of the human being, and removing the dilemmas that were caused by the rarity-claim, since it will be shown to hold only for a very specific form. These conclusions will also be helpful when we move to chapter 4 where I will perform a thorough examination of the notion of Gesinnung in order to explain the motivational force that underlies all maxims and which thereby is the ultimate origin of the worth of any human being as well as of their deeds.

**II: Good and Evil**

After these preliminaries, which most of all aim at elucidation of the human being’s basic constitution (as well as the feelings and powers that he possesses), I will continue with a more specific analysis of some of the more significant implications of the complex nature of the human being. I will push the limits a bit further and take a careful look at the ultimate origin of all specific choices and thus of all maxims and actions. While this was briefly touched upon already in chapter 3, the need to widen the perspective and
study this origin (or ground) in relation to man’s moral nature in general will become apparent. It is this ultimate ground that determines whether the human being can be regarded as good or evil by nature. As mentioned above, this is an issue where Kant and Rousseau seem to be presenting two entirely different premises, though it has also been argued that the difference is merely terminological and that they actually defended one and the same doctrine. Rousseau declared that man is truly good by nature but was converted to evil by social corruption. Kant considered man as naturally evil because of an inner source of corruption that is present at birth in everyone. Still, both held that we are rational beings with a free will allowing us to always choose to do the right thing. Kant also added that we can also always do so for the right reasons. At the same time, they unanimously saw the human being’s misuse of these powers and they both tried, from their respective outlook to comprehend the reasons behind this tragedy. In chapter 4-6 I will enter this difficult area and try to make sense of Kant’s and Rousseau’s respective ways of explaining the origin of good and evil. This will require a deepening of several of the topics we touched upon in Part I.

Chapter 4 will deal with two distinct though related problems. First we need to consider Kant’s idea of the universality of the evil propensity and second, how this ultimate ground is to be understood as existing side by side with our genuinely good predispositions through which we are said to be free to act from the moral law alone. Kant’s apparently strong claims about evil have inspired readers to search for supporting arguments. Such a reader is destined for disappointment, for the most frustrating thing here is not Kant’s provoking claim that “there is no one righteous” but rather his omission of any real argument for it. On the contrary, he frankly states that we can “spare ourselves” of such a “formal proof” because experience gives us what we need. (Rel. 6:32f.) Plainly, since empirical observations cannot account for the universality of evil in human beings, Kant is skating on thin ice. Since an argument of this kind seems to need an a priori ground, readers have generally dismissed Kant’s appeal to experience as a bad excuse for his inability to come up with a reliable ground. Alternatively, they have tried to reconstruct the proof that they believe may be implicit in Kant’s previous discussion. Others, like Wood and Anderson-Gold have simply claimed that Kant’s conception of evil actually is empirical and that it therefore can be explained in anthropological terms. According to Wood, Kant’s conception of evil is essentially the same as he has elsewhere discussed under the name of “unsociable sociability” [ungesellige Geselligkeit]; one of the most basic but at the same time powerful social energies that the human being must handle. This tendency, which occurs as soon as the human becomes aware of himself as part of a social order, consists in the inner tension between conflicting feelings. While man feels the need for others he will, at the same time, also experience strong feelings of suspicion and agitation towards
them. As we see, this tendency is fundamental for the understanding of both the social, and the moral, being.

I will present and briefly comment on these attempts and conclude by saying that although I agree that Kant’s general argument seems to need an a priori foundation, I think that the discussion has focused too much on what I take to be a quite innocent comment on Kant’s part. The claim “that we can spare ourselves” of the formal proof is not a simple subterfuge caused by his inability to provide such a proof, but rather a rhetorical strategy that once again shows that Rousseau was still in his mind.

A second topic that is brought about by the appeal to an ultimate foundation of our actions (whether good or evil) is Kant’s notion of Gesinnung. Kant’s theory of evil is complex but one thing seems clear; namely, men are evil because they are corrupted in their very nature and this internal evil ground corrupts every choice that the agent will ever make. Apparently, Kant needs to establish a common ground in human nature from which all morality of the person could be derived. In order to better understand the importance of this move one may recall one of Kant’s best-known statements from the Groundwork:

An action from duty has its moral worth not in the purpose to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend upon the realization of the object of the action but merely upon the principle of volition in accordance with which the action is done without regard for any object of the faculty of desire. That the purposes that we may have for our actions, and their effects as ends and incentives of the will, can give actions no unconditional and moral worth is clear from what has gone before. In what, then, can this worth lie, if it is not to be in the will in relation to the hoped for effect of the action? It can lie nowhere else than in the principle of the will without regard for the ends that can be brought about by such an action. (GMS 4:399f.)

Although Kant does not give any detailed explanation of maxim in this work, we are at least told that it is the subjective principle of the action or of the will (or willing/wanting).22 This principle is explicitly distinguished from the purpose or ends that we have in mind [Absicht] as well as from the effects that the action may bring about. Two different agents can thus perform what appears to be the same action while their maxims may be entirely different. The maxim explains the nature of the action and, if we were able to examine it, we would thus find the motivation (and so the moral content) behind every visible deed. Unfortunately, Kant also declares that these inner principles are epistemically inaccessible to human beings. We cannot observe maxims and we cannot deduce them from a specific act. We can never know for sure if an action, including our own, was actually done from duty and not from

22 Ger. “Prinzip des Wollens”.

23
any hidden motivation based on inclination. (GMS 4:406 cf. Rel. 6:20). Because of the obvious importance of the underlying principles of the actions there are reasons for examining the ground of the very adoption of them. Obviously, since Kant holds that we, as rational beings, are responsible for our decisions, maxims do not just occur in our minds. Rather, all maxims are freely chosen and this means that they must somehow have an origin in our innermost nature. In other words, in order to reach the ultimate origin of morality we must seek the source of it somewhere beyond the maxims themselves.

Additionally, it appears that maxims must also have their ultimate ground; a ground that determines – or at least affects – which maxims we eventually will choose. This ground must somehow lie in our will. This ultimate ground is generally identified as our Gesinnung – a moral “disposition” or “attitude” that decides which incentives to act upon and hence determines whether we are good or evil human beings. Ethics, we are told, “is a philosophy of Gesinnung”. (VM [Collins] 27:299) Kant’s most detailed analysis of this ultimate ground is found in the Religion, but his highly confusing arguments have stimulated many commentators to attempt intelligible reconstructions of it. In spite of the obscure nature of Gesinnung, one interpretation has, however, become more or less standard. If one looks up the term Gesinnung in recent secondary literature on Kant’s practical philosophy, one will normally find it defined as the agent’s highest (or supreme) maxim [oberste Maxime] that serves as the ultimate ground of the adoption of all further maxims. This idea has been widely accepted especially after Henry Allison’s careful analysis in Kant’s Theory of Freedom (1990). Arguments for this reading are usually derived from some passages in the Religion in which Kant says that Gesinnung is “the first subjective ground of the maxim” (Rel. 6:25) and that the first ground of the adoption of maxims must itself be a maxim, for otherwise the ground of this adoption could be traced back to mere natural causality, and this would contradict freedom (Rel. 6:21).

While these lines seems to provide a solid ground for the standard interpretation one serious complication of this popular reading is that it is difficult to convincingly claim that there is an obvious and important distinction between Gesinnung and the propensity to evil. A closer look at the arguments cited in support of this reading reveals that Kant does, in fact, not explicitly say that Gesinnung is our highest maxim. In order to reach this conclusion one needs to combine different remarks from various pages of the Religion in which Kant talks about a first, or a highest, ground. However, while the highest ground is indeed sometimes identified as Gesinnung, this first, ultimate or highest ground is frequently also said to consist in our propensity to evil. This complication needs to be taken seriously.

The Gesinnung-Hang problem is a major topic in recent scholarship and although some commentators accept that the two notions tend to coincide, most readers argue that they are (and indeed must be) separated since Kant
would otherwise not be able to account for a “change of heart” [Herzenänderung]. For Kant, the “change of heart” is the fundamental revolution in the human being’s moral nature through which a new and better man is brought about. I will argue that given the definitions in the Religion, Gesinnung and Hang are indeed equated but I will also show that this does not affect the possibility of the change of heart.

Moreover, the technical definition of Gesinnung from the Religion is not the only one figuring in Kant’s writings. More often he seems to regard it as more or less equivalent with moral character and as the only thing that can give maxims and human beings moral worth. Since moral character is generally identified with Denkungsart, Gesinnung and Denkungsart seem to be the same. While this is commonly accepted in secondary literature, critics of this idea generally refer to Kant’s claim that character and especially Denkungsart are said to be incredibly rare while Gesinnung exists in everyone. If combined with the problem just mentioned, namely the supposed Gesinnung-Hang identity, we see that also the human character, via Gesinnung by definition corresponds to a moral propensity which, at least in the Religion, seems to be a propensity to evil. As we will see, even further notions may be added to this list. With regards to this apparent dilemma, I will argue that the problem here is that critics often overlook the fact that neither Gesinnung nor character, are uniform concepts in Kant’s writings. Once one has made the relevant divisions, several of the puzzles that have obscured the discussion of Religion will more or less disappear.

Since one of my aims with this investigation is the detection of possible Rousseauian influences on Kant’s moral anthropology (and thus also on his moral thinking in general) we need to study systematically the foundation of morality and our awareness of good and evil. Kant identified respect [Achtung] for the law as the sole incentive [Triebfeder] to morality and this “moral feeling” or immediate awareness of obligation is given to us by our intellect. If there is a corresponding principle in Rousseau’s moral psychology it is the divine voice conscience.

While a chapter devoted to Rousseau’s moral psychology might seem a digression from the general outline of this thesis, I will show that we have much to learn from his ideas about the origin of passions and how we may handle the negative effects that may arise from our sensuous nature. With a correct understanding of Rousseau’s idea of conscience and how it interacts with our nature as a whole, we will have a tool for receiving an understanding of the true origin of the various expressions of sentiments, passions, agitations and judgements that will inescapably arrive in the mind of a social being. Much of this will also be relevant when we will discuss the effects of social passion in chapter 6 and 7.
It is clear that, for Rousseau, conscience is man’s most reliable principle and, in contrast to reason, it is said to never deceive. As we will see, this feeling – the voice of the uncorrupted soul and of true nature – is of greatest importance for Rousseau’s theory of man’s natural goodness. Still one might wonder how the human being, if endowed with such a marvellous and infallible guide, could turn into the fallen being that he now is. If conscience, as Rousseau says, really make us love the good – why do so many of us willingly choose what is truly evil?

A natural answer is that also this guide, as with most of our natural sentiments, has been corrupted by our general moral fall. Still, an acceptance of the possibility of a corrupted conscience seems to contradict Rousseau’s claims about the infallibility and eternity of this divine principle. This apparent dilemma does not become less awkward when we recall that Rousseau acknowledges that social relations and comparison are the necessary conditions of the very existence of any genuine (non-trivial) idea of conscience. Since the idea of a corrupt conscience would obviously contradict Rousseau’s optimistic view, I think that there are reasons to analyse this problem.

Accordingly, in Chapter 5 I will show that those who argue that this inner guide is susceptible to corruption are mistaken and the reason why passions caused by the social form of self-love (amour propre) can blind us is simply that a social being no longer understands the original voice of nature. Now, while the very idea of a “pure voice of nature” that might not – at least not in its celestial form – exist in the state of nature may seem contradictory, I believe that Rousseau’s theory can resolve this apparent dilemma. I will also show how this feeling is related to the social process and to Rousseau’s idea of human freedom and moral responsibility.

In Chapter 6 I will continue the discussion of Wood’s hypothesis of the anthropological components in Kant’s theory of evil. If we inquire into human nature we will find that there is one sentiment which appears in more or less every aspect, namely self-love. The concept of self-love is central in all theories of feelings and motivational psychology since it involves a direct reference to oneself and one’s own interest. It is commonly believed that self-love makes up a radical contrast to moral thinking since the striving towards satisfaction of one’s own desires or interests indicates the presence of an egoist disposition, and we may thereby frequently ignore more sublime commands. Yet self-love is clearly one of the most natural and powerful of our sentiments. Moreover, it is important to be aware that what is commonly referred to as “self-love” is not a simple passion but comes in different shades.

Rousseau’s distinction between amour de soi and amour propre is familiar and widely discussed. While the former consists in a predominant feeling of self-respect and care for one’s preservation, the latter is relative, artificial,

23 E IV 285/595; RSW IV 1029.
developed in society and cause of all competition and mutual hostility between men. Even if it is tempting to focus on the difference between them one must not overstate the unnaturalness or dangers of the latter. The tendency to reduce *amour propre* to either *pride* (usually in the unjustified sense of arrogance or of finding oneself superior) or *vanity* is unfortunate, since it fixes the exclusively negative understanding of this central and – at least sometimes – constructive passion. Although several attempts have been made to alter this reductive interpretation, *amour propre* is still generally considered as the “bad” form of self-interest.24

So what does Kant say? After having connected happiness with the fulfilment of one’s own inclinations, Kant divides this general self-regard [Selbstsucht] into *self-love* [Eigenliebe] and *self-conceit* [Eigendünkel]. In their basic meanings self-love is said to consist in elementary benevolence towards oneself (*philautia*) while self-conceit is satisfaction with oneself which takes the form of an idea of superiority (*arrogantia*). (KpV 5:73) In other works, Kant provides a number of further definitions of self-love that complicate his picture of this principle. We will look closely at some of them, especially the distinction that he draws in the *Religion* between mechanical (animal) and relative (rational) self-love.

Now, as was mentioned above, Wood argued that Kant and Rousseau, although explained in seemingly opposed terms, actually offered identical theories about human nature. More far-reaching is Wood’s further examination of self-love and the *social passions* that we inevitably develop as soon as we live together with other human beings. These passions are caused by our “unsociable sociability”. Among these passions Kant counts several forms of covetous strivings, especially for honour, wealth and power. All these passions share the same basis, namely *self-love*, but Wood goes further and states that they are all, in fact, *identical* expressions and he adds that they are also equivalent to the Rousseauian term *amour propre*. When discussing Kant’s idea of nature and culture, Wood calls attention to Rousseau’s influence and states that:

> [amour propre] is, of course, the same trait as “self-conceit”, or “ambition” – the sense of comparative self-worth and the desire to achieve superior worth in the eyes of others – which constitutes, in Kant’s view, the radical propensity to evil in human nature. (Wood (1999) p. 291)

Elsewhere, Wood declares in the same explicit manner that Kant virtually understands *everything* in human life as expressions of “unsociable sociability”, self-conceit or the propensity to radical evil in the human nature. These qualities are simply said to be “three names for the same reality”.25 Social

---

24 For non-reductive interpretations of *amour propre*, see e.g. L.D. Cooper and F. Neuhauser.
passions, Wood says, are related in that they are all grounded in one single delusion, and hence, they are the root of all evil.\textsuperscript{26} Accordingly, if we accept Wood’s suggestion it seems as if we must conclude that “unsociable sociability”, self-conceit, ambition, the propensity to evil and \textit{amour propre} denote one single human characteristic. An effect of this is what Wood, followed by Jeannine Grenberg, describes in terms of the \textit{comparative-competitive tendency} of human nature.\textsuperscript{27}

Admittedly, it is not at all far-fetched to see similarities between Kant’s and Rousseau’s dual conception of self-concern and the linking of Kant’s division from the \textit{Religion} and Rousseau’s dualism is also, except for some minor details, more or less correct. However, far from everyone seems to be fully aware of the implications that follow if this analogy is also applied to the distinction from the second \textit{Critique}, namely between \textit{self-love} and \textit{self-conceit}. While it is common to draw parallels between this moral distinction both to Kant’s anthropological distinction from the \textit{Religion} and to Rousseau’s division between \textit{amour de soi} and \textit{amour propre}, I will argue that this can only be done at the cost of considerable damage to both Kant’s and Rousseau’s idea of man’s most social passions. To say that self-conceit is the same as our rational self-love, or Rousseauian \textit{amour propre}, is to reduce these fundamental and essential human features to their worst possible manifestations.

In order to show that this pessimistic conclusion is not the only one possible, I will here show that the “\textit{one and the same-reality claim}” does not hold for self-conceit, nor does ambition directly correspond to any of the notions involved in our study. Moreover, the analysis will also show why \textit{amour propre} cannot properly be said to be equivalent to self-conceit even though it may indeed be equated with unsociable sociability as well as with the comparative self-love that makes up our predisposition to humanity.

In the last chapter I will assess Kant’s attitude towards feelings and non-moral grounds of motivation. It is clear that a rationalistic theory of morality that, like Kant’s, appeals to objective laws needs a more secure foundation than mere sensibility. Furthermore, as we learned from Kant’s general idea of non-moral acting, all such principles could be summed up under the principle of self-love. If we recall that the “principle of self-love” is equated with the “principle of one’s own happiness” there is no wonder that Kant’s theory of feeling-based actions has caused much controversy. From these premises it seems to follow that all desires, however different, share the same basis in a fundamental desire for happiness. This, in turn, suggests that Kant held benevolence and sympathy as being on the same footing as greediness and desire for honor. Because of this, Kant’s view on compassion has generated much dissatisfaction and critics love to disapprove of this “intuitively im-

\textsuperscript{26} Wood (1999) p. 290.
\textsuperscript{27} Wood (1999) p. 135, Grenberg ch. 1 (passim).
plausible” view. On the other hand, commentators friendly to Kant spend much time and effort in trying to avoid this idea in order to make it more agreeable to modern moral canons. The traditional understanding of Kant’s moral theory is that it requires that we eliminate all influences from arbitrary and feeling-based principles and subject ourselves solely to the objectively grounded command of the moral law. In other words, this is supposed to be the only way by which it is possible to act non-egoistically (i.e. not from self-love). Those who want to fortify this strictly rationalistic view frequently refer to Kant’s familiar discussion of the philanthropist in the *Groundwork* and to the *Critique of Practical Reason*. Here we find the passages where he treats sympathy as a slavish, blind and irrational feeling and where the “cold-hearted benefactor” seems to be the archetype of a moral agent. (GMS 4:398 ff. Cf. also KPV 5:82) The idea that an action can have genuine moral worth only if the motivational ground of the action is entirely free from admixture of feelings or private affections was ridiculed by Schiller already at Kant’s lifetime and further criticized in our time.

In order to nuance this discussion, **chapter 7** will develop Kant’s theory of sensibility in ethics. Those who are only familiar with the *Groundwork’s* theory of what constitutes moral worth might be surprised to see that Kant in §35 of the *Metaphysics of Morals* literally introduces a duty to cultivate our natural compassionate feelings by making ourselves aware of the miseries of sickness and poverty. Compassion is now, in Kant’s own words, said to be “one of the impulses that nature has implanted in us to do what the representation of duty alone would not accomplish.” (MS 6:457) Since Kant had repeatedly stressed the possibility of always choosing duty before inclination, this is obviously a somewhat peculiar and even “un-Kantian” twist. Kant does not say anything more about this duty, and the attempts to make sense of it are numerous. However, to my knowledge, no one has yet studied this passage by making use of Rousseau, but I believe that we have good reasons for applying his ideas here. My main reason for this is that Rousseau’s three maxims from book IV of *Emile* give us what we need in order to show that, while compassion cultivation may be a vital step towards human perfection, we need to widen the perspective in order to see how this cultivation is possible in the first place. The application of Rousseau’s maxims to Kant’s outline of compassion discloses the underlying conditions for the possibility of compassionate and moral thinking. While Kant’s proposal in §35 at least initially may have appeared to be harmless, we now see that this cultivation requires that we compare our state with others; a process that we have seen to be hazardous since it, because of the comparative-competitive tendency, too often is the cause of unwanted passions, like inflamed amour propre, ambition or self-conceit. This, I will argue, is why the most significant of the conditions of the possibility of cultivating compassion is humility. Even if compassion does not play nearly the same role for Kant as it does for Rousseau, who frequently describes this feeling as one of
the most natural and important in our progress towards goodness, Kant does say things indicating that this feeling actually plays a role in our moral life. Furthermore, it is hardly a coincidence that the word human is closely connected to the term humane in several languages and this vocabulary is, as we will see, also applied by Kant himself.28

As this general overview shows, the area in Kant’s thinking which is labeled anthropology obviously covers a wide range of topics which all concern the understanding of human beings and their powers, needs, feelings, wishes and depravities. There is something called nature, but there are also culture and morals. Whether our challenge as human beings is, as Rousseau says, a matter of remaining in accordance with nature or, as Kant says, a matter of “taming the savagery” can be discussed but, as we will see, the difference may not be as great as initially appears. It is all a matter of finding the source, or essence, of true humanity and finding out whether this source is internal or external to the human being.

Kant himself ended the Groundwork by saying that when reflecting on the categorical imperative we will comprehend nothing but its incomprehensibility. (GMS 4:463) A similar dejection may, at least initially, be felt when one encounters Kant’s moral anthropology, for every careful reader of the Religion and Anthropology will become aware of the difficulty one faces in reaching a full understanding of his idea of our inner nature. Although the awareness of the mysterious nature of the moral law might be “all that can fairly be required of a philosophy that strives in its principles to the very limits of human reason”, we shall not give up on the possibility of understanding anthropology. It is, after all, the basis of all applied moral philosophy and a correct understanding of our natural capacities would help us to use these “subjective conditions” in a better way. Most of all, we must realize that the impurity of this discipline does not only involve hinders to morality but also the feelings that may facilitate our moral thinking. By highlighting the positive effects of our sensibility, and by focusing on the way our empirical nature and rationality interact, one can no longer escape the fact that the question “what is the human being?” both requires and deserves to be taken seriously.

Rousseau and Anthropology in Kant’s Philosophical Timeline

Many are those who have searched for the ultimate origin of what is generally thought of as Kant’s so called “mature” and “final” formulation of ethics.

28 Ger. Menschheit/Menschlichkeit; Fr. humanin, humanité; Lat. humanitas.
A good number of them also assign a significant role to the way Rousseau is portrayed in them but, more often, Rousseau is said to be a figure who undeniably inspired Kant during his early years but who was more or less abandoned at the time of Kant’s critical turn. In contrast to Rousseau who, at least according to himself, continuously had defended one and the same principle, Kant’s career involves a number of more or less radical turns, shifts, stages, or even “revolutions” of which some are better known than others. We have all heard how he, because of Hume’s sharp-minded insights, awoke from his dogmatic slumber and started to think critically about the way by which we interpret the world. Yet while the discussion about Kant’s new insights is immense, what he actually turned away from is not always spelled out. Much happened during the fifty years or so that Kant philosophized and lectured on various subjects. Uncountable works have been written on Kant’s intellectual development over this period, most of which almost exclusively focus on the formation of his ethics and the influences that might have led to its final form.

Nonetheless, regardless of what may be called Kant’s early, late or mature philosophy, there are themes that can be seen as permanent threads running through his entire production. That Rousseau and “Rousseauian topics” are such threads becomes clear if one takes a look at Kant’s timeline.

Kant’s philosophical chronology is generally divided into a pre-critical and a critical period, where the Critique of Pure Reason serves as the milestone in his philosophical development. However, although it is fairly clear that Kant’s critical thinking adds an entirely new perspective to his philosophy, one cannot simply draw a line at 1781 and say that everything that was published before is “pre-critical” and, therefore, the philosophy of an immature intellect. Even if the Critique of Pure Reason was said to be written in less than five months, Kant had reflected upon these ideas for twelve years. Consequently, most of the materials from the 1770s contain “critical” matters and Kant had already at that time made the shift from his pre-critical era. Furthermore, Kant certainly did not totally abandon the arguments from the Groundwork, even though the works from the late 1790s, particularly the Metaphysics of Morals (1798), seem to make room for at least some feelings also within the moral character.

29 In his Letter to Beaumont Rousseau declared: “I have written on various subjects, but always with the same principle: always the same morality, the same belief, the same maxims, and if you will the same opinion”. (LB 22/928)

30 For discussions of Kant’s early philosophical development and hypothetical periods in this progress, see e.g. the works of Hoffding, Foerster, Menzer, Schilpp, Schmucker, Schweiger Henrich and Beck (especially Beck 1969). For some more recent discussion of Kant’s life and early philosophy, see e.g. Zammito (2002), Kuehn (2001) and Rischmüller’s comments to Kant’s Bemerkungen.

31 See Kant’s letter to C. Garve in 7 Aug. 1783 and the letter to M. Mendelsohn 16 Aug. 1783. Hoffding makes this point in his Kontinuiteten i Kants filosofiske Udviklingsgang (1893) p. 6.
Harald Høffding writes both in his text “Kontinuiteten i Kants filosofiske Udviklingsgang” (1893) and in his Den Nyere Filosofiens Historie (1904) that Kant’s development regarding moral philosophy should rather be divided into three periods where the first covers the years 1762-66, the second 1770-80 and the third 1785 and forward. In his paper “Rousseaus Einfluss auf die definitive Form der Kantischen Ethik” (1898) it becomes clear that it is the third era, namely the one from Groundwork and onwards, that he believes to express the definite form of Kant’s ethics.

Although the division into periods might look arbitrary, it may still be advantageous to make use of these hypothetical stages when attempting to analyze the supposed influence of Rousseauian thoughts over the course of Kant’s thinking, especially since his explicit glorification of Rousseau is limited to the so called pre-critical period. When Kant refers to Rousseau in his historical and anthropological works from the critical era, he does praise his forerunner for his insightfulness. Still, one must not overlook the fact that he also frequently uses Rousseauian thoughts in order to point out the problems of this way of thinking. We will see several examples of this in the forthcoming analysis.

With regard to ethics and the possible influence of Rousseau, there are a few things that ought to be taken into account. Although the first occurrence of Rousseau’s name within Kant’s corpus appears in the letter from Hamann 1759, Kant almost certainly read the Discourses when they appeared at the beginning of the 1750s. Rousseau was well known in Germany at this time and all his most influential works were almost immediately translated into German. Lessing reviewed the first Discourse 1751. Kant’s first work that directly discussed moral philosophy was a section in an early prize essay called Inquiry concerning the distinctness of the principles of natural theology and morality which should soon be followed by the Observations. It was also at that time that Kant taught moral philosophy for the first time at the university at Königsberg; more precisely in the winter semester 1763-64. One of the students who attended these first lectures was Herder whose notes are preserved. In these notes, the traces of Kant’s fondness of the moral sense school are still visible, and here we may also find several references to Rousseau. The same holds for other early lectures, for example, the notes

---

32 This scheme essentially mirrors the ones found in the studies that Friedrich W. Foerster and Paul Menzer under similar titles presented almost at the same time. See Foerster Der Entwicklungsgang der Kantischen Ethik bis zur Kritik der reinen Vernunft (1894) and Menzer “Der Entwicklungsgang der Kantischen Ethik in den Jahren 1760-1785” (1898).

33 For comments on Rousseau’s influence in Germany at the early 1750s, see “Jaumann: Rousseau in Deutschland: Forschungsgeschichte und Perspektiven” in Rousseau in Deutschland: Neue Beiträge zur Erforschung seiner Rezeption. Cf. also the contributions of Kronauer and Velkley in the same volume and Guthke.

34 Before that, Kant had touched upon some moral issues, for example the idea of freedom in his doctoral thesis Nova Delucidatio (1755) as well as in the Beweisgrund (1762).
taken by Powalski and Collins, but in later fragments, as for instance, those of Vigilantius (1793) they are rare.  

The *Prize Essay* was probably written at the end of 1762 for the contest of year 1763. It was later published 1764, in the same year as the *Observations*. Commentators disagree whether Kant, at the time he wrote the *Prize Essay*, could have been influenced by *Emile* to any significant degree. Rousseau wrote three of his major works at that time: *Julie* in 1761, the *Social Contract* in the spring of 1762 and *Emile* in the summer that same year. Theoretically, Kant could have read all these works during the first year and Rischmüller, for example, regards the idea that Kant would not have been familiar with the *Emile* until he wrote the *Observations* as unlikely. Others find that the “striking difference” between the *Prize Essay* and the *Observations* – and especially the remarks to it – show that Kant must have received some new influences.  

It is not a coincidence that the majority of Kant’s references to Rousseau relate to the theory of human nature or cultural evolution since Rousseau’s theory of the character of humanity was said to be “the most distinguished of all”. (*VA* [Friedländer] 25:675) Another work from the same early era that obviously bears the traces of Rousseau is the essay *On the Maladies of the Head* (1764) in which Kant touches upon the topic of corruption by means of society similar to the way Rousseau portrayed it in his two first *Discourses*.  

The period between 1770 and 1781 has often been called Kant’s “silent decade” since only two short works and one review appeared during these years. On the other hand, there are a huge number of reflections from this period and he still taught regularly at the University. But, most of all, this was the time in which he developed what would be his greatest book – the *Critique of Pure Reason*. Although the *Critique* was certainly not meant to be a work on moral philosophy, one of the conclusions of this critical evaluation of the powers of reason was that pure theoretical – or *speculative* – reason could never transcend the limits of possible experience. In order to get the idea of things like *freedom*, *God* or an *immortal soul* – which were the postulates that Kant needed for his arguments for the possibility of the highest good – he needed to show that pure reason can also be *practical*.

---

35 Although L. Collins’ notes have typically been dated to 1784 (and hence contemporary with the *Groundwork*) they are said to bear striking similarities with a group of notes from 1774/75 or perhaps 77/78. Hence, it is more likely that they are from this earlier semester, perhaps already winter 72 when he also studied anthropology. For discussion of this dating see the excellent database “Kant in the classroom” from Manchester University where there are further references. (see http://www.manchester.edu/kant/Home/Index.htm).

36 Rischmüller p. 150. It is sometime noted that Kant could read French so he did not need to await the German translations.


38 After his *Inaugural Dissertation* (1770) Kant reviewed Peter Moscati’s ideas concerning the essential corporal differences between animals and human beings (1771). The two published texts are a note on the different races of human beings (1775) and a comment on Basedow’s school of philanthropy (1776-7).
Kant attends this problem at the end of this work, in the section called the “Canon of pure reason”.

Kant’s first real work tackling in a detailed manner the true foundation of moral philosophy was the *Groundwork for the Metaphysics of Morals* (1785). In this highly metaphysical work he stresses the need for establishing a pure and transcendental origin of morality. Since the work actually never reached the point he wanted, namely to advance (successfully) from a purely metaphysical theory to a “critique of pure practical reason”, he continued the work in the *Critique of Practical Reason*. (1788) About the same time he also established what was required for the possibility of orienting oneself in thinking by explaining the significance of a subjective “need of reason”, namely reason’s felt need [Bedürfnis] to assume the existence of certain ideas in order to be able to navigate within an area that had hitherto been “filled with dark night”. (DO 8:137) He also developed his concept of maxims and their determining ground and discloses the principal failings of all previous attempts to reach the ultimate understanding of the highest good. In spite of some central divergences between the *Groundwork* and the *Critique*, the main premises remain the same and both declare that there is only one incentive to morality, namely respect, and that the moral person must regard himself as a being entirely free from natural causality – that is, as a *homo noumenon*.

One looks in vain for Rousseau’s name here. If one’s reading of Kant is limited to these “critical” works, it definitely appears as if he had found a line in his moral thinking in which he neither seemed to be charmed by beautiful prose, nor to be fascinated by stimulating reflections on human history. On the contrary, Kant came to argue that ethics must be based on purely metaphysical principles and not on the capricious features of imperfect beings; in other words, ethics must in no way borrow anything from human nature. However, as Høffding’s observes, one must not neglect the fact that the *Groundwork* – the chief and most abstract of Kant’s moral writings – was written between two essentially anthropological works namely the *Ideas for a Universal History* (1784) and the *Conjectural Beginning of Human History* (1786). Obviously, Kant had not lost his interest in practical anthropology, and he could apparently easily shift focus from pure to applied ethics.\(^3^9\) The arrival of these two works could thus, in some sense, also be seen as a “return” to his interest in history; an interest that he had formerly expressed in his early thoughts about history of the world in the *Allgemeine Naturgeschichte*, which now became an interest in the history of man and society.

In the third *Critique*, the *Critique of the Power of Judgment* (1790), Kant did, indeed, discuss teleology and the different kinds of cultivation that human beings must undertake in order to become the fully human being he is

---

\(^{3^9}\) Høffding (1893) p. 13.
destined to be. Yet, Kant’s most straightforward return to the theory of man as moral being and the problem of agency came with the publication of the *Religion within the limits of mere Reason* (1793). While the principle that will be forever associated with it, the principle that claims that “man is evil by nature”, seems to express the opposite to the one we find in Rousseau’s writings, the *Religion* is a highly Rousseauian work. It is, like the *Second Discourse*, essentially a *theodicy* explaining man’s fall into evil and why this fall could take place in a being whose original predispositions always must be considered as entirely good. Kant’s answer is, not surprisingly, that evil originates in the human being’s *free will* and that no one else can be blamed for this shortcoming. The *Religion* is also an anthropological text which embodies much of what Kant wanted to put under the label *moral or practical anthropology*. The *Religion* thus anticipated much of what should later appear in the book that is generally regarded as Kant’s real work on anthropology, namely the *Anthropology from a Pragmatic point of View* published in 1798.\(^\text{40}\) Still, as Norbert Hinske points out, if one really wants to understand Kant’s anthropology and his intent behind it, one must not believe that it is enough to read this late publication from 1798. Rather, in order to get the full scope and complexity of Kant’s thoughts on this matter one must consider the vast material from Kant’s entire production, including his work on history and the lecture notes as well as his unpublished reflections.\(^\text{41}\) Anthropology is, after all, not a distinct theme but must rather be seen as a thread that connects most of the topics Kant dealt with, even though he sometimes argued that we must, at least temporarily, take no account of it.

Another discipline tightly connected to the science of man is *pedagogy*. Since man is by nature unrefined, he is in need of improvement in order to be the full-fledged being that providence has destined him to be. *Emile*, which justifiably can be regarded as Rousseau’s magnum opus, is probably one of the best-known philosophical works on education that were written during the eighteenth century and, as we have seen, Kant read this work with pleasure and passion. Also, Kant’s corpus includes a work called *Pedagogy* and he lectured on education.

Kant’s *Pädagogik* was published 1803 and was edited by Kant’s friend T. F. Rink who also published Kant’s manuscripts on *Physical Geography*. Yet, there is much debate with respect to the reliability and origin of the work. While the manuscript is supposed to reflect the lectures on pedagogy that Kant had held some decades earlier, it has been argued that the text is fairly un-Kantian in both style and structure, and Traugott Weisskopf, the author of the most extensive commentary to the *Pedagogy* even claims that the work

\(^{40}\) This text is generally believed to be based on Kant’s own manuscript for some of his latest semesters as professor. It seems, in fact, as if Kant had planned to publish a real work on anthropology earlier than that. Some of his correspondents claim to be looking forward to it already in 1794. Cf. Letter von Mellin no. 622 (11:498) and to Stäudlin no.629 (11:508).

\(^{41}\) Hinske 425, see also Schwaiger p. 100.
should be removed from Kant’s corpus. The Pedagogy, Weisskopf says, seems basically to be a compilation of parts from Kant’s lectures on ethics and anthropology combined with some loose reflections plus various fragments from Rousseau’s Emile. The presentation, he says, is generally unsystematic and seems to be stylistically reworked throughout. Yet, he admits that since the original manuscripts of Kant’s lectures on scholastico-practicum are no longer available, we cannot say anything about Rink’s edition except mere hypothetical conjectures. Yet, one may also, as Lewis White Beck does, regard the presence of old matter in the Pedagogy as a sign of an echt-Kantianische origin of the views presented. Since the Pädagogik is frequently cited in modern discussions of Kant’s educational philosophy I will, therefore use material from it in my discussion.

Finally, let me briefly say something about the area in which Kant’s indebtedness to Rousseau is impossible to ignore, namely political philosophy. The end of the 1700’s was a turbulent time, especially with regard to the political happenings in Europe. Kant was, of course, not indifferent with regard to this and published some treatises that touched upon topics such as revolution, just government, freedom of speech and the possibility of peace. Kant’s work Perpetual Peace not only mirrors the title of a work of Rousseau but it also applies the Rousseauian notion of a general will [volonté général], a term which came to play a central role in Kant’s idea of justice and civic right. It is clear that this notion is essential for Rousseau’s idea of the just civilian society and that Kant’s adoption of it shows that he found Rousseau’s ideas attractive. Now, since the “general will” bears within its concept the idea of oneself as the author of the laws that are commanded in the name of this general willing, a subjection to such laws is an expres-

---

42 Weisskopf p. 349.
43 Weisskopf p. 287ff.
44 Weisskopf pp. 330, 331. The lectures in question were held four semesters; 1776/77, 1780, 1783/84 and 1786/87. For a discussion of the “unsystematic appearance” and the philosophical reliability of Päd. see also Louden (2002) p. 34ff.
46 It may here be noted that Felicitas Munzel’s recent and voluminous book Kant’s conception of Education (2012) does not mention anything about the possible unreliability of the work.
47 Among Kant’s political works one can mention the small essay What is Enlightenment (1784), “On the Common saying: this may be correct in Theory but it of no use in Practice” (1793), Towards Perpetual Peace (1794) and the first part of the Metaphysics of Morals – the “Doctrine of Right”. (1797) as well as part II of the Conflict of the Faculties.
48 Rousseau’s two works Abstract of Monsieur de Abbé de Saint Pierre’s Plan for Perpetual Peace (1761) and the posthumously published Judgement of a Plan for Perpetual Peace discussed Saint Pierre’s work Project pour render la paix perpétuelle en Europe. (1713) Kant was certainly well aware of these writings since he frequently mentions Saint Pierre and Rousseau together. In the Theory and Practice Kant’s reference to them undoubtedly alludes to their ideas of peace. TP 8:313. Cf. also IAG 8:24.
sion of one’s freedom. 49 Already in the Dreams of a Spirit-Seer (1766) Kant refers to this general will though not in terms of a guide for rightness in the political order. Rather, it is given the role of a moral guide, or a moral feeling [sittliche Gefühl] aimed at making our private wills agree with this general will. (TG 2:335) Yet, he added, this is only the effect of a process whose causes we do not yet know. Nevertheless, this is not that far from the theory that we all know, namely the idea of a good will whose principle is a categorical imperative that, in turn, is accessible to our understanding by the insight of our autonomy.

This is certainly a call from within. In a way, this awareness of ourselves as rational and free can be seen, not directly as the end, but rather as the closing of the circle in which human history has progressed. Surely, Kant might have had a “Rousseauian turn” during the early 1760s that was later followed by a transcendental /critical one. This latter turn, however radical it may appear, is not a sign of a complete turn-away from Rousseau. Man and nature are ongoing themes in Kant’s long life as thinker, teacher and human being.

49 Cf. Rousseau’s claim “to follow the law that one has made oneself is freedom” (SC I:viii) and Kant’s statement in the Groundwork that “a free will and a will under moral laws is one and the same” (GMS 4:447).
PART I: Natural and Moral Beings
1. The Science of Man

At the Temple of Delphi there was a short and clear-cut inscription – *know thyself*. This decree, which “contained a precept more difficult and more important than is to be found in all the huge volumes that moralists have ever written” expressed Rousseau’s earliest motto. *(2D 12/122)* While he did never explicitly explain *why* these words appealed so much to him, his overall philosophy shows that what they expressed really mattered to him. Perhaps it is a coincidence that Rousseau’s favourite scientist – Carl von Linné – put this phrase next to the headline “homo sapiens” in the *Systema Naturae*, indicating that self-knowledge was inseparable from the concept of man.*[50]* Although Rousseau should later give up on this task (it turned out to be much more difficult to follow than he had thought)*[51]* it essentially expresses his main premise that the present man does not know his true nature. Worse still is that man is also completely unaware of this lack of knowledge. This sentence also served as the sole maxim by which Rousseau found it possible to answer the question asked by the Academy of Dijon, namely “*What is the origin of inequality among men, and is this authorized by natural law?*” For Rousseau it was evident that one cannot understand the present state and its problems without knowledge about the chain of events that had gradually led to this condition. We must also ignore for a while all those books that only consider man as *he has made himself* and look at the man who is void of everything but his most basic instincts. Rousseau’s reflections on this topic ended up in the controversial text usually referred to as the *Second Discourse*.

Rousseau begins the *Second Discourse* by identifying the science of man as “the most useful but still least advanced of all human knowledge”. *(2D 12/122)* The same opinion is echoed in the earliest notes from Kant’s lec-

---

*50* Linné (or sometimes Linnaeus) was a Swedish natural scientist whose works on biology and systematics are still generally regarded as revolutionary. The *Systema Naturae (On the System of Nature)* appeared for the first time 1739; the 10th edition 1758. We know from the *Confessions* and the *Reveries* that Rousseau adored Linné and that he had a copy of his book but it is not clear if he had read it already when he wrote the *Second Discourse*. The idea of self-knowledge for man and his relation to himself and God also appears in Linné’s latter text *Nemesis Divina* (“If I cannot take measure of myself, it is hardly strange that I should be unable to grasp God”).

*51* See *RSW* IV. During this walk Rousseau explains his newfound fascination for the concept of *truth* and therefore he replaced his older motto with the Juvenal’s principle *vitam innpendere vero* – I consecrate my life to truth.
tures on anthropology wherein Ludwig Collins writes that nothing could possibly be more interesting for us than this highly neglected science. (VA [Collins] 25:7) Although the term “anthropology” may have been fairly new at that time, it is obvious that human nature and its capacities have been some of the most fundamental topics in the entire history of philosophy. No matter how abstract and metaphysical the questions of philosophy may be, we have to relate them to one or another conception of human soul, mind or understanding. Many of the most read works at this time discussed the science of man and his intellectual, social and moral faculties. For example David Hume said:

’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. 52

This agrees well with Kant’s idea that metaphysics, morality and religion are in fact anthropological disciplines since they can only be understood as related to the human being. 53 In Pilau’s anthropology notes we learn that studies of human nature provide us with principles that have great influence in morals, religion and also education. (VA [Pilau] 25:735) The last addition is important and leads to the idea of a vocation, or ultimate purpose, of human existence. For Kant there was no doubt that there had to be such a vocation, but his description of it is not as clear as one would want; in particular, his idea of the necessary progress towards it generates numerous new difficulties that are often left unresolved. However, the import of anthropological research is unquestionable for in the Conflict of the Faculties we are told that:

Philosophy is the science of the human being, his representations, thoughts and actions; it should present all the components of the human being both as he is and as he should be, that is, in terms both of his natural functions and of his relations of morality and freedom. (SF 7:69)

Moreover, in his Pedagogy Kant frankly states that “the human being can only become human through education”. (Päd 9:443) He continues:

It is the human being himself who is supposed to first develop his predispositions towards the good. Providence has not placed them already finished in

52 Hume Treatise Introduction §4.
53 Again recall the fourth and anthropological question (What is the human being?) to which the former three (What can I know? (metaphysics) What should I do? (ethics) What may I hope? (religion)) were said to be essentially related. (VL [Jäsche] 9:25)
him; they are mere predispositions and without the distinctions of morality. The human being shall make himself better, cultivate himself, and, if he is evil, bring forth morality in himself. (Päd 9:446)

Thus it becomes clear that Kant’s moral anthropology does not essentially deal with the question what human beings are but rather what we ought to be. Still, since the capacity of fulfilling these commands depends on our natural constitution, we need to know the nature of this constitution too. Kant clearly states that moral progress and the fulfilment of one’s vocation is a duty and duty commands nothing that we cannot do. However, given his pessimistic idea that there is a deeply rooted propensity to evil within us and that men are made of “such crooked wood from which nothing perfectly straight could ever be timbered” it is not easy to see how we ever could be good enough to meet those demands. (Rel. 6:100; IAG 8:23) Kant’s ideal for the moral person is high; his idea of the highest good of the human species is even higher. I will return to these questions, but before that we need to look at the general nature of the human being.

In the following discussion I have chosen to highlight the three main aspects or stages that illustrate the several features of the human being and which influence the understanding of more specific traits and dispositions. These aspects are the natural, the social and the moral, which applies to three main forms of the human being. For Rousseau, these headlines essentially express the historical evolution of the species from the unrefined naturalness of the savage to the human being as citizen; a being that is prepared for a moral way of living within his present civic state. For Kant, they represent different aspects of the individual human being that are simultaneously present in everyone and which explain man from his lowest to his highest predisposition. Yet, as will become clear, both perspectives can be found in each philosophical system so the difference just mentioned is nothing but a simplification. Also Rousseau saw the human being as composed of sensibility and rationality – as a being who is constantly affected by conflicting impulses and someone who needed ruling principles for the right conduct of its own will. Moreover, even Kant had a great interest in human history and especially the foundation of the legitimate political order. Both agree that since virtually everything that is essential to the true human condition can only be developed in a social context, one must also consider the path from crudity to civilization and the conditions that made this transformation necessary.

Finally, before entering into the discussion about the different aspects of the human being, it seems reasonable to say something in advance about another term that has been frequently present in connection to each aspect of man, namely nature. As we will see, the term can be used in a number of ways which differ both with respect to its applicability to the various forms of man and with regard to its normative status. While for Rousseau “nature”
was a thoroughly positive epithet that represented innocence and sincerity and whose existence – regrettably – had vanished in the present man, Kant’s conception of the term is far more complex.

In one sense, nature can be interpreted purely anthropologically and then it would, as Kant says, “stand in direct contradiction to the predicates morally good or morally evil” since nature is the opposite of freedom. If nature is taken in this sense, man is neither good, nor evil by nature for he is by nature not a moral being. (Päd. 9:492) However, Kant says “nature” may also, at least in the connotation of the nature of a human being, refer to the “subjective ground” of the “exercise of the human being’s freedom in general” and thus be the explanation of all good or evil deeds expressed in the world of sense. (Rel. 6:21) In this sense, man must be considered evil by nature. This reminds us of Kant’s familiar distinction between the human being as phaenomenon and as noumenon: between man in the world of sense (nature) and man in the world of understanding (the intelligible world). While the former is subject to the causality of nature, the latter is not. This is why morality, and accordingly the source of good and evil, must not be sought in anything that belongs to man as a merely natural being but only when he, through freedom, realizes that he stands under moral laws. Still, though that which belongs to mere “nature” should not be dismissed as unimportant, it cannot – by itself – lead us towards the essence of the human being.

1.1. Natural Beings

For Rousseau, the doctrine of man is the heart of his philosophy. “It is of man I want to speak,” he says in the introduction to the Second Discourse. The primacy of anthropology becomes even more obvious when he, in the Social Contract, insists on the importance of creating a society fit for men instead of forcing human beings to conform to artificial laws or conditions. Moreover, the whole discussion in Emile concerns the proper upbringing of an undestroyed soul and here Rousseau teaches us how to cultivate the good seeds already embedded in the child without thereby also planting any dangerous weed.

The title page of the Second Discourse pictures a frontispiece accompanied by a sentence from Aristotle’s Politics, telling us that “not in corrupt things, but in those which are well ordered in accordance with nature, should one consider that which is natural”. 54 Obviously, Rousseau did not think that our present (corrupted) state could provide us with those genuine objects, beings and conditions that one would need for such an investigation. There-

---

54 Rousseau has transcribed the quote in Latin: Non indepravitas, sed in his quae bene secundum naturam se habent, considerandum est quid sit natural (from Politics 1254a 36-7). The English translation is from the editorial note in the Collected Writings vol. III.
fore, Rousseau asks, “what experiments would be necessary to achieve knowledge of natural man? And which are the means for making these experiments in the midst of society?” (2D preface 13/123-4)

In order to answer the question on the origin of the vices of the present state we must find a starting point where men had no idea of inequality. The only state that could fulfill these premises was a state of solitude. Rousseau’s experiment thus begins by his imagining the most simple and primitive state in which only the most basic operations of the human mind were present. It is an ideal world, Rousseau says, “similar to ours but yet altogether different”. (RIJJ I 9/668 emphasis added)

The duties of man in the state of nature are always subordinate to the care of his own preservation, which, Rousseau says, “is the first and strongest of all.” (PF 17/475) The natural man, or simply the savage, lived a solitary life in the woods, peaceful and happy as long as his basic needs were satisfied. He saw himself as an animal among the others and if he occasionally met someone else, he ran away rather than fought. He was, as we are now, led by passion but not by the same as moves us. Natural man is basically affected by one main sentiment – the sentiment of his own existence. (2D II 66/193)

Tied to this sentiment of existence was the care for his own well-being. As Rousseau says, all animals possess the faculties needed for their preservation; only in man are these faculties excessive. (E II 81/305) Before man developed his fatal and excessive sense of his own importance he was guided by this self-preservation and proper self-respect, two principles that constituted “love of self” in the most basic sense, namely amour de soi. This love of self is the original passion and all further principles of the soul are nothing but modifications derived from it. This notion does not include any sign of one’s own priority to others since there were no such things as social interaction or property and hence there could be no envy or jealousy between men. In this sense, self-love could by no means be called egoistic; it was rather the principal rule according to which all men guided themselves. To see, to feel and to will or not will were the only operations of the savage soul. His desires never exceeded his physical needs. Since his sentiment was limited to the present, he had no thought about the future. (2D I 28/144)

In addition to amour de soi, there was another basic feeling existing in all human beings, namely natural pity expressed through the “natural repugnancy of seeing another sensitive being suffer or die”. This is revealed through the fact that we all respond with distress to the sight of a dead body and even brutes show significant signs of this instinct. (2D intro. 15/126.) As long as man was led by these two principles alone, he would never hurt anyone else and he would respect others as much as he respected himself. Even if this primitive natural goodness [bonté naturelle] was not capable of giving us the sublime maxim of rational justice – “do to others as you would have them do

55 E IV. 213/491, RJJJ I 10/670, II 113/805, LB 28/935.
unto you” – it taught us a more useful principle, namely “do good to yourself with as little evil as possible to others”. (2D I 37/156) Natural pity is thus the impulse which hinders us from doing harm to others and it can be seen as an extension of the instinct of self-preservation to a preservation of the whole species. As a natural instinct, it can be regarded as a voice totally independent of reflection or a voice that also served as a substitute for moral norms. Thus we can, as Rousseau puts it, “set down as an incontestable maxim that the first movements of nature are always right”. Since there is no original perversity in the human heart there can be no vices in it of which it cannot be said how they entered.56 This state was simply characterized by authenticity and transparency and it was, thus, far from the Hobbesian picture of a state of “continual fear and danger of violent death, and the life of man, solitary, poor, nasty, brutish, and short”.57 This, according to Rousseau, “horrible and absurd doctrine” was the result of Hobbes’ confusing the natural man with the man he had before his eyes.58 The man of today bears only vague similarities with the original man, like a statue that has been eroded by time, wind and water.

A common criticism against Rousseau’s basic premises is that there is no support for his idea of man as a naturally unsocial being and that today’s science rather points towards the opposite. Rousseau’s picture of the state of nature is thus said to be based on a false premise that, even were it true, would be impossible to substantiate. To this one may reply that Rousseau did, after all, not claim that this state had ever existed – not even before the fall. On the contrary, he replied that all such empirical facts could simply be “set aside” since they “do not affect the question”. (2D I 19/132) What Rousseau wanted to do was to present an idea of the true state of nature. This was a theory that aimed to correct the mistaken attempts presented by Hobbes and Locke who, by the notion of a “state of nature”, had rather pictured a social state in its early and lawless form; a state in which vices of culture had already developed. Their misguided ideas of the natural state was the reason why they both saw the need for a sovereign, or a civil government, to cope with the prevalent evil and hostility among what they took to be natural men. Admittedly, Rousseau also found himself obliged to accept the need for a social pact, but he did not agree that this need was a consequence of any natural inclinations of men. Still, Rousseau was, of course, not the only one who had concerns about the conflicts between men in the social state. Kant would take up much of it in his idea of unsociable sociability. Moreover, before him, Montaigne had in his reflection “on Solitude”

56 E II 92f./322. Similar FM 38/97.
57 Hobbes Lev. XIII 9. For an extensive treatise of Rousseau’s idea of authenticity and transparency, see Starobinski.
stated that “There is nothing so unsociable and so sociable as man: the one by his vice, the other by his nature”.59

Arguing from hypotheses about the earliest stages of human history is not unique for Rousseau. While anyone presenting such ideas would generally run the risk of becoming the enemy of the church and the Bible’s monopole with regard to this topic, conjectures about human history were not uncommon at this time. Kant wrote two works on human history in which the influence of Rousseau is impossible to ignore: *Idea for a universal History with a Cosmopolitan aim* and *Conjectural beginning of Human History*. When these works were written Kant had recently reviewed Herder’s *Ideen zur Philosophie der Geschichte der Menschheit* in which he commented on his colleague’s way of passing through this “untrodden desert”.60 His own works on the history of mankind could thus be seen as his own pathway into a new area of thinking. A central conclusion that Kant reached during this research was the one that seriously disturbed Herder; namely that the human being is an animal that needs a master.61

As will become clear in the following sections, Kant held that a human being cannot stay in an undeveloped, or half-finished, phase. That having been said, he did maintain that man, in his crudest form requires various forms of culture before he is ready for a just social state. Thus, man’s destiny to enter a social order disclosed the need for taming the savagery and of cultivating the germs that are exclusive to human beings. According to Kant, we must assume that the natural vocation of humanity consists in continual progress toward the better, for it is only on that premise that we may love the humankind considered as a species. For, he adds, as we necessarily hate what is and remains evil, a progress is required in order to allow for the love of humankind. Fortunately, the assumption of a possible progress seems to be more plausible and an endless regress is, Kant says, impossible simply because one would sooner or later hit the ultimate bottom where the human species would destroy itself. Moreover, human history can neither consist in a standstill, nor in an oscillation between good and bad states since building up just in order to once again fall back would resemble the struggles of Sisyphus whose crimes we no longer remember. The latter alternative would, in fact, not be different from the standstill, and because of the fruitlessness of all moral efforts, striving towards the better would thus be a ridiculous waste of time. Still, since the human vocation requires the highest form of culture (i.e. moralization) we are, therefore, allowed to assume that the human species is continuously progressing. (*SF* 7:81, *TP* 8:307-8)

60 See AA 8:64. Herder’s work appeared 1784.
61 IAG 8:23. In a comment to Herder’s dissatisfaction with this pessimistic, misanthropic and even evil view, Kant proclaims the importance of admitting the exploring writer full freedom in this examination. *Review of Herder* 8:64.
Yet, for the moment, we are only interested in the natural being. In contrast to Rousseau, Kant did not see the purely natural man as an ideal but rather an idea. A Rousseauian state of nature could, in Kant’s view, be equated with the so-called “golden age” about which the poets wrote their romantic stories. These lost paradises, which most of all reflected the human species’ dissatisfaction with its current state, became in these verses mere objects of “empty longing”. The reason why all wishes for such states had to be seen as empty was that everyone, after all, was well aware that such a pleasant condition was nothing but a “shadow image” about which we know nothing – not even if it had ever existed. Since Kant did not begin at this fancied stage, his savage man also differs significantly from the one offered by Rousseau. When Kant talks about savages [Wilden], he refers to those men who existed in the pre-governmental state (before legislation) or to such human beings who may still be living in primitive countries. Such men are brute and uncivilized – only marginally different from animals.

Nevertheless, we must not forget that all human beings, even the most cultivated ones, are still partly animals, since no one can ever escape the fact that he is a living being with both natural and artificial needs. The satisfaction of these needs is secured through our predispositions to animality [Tierheit] and humanity [Menschheit] – both of which can be seen as forms of self-love. The first, which does not require reason, is purely physical or mechanical and it merely provides us with the drives that we need in order to protect ourselves and to propagate the species. This natural self-love also involves a basic instinct to seek company and enter societies. (Rel 6:26-7) In the Anthropology, the predisposition to animality is called technical and is contrasted to the pragmatic and the moral ones.

Kant’s relatively spare outline of the idea of a “natural man” is probably a result of his view that the essence of a human being is not the mere living and fulfilling of basic needs in a physical world. It is also, he says, difficult to imagine the very concept of such a first human being. (Päd 9:447) In contrast to Rousseau, who did not “anatomize” human nature further than explaining some basic distinctions, Kant presented a detailed plan over the human being’s general constitution including the mind’s several parts and their mutual relationship. Accordingly, when discussing Kant’s anthropological notion of “human nature” one must distinguish between what belongs to our nature as mere living beings and what constitutes our nature as rational, free and moral beings. In the first group, which is generally labelled sensibility, [our “Sinnlichkeit” or “Sinnesart”] we find natural aptitudes or simply “nature” [Naturell], instincts, talents, temperament, predispositions [Anlage],

---

62 MAM 8:122. For a discussion of this idea, see Galston pp. 72-3.
63 Anthr. 7:322ff. In the Religion, the moral predisposition is named personality.
mind [Gemüt], soul [Seele] and heart. In this group, one may perhaps also include mentality [Gemütsart, Gemütsverfassung]. When inquiring into the moral being, one needs to be familiar with concepts such as Gesinnung, moral or intelligible character, will [Wille] and faculty of choice [Willkür]. We will return to the function of these faculties in the following chapters.

Even if Kant is often thought of as a systematic thinker, it is not always easy to orient oneself in his terminological jungle. Some of these notions can be seen as umbrella terms which include a number of the others and some of them are also more or less identical. For example, sensibility (which is sometimes identified with our entire sensible or empirical character) includes our natural aptitude which, in turn, covers (among other things) talents and temperament. Talents, temperament and predispositions are gifts of nature or fortune while character has to be acquired. However, also the natural aptitudes can and ought to be refined by means of culture and this is the first step from brutishness to civilization. As we will see, not even the highest degree of cultivation of talents and temperament can make the human being moral. In order to reach the highest form of humanity we need moral culture or what Kant calls moralization.

1.2. Social Beings

The natural man in Rousseau’s state of nature had no excessive needs; as long as he was sated, free from pain and had well slept, he was the happiest being on earth. While Rousseau explicitly left out all speculations about whether the original man had fur and claws, if he grazed or went on all fours, Voltaire’s well known and ironic description of the Second Discourse as “a work against the human race” which reduced men to stupid animals, expresses the caricature so commonly attributed to Rousseau’s thoughts. While Rousseau, indeed, regretted that he had lost “a little bit” of his own stupidity he made clear that a return to this state of mere animal living was not what he wished for. In fact, the evils that we had brought upon ourselves through the path to this new way of living could, in some sense, be likened

---

64 Even if Kant generally stresses the importance of not conflating the terms “mind” and “heart” with our moral nature, he sometimes uses them in a more inclusive or perhaps “poetic” sense. A good heart, for instance is usually not enough for being a morally good human being since only goodness of character is fit for this. Elsewhere, heart is more or less identical to our moral self or character and “a change of heart” is the only way of becoming a truly moral being. I will return to this in chapter 4.

65 Voltaire, “Letter to Rousseau Aug. 30, 1755 contemporary”. In this letter, Voltaire made his legendary comment that the Discourse made him want to “walk on all fours” even though it was more than 60 years since he gave up this habit. (CW vol. III 102/OC III 158) To this Rousseau replied that he knew no one who stood as steady on his two feet as Voltaire and that it would require a miracle to make such a genius mind fall back into animal stupidity. For Rousseau’s reply, see CW vol. III 105/OC III 227.
to a knife that must be left in the wound in order to prevent the stabbed person from bleeding to death. What he regretted the most was the sharp turn that humanity had taken. In this learned century, Rousseau said, we adopted instructions from those who, in fact, were unfit to give them, as if the lame would wish to teach others how to walk. Science, he said, is perhaps not made for man since his limited mind makes him incapable of handling it. Accordingly, Rousseau concluded: education generally "corrupts [man’s] morals, impairs his health, destroys his temperament, and often spoils his reason: if it teaches him something, I would consider him very poorly compensated".

The simple reading of Rousseau’s picture of the nature of the natural man leaves us with the question how and why the transition from nature to society could take place. If man actually is (or at least was) such a timid, modest and happy being that the original premise claims, it is hard to see why he would willingly choose to leave this state of peacefulness and make himself a subject of social bonds with all the costs involved. This is especially confusing if we recall that Rousseau explicitly said that all characteristics of men could be explained “without introducing the idea of sociability". However, this early description does not reflect the whole story and one element that is sometimes overlooked is Rousseau’s later claim that “it cannot be doubted that man by his nature is sociable or at least made to become so”. If one considers the last part of the quote, that we were made to

---

66 See Rousseau’s reply to Voltaire Sept. 10 (1755) CW vol. III 106/OC III 227.
67 See Rousseau’s reply to comments made to the First Discourse (in CW II 38/OC III 36). In the Preface to Narcissus he asks rhetorically: “What have we gained by this?” and replies “much babble, some wealthy people and some argumentative people, that is to say, enemies of virtue and common sense”. (PN II 194/969)
68 E IV 290/600, similar in ML V 196/OC 1109. Admittedly, the reference from Émile belongs to the long section (or perhaps “interlude”) in book IV entitled the “Profession of Faith of a Savoyard Vicar” (E IV 266-313/OC 565-635). It has, however, been objected that this section, though explicit and powerful, does not express Rousseau’s own view. The fact that he puts these words in the mouth of the vicar might indicate that he just wanted to ventilate some general ideas. Indeed, there are passages where Rousseau defends himself from criticism by pointing out that some of the words that had upset people were not his own and they should not blame the mere “editor” for what the “author” says (see e.g. Letter to Beaumont and the Letters from the Mountains). It has also been claimed that Emile never heard the “profession” which indicates that the content of it was not acknowledged by his tutor (i.e. Jean-Jacques). Although I will not directly discuss this question here, I will at least say that even if arguments for a separation of the vicar’s and Rousseau’s respective views are widespread (see e.g. Melzer (1990) p. 30, (1996) p. 355; Horowitz, p. 139; Bloom, preface; Marks (2006) p. 567), there are also several reasons for accepting the vicar’s claims about at least some matters as an outline of Rousseau’s own ideas. It should also be noted that the Profession is not a review of one person’s ideas. In the Confessions we see that Rousseau composed this section from two priests that he found “truly worthy of attachment and esteem” and he flattered himself that his “imitation did not dishonour the original”. (Conf. III 119, cf. the editorial note on p. 615). There are also several similar (and almost identical), sections, especially in the Moral Letters but two also in the rest of Émile, in the Dialogues, and in the Letter to Beaumont. In the third walk of the Reveries, Rousseau explicitly talks about the Profession as a work that he is proud of and which “one day make a revolution among men, if good sense and good faith
become sociable – there seems to be no obvious contradiction involved. Even if the path to society was indeed a coincidence (and perhaps also to be seen as a misfortune), the accidental event from which our sociability was born was the simple fact that we met and had to live together within a limited area. This closeness made us aware of our weakness and dependency. At the beginning of Emile Rousseau states that “we are born weak, we need strength; we are born totally unimproved, we need aid; we are born stupid, we need judgement.” (E I 38/247) Later he adds that it is just this weakness [faiblesse], or better vulnerability, that makes us sociable since, as he puts it:

> Every attachment is a sign of insufficiency [insuffisance]. If each of us had no need of others, we should hardly think of associating with them. Thus from our very weakness is born our frail happiness. [Ainsi de notre infirmité même naît notre frêle bonheur] A truly happy being is a solitary being. God alone enjoys an absolute happiness; but which of us has any idea of it? If any imperfect being could be sufficient to itself, what according to us would he be able to enjoy? He would be alone, he would be miserable. I do not conceive how one who has no need of anything could love anything; I do not conceive how he who loves nothing could be happy. (E IV 221/503)

Our needs thus made us fit for society and the only thing that is required for the claim that we were “made to become sociable” is that we would become so under proper circumstances. Man’s sociable germ was originally not active but latent and our history shows that he had a soul that was receptive to various new kinds of feelings. Most importantly, these remarks make clear that a human being cannot be truly happy unless he cares for the things that make him content. What the savage man feels is only some rudimentary sentiments of pleasantness. As compared to the modern man, the natural man is, therefore to be considered poor with regard to genuine human characteristics.

Although Rousseau is the philosopher who is generally associated with the idea of man as an originally solitary being, this idea is, in fact, also suggested by Kant. As Kant himself puts it, there are several questions that can be asked about man in his first shape, but the most important one is whether this being was a sociable animal or a solitary one that shied away from his neighbours. To this question Kant replies that the latter is the most probable (Anthr. 7:322) However, our higher destiny requires that we break away from this crude state of mere animality. Kant’s reasons for disproving the state of nature as the appropriate state for human beings are based on his strong belief in the purposiveness of the creation. We may once again recall his strong esteem for Newtonian science which was said to establish the...
lawfulness of nature that, as we saw, via Rousseau, led Kant so see a corresponding lawfulness to human nature.\textsuperscript{69} Since we have the predispositions for higher faculties, these must sometimes develop themselves for, as he puts it, an organ that is not meant to be used or an arrangement that does not attain its end, is a contradiction in the teleological doctrine of nature.\textsuperscript{70}

The savage may be uncivilized but he is still potentially rational in virtue of his predisposition. It is thus important to distinguish between the animale rationabile and the animale rationale of which the latter is someone who has realized that he can be the master over his impulses and also uses this knowledge for the advancement of himself as well as for the entire species and society.\textsuperscript{71} Without going into the details of Kant’s theory of teleology, we must at least find out what he held to be the end of human history. It is clear that there is a vocation for the human species though this is not, as in the case of most species, a fixed one that is fulfilled more or less automatically. Human beings are complex but most of all they are unpredictable. It is because of this capriciousness that Kant found himself forced to conclude that the human being needs a master. This, Kant says, does not mean that man is like the cattle – which must subject to their herd – but rather that he is like a bee, namely a creature that can only fulfill its true purpose within a society.

In short, Kant says, the natural vocation of humanity can only be reached in a civil state and this vocation consists in the highest culture, no matter how strongly our brutish nature tries to resist it. On the other hand, not even a natural vocation is reached automatically; on the contrary, it requires the help of inclinations (acquired habitual desires) whose powers exceed the force of our natural (and innate) instincts. According to Kant, the most important inclinations for this progress towards the vocation of our species are, quite paradoxically, unsociability and envy.\textsuperscript{71}

The fact that envy and unsociability – which often takes the form of hostility – are the main springs for the progress towards the better (morality) demonstrates with clarity that the pathway towards the highest vocation is not straightforward. Even if Kant says that the question whether the species has gained or lost through this process is no longer relevant, we must not ignore the obvious consequences of it. These considerations obviously echo Rousseau’s pessimism about the social process. We cultivate the raw nature in order to reach something better; men need strength and judgement in the same way as a tree needs soil and water. Yet, Rousseau’s opening lines of Emile – “Everything is good when it leaves the hands of the Author of things; everything degenerates in the hands of man” – clearly shows his wor-

\textsuperscript{69} Cf. again BB 20:58-9 (quoted in the introduction).

\textsuperscript{70} IAG 8:18. For a fuller account of the role of teleology in Kant’s idea of human history, see Alison (2009).

\textsuperscript{71} Refl. no. 1521 (15:885) Loses Blatt. The dating is vague but is said to be from 1780s so it is contemporary with the historical works as well as with the Groundwork.
ries about our encroachment on nature. (E I 37/245) What unifies Kant’s and Rousseau’s depiction of the hazardous aspects of social progress is the fact that all these elements in one way or another have to do with our social passions. Passions are strong desires; according to Kant they are often strong enough to make man lose control over himself and thus they may develop into vices. Vices of nature consist in the misuse of our natural inclinations. The vices of culture follow basically the same pattern since they arise from the excessive desire for ends that automatically arose in the social state. They are the result of our predisposition to humanity, which is not merely physical but also involves reason.

Reason is a double-edged sword that must be handled with care. It elevates us above the rest of the creation and lets us see things in a new manner. The natural man did not care much about other men. If he met another human he would hardly react differently than if he had met an animal. However, when savage man left his woods and entered a life among other people, he suddenly saw himself as standing in relation to them and he immediately began to evaluate his own merits by making comparisons. This is the “origin of inequality” looked for by Rousseau in the Discourse. More exactly, it is the origin of our awareness of natural inequality.

Therefore, one of Rousseau’s most central arguments was that one must not confuse natural and moral inequality. The former, which consists in the differences of physical attributes such as strength, health length and age, obviously existed also in the natural state. The latter, which is also called political, is based on convention and on the non-natural advantages that men could obtain by means of inequality in natural skills. As already mentioned, since men now assembled within small areas, it became possible to study one’s fellows in a way one had not been able to do before. One suddenly became aware that some were stronger, some more skilful and some more attractive than others. It is not difficult to understand that such insights must have been shocking for those who had never reflected on such relations before, and it is thus easy to imagine how feelings like vanity and contempt could arise in those who were superior, and how shame and envy were born in those who found themselves to be inferior. (2D II 47/170) Kant and Rousseau unanimously state that comparison, that is, relative evaluation is the root of all evil since it brought forth a new mode of existence: existence as an object in the opinion of others. The man who had previously never doubted his own feelings or his own worth could now only evaluate himself in comparison with others. Furthermore, man, who until now had been governed only by true needs – which never extended his powers – now found a new unlimited faculty: the power of imagination.

The natural question here is where these foreign passions of competitiveness actually came from. If it is true, as Rousseau declares, that all sentiments in the human soul are born out of amour de soi and pity, also these potentially violent passions must have their origin in them. Yet, Rousseau’s
answer is that all social drives have their root in a transformed kind of self-love, namely in what he calls *amour propre*.

*Amour propre* is probably one of the most controversial and widely debated concepts in Rousseau’s moral psychology. The difference between natural love of self (i.e. *amour de soi*) and this novel feeling which often manifests itself in vain pride and the desire to dominate (*amour propre*) is highlighted in a footnote to the Second Discourse in which Rousseau says:

> *Amour de soi-même* is a natural sentiment which inclines every animal to watch over its own preservation, and which, directed in man by reason and modified by pity, produces humanity and virtue. *Amour propre* is only a relative sentiment, artificial and born in society, which inclines every individual to have a greater esteem for himself than for anyone else, inspires in men all the harm they do to one another, and is the true source of honour. This being well understood, I say, that in our primitive state, in the genuine state of nature *amour propre* did not exist; for each particular man regarded himself as the sole spectator to observe him, as the sole being in the universe to take any interest in him and the sole judge of his own merit. (2D note 91/219)

*Amour propre* is unnatural in the sense that no one is born with it. It is undoubtedly true that Rousseau cautions against the excessive effects of *amour propre* and that he occasionally calls it the “wickedness of man” or “the principle of all wickedness”. (E I 67/289; RJJJ II 100/790) This is also the most commonly referred to passage and this note is quite often thought to express the definition of *amour propre*. It is, however, important to be aware that the definition given in the Second Discourse is insufficient in order to explain the nature and psychology of social man. This explanation, though seemingly important, clear and concise, does not tell us much about the role of *amour propre* in our social life. One must not forget that Rousseau’s aim in the Discourse is to highlight the unnaturalness of *amour propre* from the “state of nature point of view”. In order to recognize the full meaning of the term, it is important to be aware that *amour propre* is not a static concept and, as we will see, it must be regarded as highly natural in our present state. In social contexts we cannot avoid looking at others and being affected by their actions, merits or defects. As a result of these impressions we increase, or reduce, our worth in relation to them and we, thereby, lose our original sense of our absolute worth. The only ones who are entirely free from this feeling are small children. However, at a certain age the idea of oneself as related to a greater whole suddenly becomes inescapable, and from that moment the child must realize that he is not the only one who has needs and wants. Moreover – and this is the most crucial effect – he realises that he has to compete in order to reach his desired position. Thus *amour propre* begins to develop and with it all the passions that depend on this principle. Appropriate education will decide whether the passions that arise from *amour propre* will be humane and gentle or cruel and malignant; that is, if they will be
passions of beneficence and commiseration or of envy and covetousness [passions de bienveillance et de commisération, ou d’envie et de convoitise]. (E IV 235/523) In other words, _amour propre_ is a passionate drive that necessarily develops in social contexts but one must not believe that _amour propre_ necessarily must take the form of an aggressive feeling or that it always degenerates into vanity and conceitedness. Even if it is tempting to focus on the seemingly sharp difference between _amour de soi_ and _amour propre_ one must not overstate the dangers of the latter, and it will be clear that without it, we would be fairly uninteresting beings.

There can be no doubt that it was society combined with men’s novel appetite for deceptive goods that brought the idea of _status_ into the human soul. Gauthier hints that he takes Rousseau to be saying that the art and sciences caused this corruption by bringing “complexity” and “luxury” into our minds and with this dissoluteness. Still, as Delaney makes clear, Rousseau did not see the arts and the sciences as bad _in and by themselves_. The problem was that they appeared in a social condition in which _amour propre_ already prevailed. The effect of a society influenced by arts and science was that man sacrificed his integrity for approbation. So, instead of doing what they themselves thought would be good, useful or right, the artists and scientists did what they believed other people would like. In other words, they put public esteem before their sense of morality.

Also Kant recognized the problem of social antagonism and, although he did not adopt the term _amour propre_, it is likely that Rousseau played a great role in the genesis of his idea of the antagonism he found in the social order. According to Kant, man has a natural tendency to enter into society, primary since this condition alone makes us able to develop our human characteristics. As we have seen, we are destined by our reason to live among other men, and by means of science and arts we must _cultivate, civilize and moralize_ ourselves. We must be _educated to be good_ for the animal drive will always remain inside us. In order to become worthy of being happy, these obstacles have to be fought. Thus sociality is the natural step for mankind, but, as Kant himself admits, it cannot be denied that human beings also have a strong tendency toward isolation. This antagonism – the conflict between companionship and individuality or between intolerance and dependency – expresses man’s _unsociable sociability_.

In spite of its importance for Kant’s anthropology, the discussion of this tendency is restricted to a few lines in the fourth proposition from the _Idea of a Universal History_ where Kant describes it as the propensity of human beings “to enter into society, which, however, is combined with a thoroughgo-

---

72 Gauthier p. 3.
73 Delaney p. 77.
74 This is the thoroughgoing theme of the _First Discourse_.
ing resistance that constantly threatens to break up this society.” These conflicting feelings awaken the powers that had so far been merely latent in the mind of man; through this, man’s indolent laziness was replaced by ambition [Ehrsucht]. Although this passion, together with his greedy and violent inclination to wealth and power [Habsucht and Herrschsucht] put man in opposition to his fellows, it was only by means of these egoistic drives that man was able to develop the skills that were prescribed by his vocation. As Kant put it, this antagonism was not only the “first true step from crudity toward culture” but also the sole way by which a human being makes himself worthy of his humanity by struggling actively with the obstacles that constantly tear him back to a cruder state. Since the human being was merely passive and subject to instinct in this cruder state, we should be grateful for our asocial traits for, as Kant says, without the influence of envy, emulation and vanity, all our excellent inner capacities would forever slumber undeveloped. (IAG 8:21; Anthr. 7:324)

Although social passions are unavoidable and thus entirely human, they may easily develop into vices and if these vices reach a high degree, they can be considered extremely evil and even be called devilish vices. (Rel. 6:27) Accordingly, in spite of the usefulness that this social progress had for mankind as a whole, human beings had to suffer much from what followed from it. Here, Kant is obviously drawing much from Rousseau, since he adopts the idea of a sudden restlessness that annoyed the human heart when man was thrown out from his harmonious state of individuality and found himself forced to estimate his own person in relation to others.77 Now the frustrating

76 IAG 8:20. Although the very notion “unsociable sociability” is generally associated with the “fourth proposition” in the IAG from 1784, the essence of this idea is explained about ten years earlier in notes from Kant’s lectures on anthropology (1772-3). Friedländer writes about the inclination to be in a society and to form families as well as larger communities, but he adds that there will still always be an inclination that aims at oneself, namely the desire for domination, honor and other “vanities”. These inclinations are indeed “social” (they exist and can only be satisfied in a society) and they involve the desire of having a big circle of admirers. Yet, since these inclinations aim to keep others at a certain distance from oneself, they are at the same time to be considered unsocial. (VA Friedländer 25:585-6) They are based on distrust and jealousy, which correspond to two related (though reverse) attitudes and while the former leads one to defend what one already has the latter makes one strive to obtain things that one does not yet have. These two passions are in a sense related to the fear of being devoid of something important and especially of something that others have. There is also a short note about this tendency in a reflection dated somewhere between 1785-89 (Ref. no. 989, 15:433) For good discussions of this aspect of humanity, see Wood (2009), Schneewind (2010) and Ferguson.

77 This pattern bears also several similarities to Mandeville’s idea of the important role for “unsociability” and egoism in human evolution; an idea that was expressed in his highly popular book The Fable of the Bees, a work originally consisting of a single poem but which gradually, through countless editions with new additions, grew to a two-volume edition of comments and miscellaneous reflections. The poem, named “The grumbling hive, or knaves turn’d honest” illustrated the counter-productivity of unselfishness in society since, without egoistic striving, the previously wealthy and flourishing society would simply collapse. This conclusion was expressed in the subtitle of the Fable, namely that “private vices” is the condi-
paradox of human nature was an undeniable fact. Driven by social passions like ambition, greed and addiction to power, man had to realize that he so desperately needed the same beings that he at the same time struggled to tolerate. (LAG 8:20-1)

This tendency overlaps almost perfectly with our rational self-love, which constitutes our predisposition to humanity. As reason awoke we became able to compare ourselves to others, and although we initially only struggled to reach an equal position, this striving for equality soon mutated into striving for superiority. Because of this, our passions become “vices of culture” and the sources of all injustice in human society. If successful such vices tend to produce wealth, vanity and pride; if unsuccessful they result in hate, vengefulness, shame and envy. (Rel. 6:27)

Before leaving this section there are some important remarks to be made. Because of the hazardous nature of social interaction and the comparison that is essentially connected to it, one would perhaps draw the conclusion that the most effective way to secure good behaviour is to never compare ourselves to others. Yet, as will become clear, comparison and awareness of one’s position among others is the ground of everything that can be called good in any interesting sense. Even Rousseau admits that everything would be worse without cultivation and our species was not made to remain only half-finished. (E I 37/245) We therefore have to learn to compare ourselves in the right manner; that is, we must know when – and to what extent – this interpersonal comparison is to be practised. I will return to the importance of comparison and the necessary conditions for a fruitful practice of it in chapter 7.

1.3. Moral Beings

Whether or not it ever reigned, the state of nature does no longer exist and we cannot use the idea of it as an end for human progress. We must accept that this state, which merely represented a hypothetical state of ignorance and innocence, is lost forever and that we must meet the demands of this new state: a state in which we have to exist among others. For Rousseau, society begun with the mutual love between two individuals, and from this bond the first families were formed. Families gathered in the same area, and in this way, small communities arose. This step was in many respects to be considered an advantage since it introduced us to a state of reciprocal love and friendship which made it easier to secure our true needs. Yet, as we have

tion of “public benefits”. Although the Fable is not featured in the list of books present in Kant’s library (Warda), Kant probably read it, as did almost everyone at the time. His references to Mandeville are, however, only sporadic and fragmentary. For a discussion of Kant’s possible indebtedness to Mandeville, see Hundert p. 55.
seen, it had also become a state of dependency and antagonism. Through the social, the non-rational savage transformed into a man who used his reason, a man who reflected and deliberated but, unfortunately, often did so in an immoral way. The human being was still basically governed by self-interest but, since the desires born by imagination were able to transcend natural need, and since the new man now often saw his fellows – not as companions but as rivals and foes – the Hobbesian state of war had become a reality. Nonetheless, all human beings, including those who are born in the social state, are still good in their original nature; consequently, any new-born infant is still as innocent and uncorrupted as the savage man once was. Corruptness and vices are thus the result of the ideas and ideals planted from outside. They can be seen as seeds of weeds that tend to take over the child’s soul which is continuously susceptible to any kind of influence. In another passage of Emile, Rousseau explains with his usual expressive vocabulary the way one must proceed in order to avoid this danger:

Do you wish, then, to excite and nourish in the heart of a young man the first movements of nascent sensibility and turn his character toward beneficence and goodness? Do not put the seeds of pride, vanity and envy in him by the deceptive image of the happiness of men. Do not expose his eyes at the outset to the pomp of Courts, the splendour of palaces, or the appeal of the theatre. Do not take him to the circle of the great, to brilliant assemblies. Show him the exterior of high society only after having put him in a condition to evaluate it in itself. To show him the world before he knows men is not to form him, it is to corrupt him; it is not to instruct him, it is to deceive him. (E IV 221-1/505)

This leads us to the fundamental principle of Rousseau’s theory of education; namely, the first instruction must be purely negative. Such education does not consist in teaching the child virtue or truth, but rather in guiding it away from vice and deceptive ideals. In Rousseau’s view, very few of the things and ideas that the unprotected mind of the child will be exposed to in the world are related to humanity as such; on the contrary, they ought to be seen as abominations of nature. Even if it might be difficult to fully accept Rousseau’s frank claim that the only thing essential to man is his mortality, we must admit that his general account is quite reasonable. After all, most of us will never become as rich, famous and heroic as the men we read about in books or find portrayed on the scene in the theatre. (E IV 221-2/503-4) The ordinary man is weak and vulnerable. What he needs is not visions of a fantasy world but the tools necessary for proper self-knowledge. Again we are reminded of Rousseau’s motto – to know oneself. This is why Emile’s tutor could only recommend one single book for the young – namely Robinson Crusoe. (E III 184-5/454-5)

It is true, Rousseau admits, that the choice of this book involves an apparent contradiction since on his isolated island, Robinson cannot be seen as a
social man and his condition is thus far from that of Emile. On the other
hand, the social man should look upon the world in a similar way to this
solitary man. Robinson had to find out what he truly needed and, in so doing,
he learned to live a decent life with the means available to him. Most im-
portant, however, was the fact that there was no one there to judge him. For
the child, the early years represent the age of nature, which pre-empts the
age of intelligence.78 As was mentioned in the section about the natural man,
Rousseau held that the true movements of nature could never lead us astray.
Before the young man can handle his reason and what comes along with it
(e.g. amour propre), it is important that he does not act because someone is
observing him. Our intuition would lead us right and as long as the human
being follows this lead he will do nothing but good. (E II 92-3/322, FM
38/97)

At the same time it must be made clear that while Rousseau indeed hon-
oured the natural man for his modesty and simplicity, he underlines that this
man – which basically is an animal – is far from good or noble in ordinary
terms. On the contrary, the natural man has absolutely no sense of morality
since he does not reflect and he does not reason. His goodness – which can
by no means be called virtue or moral goodness – is only an effect of his
ignorance of property and all hierarchies created by social relations. In that
state, Rousseau says, man knows nothing but himself; “he neither hates nor
loves anything. Restricted to physical instinct alone he is null, he is stupid
[bête]. That is what I have shown in the Discourse on Inequality”.79

These comments show that Rousseau did not think that a merely natural
goodness and friendly living could qualify for being a truly moral being.
Accordingly, the goal of education cannot be the making of a savage but the
making of a man. The man, however, is not the same as the citizen and again
we may recall Rousseau’s famous dilemma that when educating a young
person one must “choose between making a man and a citizen for one cannot
make both at the same time”. (E I 39/248) Furthermore, the “man” in ques-
tion is not only distinct from the savage and the ordinary citizen but also
from the man in the pre-governmental social state. Still, Rousseau insists that
the ideal man is also the most natural man:

I will be told that I abandon nature. I do not believe that at all. It chooses its
instruments and regulates them according to need, not to opinion. Now, true
needs themselves change according to the situations of men. There is a great
difference between the natural man living in the state of nature and the

---

78 These ages, along with the last, namely the “age of force”, are described in the draft to
Emile called the Favre Manuscript. In Emile, the “age of intelligence” is generally called “the
age of reason”.
79 LB 28/936. Cf. FM 8/61 where the state of nature is called “a state of ignorance and stupidi-
ty”. See also the Social Contract where Rousseau says that instead of being a stupid animal
[animal stupide] man could now, because of his acceptance of social norms, become a ration-
al being and human. (SC I:viii 141/364)
Obviously, in forming “the man of nature”, Rousseau is not arguing for a return to the woods. It is enough to make Emile strong enough to not being carried away by the “social whirlpool” which has deceived so many of his fellows. The ideal man should be governed by nothing but his own reason and the feelings coming straight from his heart. Again, social man must be cultivated; one must twist and turn in order to reconcile natural and social law. As Rousseau, quite paradoxically, puts it: “one must use a great deal of art to prevent social man from becoming totally artificial.”(E IV 317/640)

Accordingly, in order to understand Rousseau’s paragon of the perfect man we need, as the above-quoted passage shows, to add another conception: “the natural man living in the state of society”. This, which with Cooper’s vocabulary is called “civilized naturalness”, is indeed a “secondary naturalness” but only secondary to the savage naturalness. Civilized naturalness is the highest stage of human naturalness that one can reach, and it is the closest approximation to the state of nature that is possible in our present state. The man who possesses this second nature does not live like his neighbours but rather with them. This man will be good since he will still be able to understand the call from within. The nature of this inner voice will be spelled out in chapter 5. We will see that this voice, though inner and unaffected by external influences, is still a voice of true virtue and not merely a fleeting romantic sentiment. In other words, it is a full-fledged moral guide. Again we may recall Rousseau’s claim about the “poorly compensated” enlightened man who indeed had got such things as enlightenment, knowledge, laws, morality, reason, propriety, consideration, gentleness, amenity, politeness, education. Yet, what he truly needed and what alone could sum up genuine human goodness was two small words that were ringing in Rousseau’s (as well as in any uncorrupted man’s) ears and which made him all the more confident about their magnificence. These words were truth and virtue, and if someone saw these terms as mere words, Rousseau confessed that then he had nothing more to say to him.

In insisting that it is not positive education that makes men morally good, Rousseau differs significantly from Kant who held that the human being must be actively educated to the good. Negative education, which Kant refers to as discipline, will only prevent our purely animal instincts from dom-

---

80 See Cooper p. 40.
81 See Rousseau’s “Letter to M Abbé Raynal” CW II 27/OC III 33. Cf. also the Dialogue: “Someone who can coldly contemplate virtue in all its beauty, who can portray its most touching charms without being moved by them, without feeling stuck by any love of virtue, such a being, if he can exist, is hopelessly wicked; he is a moral cadaver.” (RIJJ I 8/667)
inating our soul, but talents as well as manner and morals need much more than that. Accordingly, in order to become morally good it is not enough to let our innate good germs or predispositions develop since we also have opposing forces that constantly tend to overwhelm our thinking. (Rel 6:57) Again, since providence has not provided us with already finished skills, the human being needs – and ought to – develop them and, as Kant says, “make himself worthy of humanity by actively struggling with the obstacles that cling to him because of his nature.” (Anthr. 7:324) Indeed, Kant was clear about this already in the 1760s since he wrote:

In the human being we must distinguish between the animal, i.e., what happens in him in accordance with laws of sensibility, and the spirit, in accordance with laws of reason. His power of choice as an animal is really always determined by stimuli; yet his will is still free insofar as his reason is capable of altering these determinations of the power of choice. (Refl. 3872, 17:320)

Although it will become clear that the struggle against obstacles and conflicting forces is not essentially a battle against the inclinations themselves, we must not ignore the impact of these incentives on our will. Before man is disciplined he is wild, lack of cultivation makes him rough. Since cultivation only applies to skills and talents, it has no influence on our capacity of setting ends, namely our will. A further kind of culture, civilization [Zivilisierung] is needed in order to refine our sentiments and temperament, that is, those things that Kant in his definition of practical or moral anthropology identified as the “subjective conditions” in human nature which either hindered or helped people in performing moral duties. (MS 6:217) Yet, moral anthropology is not ethics and not even the most careful improvement of these subjective conditions can provide us with the means for actually fulfilling the supreme end of human existence. This is the task for our rational will, and only moralization affects this true source of morality in the human mind. By means of this very special kind of culture we can develop a character and form our way of thinking so that we, in all possible circumstances, will become disposed to choose the right kind of maxims. This is why Kant had to add his third predisposition – the predisposition to personality – which enables us not only to act from reasons but also from the right reasons, thereby choosing nothing but good ends. This predisposition is the susceptibility for respect for the moral law, and the capacity to make this a sufficient incentive. Such an incentive is not based upon self-love, but rather on the moral feeling itself. In the second Critique, Kant exclaims:

Duty! Sublime and mighty name that embraces nothing charming or insinuating but requires submission, and yet does not seek to move the will by threatening anything that would arouse natural aversion or terror in the mind but only holds forth a law that of itself finds entry into the mind and yet gains reluctant reverence (though not always obedience), a law before which all in-
clinations are dumb, even though they secretly work against it; what origin is there worthy of you, and where is to be found the root of your noble descent which proudly rejects all kinship with the inclinations, descent from which is the indispensable condition of that worth which human being alone can give themselves? (KpV 5:86)

Obviously, duty and the principle of personality is what elevate the human being above his mere existence as homo phaenomenon. It is this idea that links man to the higher order of things, an order that is only accessible for rational beings. As a result of this receptiveness, man is free and thus independent of the mechanism of nature; however, this involves him being subject to pure practical laws that are imposed on him by his own reason.

The intensity of this awareness of our duty is spelled out in Kant’s so oft-quoted lines about the starry heaven and the moral law and how they filled him with a never-ceasing feeling of admiration and reverence. (KrV 5:161) Neither of them needs to be searched for since they both come to us as if they were immediately connected with our existence. Now, in contrast to the sky, which is easily accessible to every eye, the moral law can only be reached from within by means of our understanding. Thus a remaining task for Kant was to explain how and why the moral law is necessarily present within us. Kant’s final explanation of the existence of a moral law within the human mind has, quite understandably, been questioned since it appeals to a pure and undeniable fact [Factum] of reason that cannot be further investigated or proved. As Kant says, the moral law cannot be found in any antecedent data – not even in freedom – and, as we have seen, we do not actively need to consult it since its law forces itself upon us as a synthetical a priori proposition. Because of this assumption, Kant may then conclude that: “Pure reason is practical of itself alone and gives (to the human being) a universal law which we call the moral law” (KpV 5:31)

However, although the idea of the moral law as an innate fact of (practical) reason indicates that it is not directly derived from freedom, it is still completely dependent on the existence of it since the very idea that there are duties requires that we must be able to fulfill them. This reasoning also leads us to Kant’s justification of freedom, or rather, his reasons for holding freedom to be true. Since the law is already within us and provides us with imperatives, we also immediately realize that freedom must exist so that we can act in accordance with these demands. Thus the relation between freedom and morality is that while freedom is the necessary condition, the ratio essendi for the existence of morality, morality is the ratio cognoscendi of freedom. (KpV 5:5n.)

Apparently, in the reasoning above Kant is now speaking about a specific form of our rational capacity, our practical reason, which alone can account for the concepts that he needed in order to lay the foundations for his moral theory. Put very briefly (this will be brought up again in the next chapter),
reason has two quite distinct faculties. In the second Critique we are told that while theoretical, or speculative, reason deals solely with the object of the cognitive faculty, practical reason is concerned with the determining grounds of our will, that is, with the production of those objects that correspond to our representations. Later in the same work we learn that the practical use of reason consists in the determination of the will “with respect to the final and complete end”. (KpV 5:120) Without this aspect of reason, we would not be able to make any judgments at all about those things that cannot be cognized through experience, including freedom. Kant was well aware that freedom cannot be proved absolutely, which is to say that freedom cannot be an object of cognition. Such a direct cognition is, however, not essential and in the preface to the second edition of the Critique of Pure Reason (1787), he states that he nevertheless can “think freedom to himself”. This move, he says, is justified by the fact that the representation of freedom at least contains no contradiction in itself, and he adds that for morality he needs “nothing more than that freedom should not contradict itself, that it should at least be thinkable that it should place no hindrance in the way of the mechanism of nature in the same action (taken in another relation).” (KrV B xxviii-xxix) The application of this idea of a mere possibility to think of oneself as free is spelled out in part III of the Groundwork where Kant says that in so doing we transfer ourselves into the world of understanding and regard ourselves as members of it. In this way we become subject to no other laws than those of morality. (GMS 4:452) The unconditional moral command, which also is the maxim of the absolute good will, is the categorical imperative.

In the derivation of this principle Kant begins by introducing the idea of an absolutely good will, linking it to the notion of duty and thereby creating the ultimate test of every possible action: Whenever I am to act, I must ask myself if I can will that the maxim – the fundamental principle of my possible deed – can be made a universal law for all rational beings. Expressed in another way, this principle also requires that I respect humanity [Menschheit] in all human beings by never treating them merely as means but always at the same time as ends in themselves. If exercised correctly, this method will necessarily rule out any claims for private happiness or principles that conflict with the objective ends of nature. Moreover, tested against a number of moral queries, the universality claim will show that one’s maxims end up in a contradiction; either a conceptual contradiction, as for the lie or for the maxim of suicide or a contradiction in my willing. The last contradiction is not practically impossible, but nevertheless impossible to will it as a universal law. In spite of the abstract character of these rules, Kant maintains that these insights are accessible to “common human reason” and even the great-

82 KpV 5:15. In the introduction to the first Critique (2:nd ed.) it is said that while speculative reason determines the object and the concept of it, the task of practical reason is to make the object real. (KrV B ix-x)
The very idea of an intelligible world and our membership of it is indeed mind-bending and requires a number of supporting premises. It is clear that in such a world, anthropology has no place since all “subjective conditions” have been removed. This world constitutes the realization of the final end, that is, “mankind under moral laws”. Yet, we are not pure noumena so hindrance to morality will always remain. So who is Kant’s moral being? Is he, in fact, foolishly performing the work of Sisyphus; the tragedy that in the long run, turns into a farce? Fortunately, there are other ways of understanding the demand of perfection.
1.4. Ideal Beings

The three portraits above tried to explain man as he is and how he ought to be. In some sense they also told us how man may have been and what he eventually will become if he follows the inner call to develop his inherent possibilities completely. In the case of Rousseau, the making of a moral man will undeniably require much work, but this task does, in fact, appear to be conceptually possible though it may be impossible given our present condition. As already said, we cannot – and ought not – return to the wood or to the savage state of nature. Rousseau’s ideal is, as we saw, the natural man living in society, the one who has no inflamed *amour propre* and who still hears the voice of his true nature. With Cooper’s words, this civilized naturalness expresses “the highest level of humanity.”85 This man lives within himself rather than on the outside and in the opinion of others. As Allan Bloom puts it, the crucial task for the educator is to make Emile (or indeed any human being) “social while remaining whole”.86

With respect to Kant’s approach to ethics one must agree with Samuel Kerstein that it is “no modest claim”.87 Kant wanted to find the supreme principle of morality which could serve as the sole compass for all ethical queries. As we saw, Kant’s conclusion is that this principle, if it exists, must be the *categorical imperative* and the sole incentive for such actions is pure respect for the moral law which requires the total abstraction from all subjective influences. This is not easy for a human being and Kant actually admits that it is not clear whether any single action from duty has ever been performed. (*GMS* 4:406-7)

Because of this, yet another type of man may perhaps be added to this overview of the human being. In Mrongovius’ notes on anthropology, we learn that another thing that differentiates man from other species is that all animals reach their individual vocation already in this life. On the other hand, human beings only make minor progresses from generation to generation since improvements that are made by one generation are added to the next. (*VA* 25:1417; cf. *Päd* 9:441) By this almost Lamarckian idea Kant seems to say that human perfection is only possible with regard to the species, and this kind of perfection is therefore possible only over a long time of constant evolution. To be sure, Kant admits that no human being will reach the state of moral perfection or the highest good in this life since this requires an *endless* progress and as we all know, human beings do not live forever.88 Because of this, the notion of man as he should be, or become,

---

85 Cooper p. 103.
86 Bloom (introduction to *Emile*) p. 20.
87 Kerstein p. 1.
88 While this is an extraordinary multifaceted topic one can at least say that for Kant, the highest good consists in happiness in exact proportion with morality or virtue. The attainment of the highest good thus signifies our *worthiness* of being happy [Würdigkeit glücklich zu
cannot properly be ascribed to the existing man. The notion of the truly moral man is, therefore, most of all an ideal. Indeed, Kant admits that the human being can never reach the moral realm. If he did, he would be holy since only a being completely free from subjective incentives (i.e. inclinations) has a will that always and necessarily accords with the moral law. Again, Kant’s description of men as made of “crooked wood” also indicates that we are – and will remain – imperfect creatures. (Rel. 6:100; IAG 8:23)

Still, Kant equally often declares that it is our “universal human duty” to elevate ourselves to the ideal of moral perfection; indeed, as he puts it, “to the prototype of a moral disposition in its entire purity”. (Rel. 6:61; cf. KpV 5:84) He frequently refers to our duty to promote the highest good since it is the necessary object of the moral will, that is, a will determined by respect for the law. (KpV 5:4) In the Religion we learn that the ultimate aim of our battle with the evil principle in our nature is the foundation of a kingdom of God on earth which can also be called the realm of virtue. (Rel. 6:93; 94)

Accordingly, the idea of perfection and the highest good, practically attainable or not in this world runs through Kant’s entire practical philosophy. This idea also forced him to show that pure reason can also be practical and that it could thereby extend itself to objects that pure speculative reason could never attain without finding itself trapped in antinomies. While this is a central theme in Kant’s critical thinking we need, for our present purpose, merely understand how these commands can be given to us as duties and imperatives. Now, if we recall the three questions that opened the second section of the Canon – “what can I know?”, “what should I do?” and “what may I hope?” – it will turn out that the first does not have much to do with our problem since it is purely theoretical (speculative) and thus restricted by the limits of possible experience. More important are the moral and theological ones since these involve reason in its practical use. It is only through this extension of reason’s capacity that we can assume freedom as well as the existence of God and the immortality of the human soul. We simply must do this since they are the necessary conditions for the possibility of the highest good that it is the final end of the moral law that, in turn, as we have seen, is given to us as a fact of reason. By postulating them for our practical purposes we do not claim to know anything about them and accordingly we cannot say that it is certain that they exist but only that we are certain that they do.

[sein]. Kant’s most detailed outline of the highest good is entailed in the dialectic of the KpV but he also discusses the topic in the Canon (KrV) as well as in the KU and in the Religion. The idea of the highest good as a necessary end of the moral law was the reason why Kant had to postulate God (only God could secure such a perfect accordance of happiness and morality) and the immortality of the soul since this end requires an endless striving towards perfection and thus we need to assume that there is a future world. These Ideas (freedom, God and the immortality of the soul) are therefore postulated for practical purposes.
1.5. Summary: *Human* Beings

In order to sum up the essential lessons from this chapter I will once again stress that it should be clear that the above outlined portraits of the human being are nothing but stereotypes and none of them actually expresses the *present* man. That neither Kant, nor Rousseau, could regard any living being as an *ideal being* should follow from their respective conceptions of a *human* being since these conceptions entail numerous imperfections that are hindrances to the completeness of the human vocation. Accordingly, the ideal being is no *human* being. The good being “under human conditions” is one that is subject to laws, both moral and socio-juridical ones. At the same time it is also doubtful if the merely *natural* being was a human being and even Rousseau, the defender of nature’s innocence and glory, admits that it was only when the voice of duty replaced physical drives and when right replaced appetite that we become truly human. (*SC I:viii 141/164-5*) So, in some sense, Kant was right in saying that the human being needs a master. Yet, this master must be found *inside*. In other words, man must learn to become *his own master*.

This insight can perhaps be said to be the fundamental idea that reappears in both Kant’s and Rousseau’s anthropological inquiries. For Rousseau, it was because of the acquisition of *moral freedom* that man became the master of himself rather than remaining a slave to mere passion. (*SC I:viii 142/365*) Kant saw the moral predisposition as the one that made him prepared to create his own way of thinking. (*Anthr. 7:285*)

Yet, the mere possibility to be one’s own master is not sufficient for always *expressing* this additional predisposition. We need to learn how to cope with conflicting incentives and temptations to transgress duties. We must understand how to use our powers correctly in order to develop them completely and purposively in order to avoid the state of “purposelessly playing nature” in which “desolate chance takes the place of the guideline of reason”. (*IAG 8:18*)

As we will see, the present man needs much and there are several obstacles he has to handle. We need to learn what to do in order to not diverge too far from the path to true humanity by grasping the tension between nature (*authenticity*) and the alienated state – a tension that can be translated into terms of *inside* respective *outside*. We will have to learn much more about this unique but also complicated being.
As we saw in the previous chapter, one of the most central divisions with regards to the human being is between what belongs to us as merely living beings and what we need to elicit in ourselves in order to also be regarded as moral. We learned that we have certain powers of thinking and reasoning but we have not yet analysed the faculties in detail. For practical purposes, some of the most important of these abilities are those that concern our motivational powers, which in human beings are (and must be) free in order to allow for moral responsibility. These powers are the will and the faculty of choice. However, the double nature of man – partly rational and partly sensual – implies that our volition is susceptible to feelings; consequently, we do not necessarily act according to the moral principle as we ought to. That is why Kant says that freedom can be regarded as "the greatest good and the greatest evil". (Refl. no. 7217 19:288)

This chapter will sketch the basic features of the faculties just mentioned, that is, the human will and its ability to choose between possible actions and, when a decision is reached, put this decision into practice. It will become clear that we must distinguish between various kinds of motivating forces and uncover the sense in which everything we do is the result of our freedom. We also need to understand why Kant found himself obliged to say that although we cannot prove or know that we are free, we can act only under the idea of freedom.  

Since Rousseau did not offer a scholarly scheme of freedom or its existence in the world, the main focus in this chapter will be on Kant. Surely, this does not mean that Rousseau was unaware of the possible problems of freedom and the human will; as we have seen, he is sometimes said to be the one who laid the foundations for Kant’s theory of autonomy. Still, as the historical anthropologist he was, he mainly wanted to discuss the man he saw. For that reason, Rousseau did not discuss human nature and its capacities in metaphysical terms. Actually, he sometimes explicitly claimed that he had “no wish at all to speak here about this metaphysical and moral freedom” and that he cared very little about knowing whether “the acts of my will are in

---

89 This central claim appeared already in a Reflection from the 1770s (see Refl no. 1021, 15: 457) but reappears as a central argument of Groundwork III.
my own power or if they follow an outside impetus”.  

This comment, written shortly after the First Discourse is, however, entirely different from many of his further comments on freedom where he seems to be fully convinced that we are our own masters. In the Reveries, for example, he says that “there is hardly any of our automatic impulses whose cause we could not find in our heart, if we only knew how to look for it”. (RSW VI 49/1050) In the Second Discourse we are told that while animals obey the commands of nature, man, though feeling the same impetus, still realizes that he is free to acquiesce or resist. It is in this “consciousness of his freedom that the spirituality of his soul is shown.” (2D I 26/142) Likewise, the importance that Rousseau assigned to the general will obviously indicates that he realized the need for a more sophisticated way of thinking. According to Rousseau, a citizen can, and often has to, restrict his own will, or personal demands, in favor of some higher kind of intelligent insight or for the common good. Still, since he never explicitly discussed the very process of motivation and choices between various motives in the structured way that Kant did, we do not know which arguments he might have had with respect to this question. 

That said, we may have to have Rousseau’s thoughts in mind when we examine Kant’s more advanced idea of freedom and especially the notion of autonomy since it is commonly said to be derived from Rousseau’s social ethics or, more precisely, from the kind of freedom that men developed in society. With these caveats in place, I believe that we may return to Kant.

2.1. The Free Being

Kant operates with a number of different conceptions of freedom and it is not always easy to orient oneself in this complex discussion. Not only are the various notions often complicated, Kant is not always entirely consistent in his use of them. Regardless of the many obscurities with respect to this notion, it is clear that freedom plays a central role in moral agency. If, for example, everything we did was the result of the laws of natural causality, our activities would have been like those of the marionettes that moved solely as a result of external forces. This illustration of natural causality is, however, not complete and according to Kant, it has deceived many thinkers into believing that the bare fact that one can explain human actions as the result of internal causes would be enough to save the idea that our actions are essentially free. To regard such a process, a process which explains our actions as

---

90 Rousseau, “Fragment on Freedom” (1751) in Collected Writings of Rousseau vol. IV pp. 12-13. For several perspectives on Rousseau’s idea of freedom in the human society, see e.g. MacLean and McDonald & Hoffmann.

91 Again, this is not meant to diminish Rousseau’s rank among philosophers. He did have things to say about human motivation and moral capacities. I will return to Rousseau’s moral psychology in chapter 5.
caused by our own inner incentives – *freedom* would rob the term of its true meaning since we thereby ignore the fact that desires and wants also have their origin in a causal chain. The simple appeal to the “internalness” of these factors would, in Kant’s view, require that we in the same manner would have to regard the movement of a clock, or a turnspit, as *free* simply because such artifacts also perform their moves “*by themselves*” without any external forces affecting them. In the case of the automatons, we all understand that these “*inner principles*” are ultimately determined by an external one, namely by the designer, or watchmaker, who organized the cogs and springs of the machinery. In other words, this is no absolute freedom since freedom in its proper sense, must be *absolute* spontaneity without qualification. (*VMet* [L1] 28:267-8) At best, one could call this phenomenon *comparative* freedom but in this case, freedom can only be ascribed to the *effect* [Wirkung] as in the case of a body in free fall. This conception is far from sufficient to explain moral responsibility, which also requires that the efficient cause must be proven to be free.

Now, it should be clear that Kant admits that the very concept of causality brings with it that of *laws*. This means that not even the freest of will can be regarded as being entirely “lawless” (i.e. *not as mere chance*) but he emphasizes that this law must be of a “particular kind” and clearly distinct from natural laws for otherwise a free will would be a non-entity [unding]. (*GMS* 4:446-7, *Rel.* 6:51n.) Accordingly, we need to find a conceivable conception of *free causality according to laws* that can coexist with the ordinary form of causality that obviously exists in the world. Now, the first lines of *Groundwork* III plainly state that the causality that we look for is our will [Wille]. In the case of rational beings, the will is free in the sense that it can be efficient independently of alien causes. This stands in direct contrast to all non-rational creatures, which are determinable solely on external grounds. Unfortunately, Kant adds that this conception is unfruitful for his purposes in this section since, being only *negative*, it cannot explain the precise relationship of freedom and the moral law and hence it cannot explain how the moral law can be a motivational ground for rational beings. (*GMS* 4:446) Still, in spite of its insufficiency for the deduction of the moral law, this freedom plays a significant role for human acting. Indeed, this negative concept is said to be the ground for the derivation of the positive one that Kant elsewhere defines in terms of *transcendental freedom* and which in the *Groundwork* is given

---

92 This so called *psychological* concept of freedom is discussed in the second *Critique* (5:96-7) but the likeness of the mental process with that of a clock, a turnspit or other mechanical things appeared already in notes from his lectures on metaphysics dated to the mid-1770s. This kind of freedom is there called “automatic spontaneity” or *spontaneitas automatica*. (*VMet* [L1] 28:267). As is sometimes noted, this appeal to “inner grounds” appears to have been Kant’s own idea of freedom in the *Nova Delucidatio* from 1755. In this work, Kant’s spokesman argues for the existence of *internal determining grounds* and uses them as arguments for the existence of a free will.
the meaning of autonomy of the will. However, as was briefly mentioned in chapter 1, freedom is one of the ideas [Ideen] of which we can have no direct knowledge but whose existence we nevertheless have to assume for “practical purposes”. Still, since the possibility of freedom makes up the entire basis of our nature as moral beings, we need to take a closer look at Kant’s justification of this assumption. The reason for this is that it involves several elements that we will need for our upcoming analyses and it contains the basis for the possibility for attributing a double character to the human being. Before that, I will give a brief introduction to the two notions of freedom and how they might be related to each other and to human agency in general.

The Groundwork’s discussion of freedom may, because of its strict opposition between autonomy and heteronomy, perhaps lead readers to believe that only autonomous actions – actions following the laws made by the rational will – are truly free and that all further actions – as being heteronomous – are caused by a chain of events in the past. Since historical factors are no longer within our power, we would then be as determined as the mechanical artifact that only accomplishes the motions decreed by natural law. This division of action into only two classes (autonomous and heteronomous) would immediately remove all real responsibility from the human being over his unlawful (heteronomous) actions since it entails that the agent was not free at the time that he performed them. Since this conclusion clearly conflicts both with common intuitions and with Kant’s general arguments about moral evil, this could not possibly be what he meant. Still, it cannot be denied that he sometimes seems to be using the autonomy-heteronomy distinction in order to mark the huge gulf between the moral and the non-moral, a difference so obvious that it is said to be cognizable also to the commonest eye. It is especially disturbing that Kant also literally claims that while the human will is linked to the notion of autonomy (though not always used in this way), the human faculty of choice [Willkür] is said to be heteronomous and thus entirely unfit for grounding obligations since it is “opposed to the principle of obligation and to the morality of the will”. (KpV 5:33) Kant speaks here of the “heteronomy of choice”, namely the dependence upon the natural law of following some impulse or inclination.

Nevertheless, a deeper analysis of Kant’s moral writings shows that he obviously wants to make room for the possibility of evil character and thus

---

93 This is, at least, what Kant says in the GMS: that the positive notion of freedom flows from [fließt aus] the negative one. (GMS 4:446). However, in the first Critique, Kant says that practical freedom is grounded in transcendental freedom and accordingly, that “the abolition of transcendental freedom would also simultaneously eliminate all practical freedom.” (KrV A534/B562). Perhaps Kant means that practical freedom indicates that there is something beyond the world of sense and that even this negative notion requires that we are also susceptible to laws of another kind that, hence, also must be of another origin than the laws of nature.
of evil choices. Accordingly, autonomy cannot be the sole conception of freedom that is relevant for moral purposes. In fact, in spite of the apparent importance of the autonomy-heteronomy distinction, it should be noted that Kant does not use the term heteronomy in works other than the *Groundwork* and the second and third *Critiques* – not even in the *Religion*, the work which contains his most exhaustive analysis of the difference between good and evil maxims and the way these originate in the human faculty of choice.94 Again, in order to find out how even our sensually conditioned (i.e. heteronomous) actions are to be considered as the result of a voluntary choice we need to deepen the idea of freedom with regard to human actions.

Now, in the second Critique, where he so carefully maps out the contrast between what is moral and non-moral, Kant returns to the terminology used in the *Groundwork* where freedom was given both a positive and a merely negative sense. Both of them, however, are now said to belong to the moral side since the independence from all matter makes up the sole principle of morality. As we see, there are many difficult arguments and central dichotomies involved in Kant’s theory of human agency. In the following I will extend the discussion of Kant’s conception of human freedom and show in which sense we are free even when we do not act from the moral law.

2.1.1. Practical and Transcendental Freedom

The just mentioned “merely negative” conception of freedom is essentially what Kant elsewhere calls practical freedom, namely our independence of coercion by sensible stimuli or, in other words, independence from everything but the moral law.95 This is an effect of the human being’s faculty of choice as an *arbitrium liberum*, that is, a free faculty of choice. Although this faculty, as a human faculty, is not necessarily in accordance with the rational laws (it is also an *arbitrium sensitivum* and thus *impurum*) it is not for that reason necessitated to act from any present stimuli. Accordingly, the fact that this negative notion was said to be fruitless for the derivation of the moral law does not mean that it is insignificant for Kant’s theory of the human being in general. Again, this form of freedom signifies the most essential aspect of the human soul: the possibility to choose between various actions and that no incentive, however powerful, can – by itself – determine

---

94 The term heteronomy figures five times in the *KU* and though Kant does not here use it in moral contexts it is said to be unfree and based on sensibility or the judgments of others. In three of these passages, it is contrasted with autonomy. (*KU* 5:294, 5:350, 5:5389). Moreover, in a fragment from his *Opus Potumum* he says that a man [Mensch] is determined through heteronomy but that he as person [Person] stands under the laws of autonomy. *OP* 21:62.

95 See e.g., *KpV* 5:93-4, *KrV* A534/B562; A802/B830 and *MS* 6:213. This is also carefully spelled out in the *Lectures on Metaphysics* (see e.g., L1, Mrongovius, L2 and Vigilantius) I will return to Kant’s definition of a faculty of choice and its relation to the will section 2.2.
the human choice. Sensual incentives can undeniably affect this power but, unless the human being has freely chosen to include this incentive in a maxim, he will not be determined to act from it. This is what is expressed by the, so-called, incorporation thesis:

The freedom of the power of choice [Willkür] has the characteristic, entirely peculiar to it, that it cannot be determined to action through any incentive except so far as the human being has incorporated it into his maxim (has made it into a universal rule for himself according to which he wills to conduct himself); only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of freedom).

Henry Allison, who also is the inventor of the name of this central passage, argues that these lines include the key to almost everything important about Kant’s idea of moral agency. At first glance, this thesis seems to articulate the description of practical freedom, namely the independence of necessitation by impelling causes. This freedom, Kant says, can in fact be proved through experience since we can observe the possibility of restricting oneself and thus act contrary to a present temptation. Yet, what this proposition intends to say is not only that we now and then may make the decision to sacrifice a personal desire in favour of what is objectively commanded by reason. What we are told is that we always have to make a decision to act in a certain way by making a maxim. Hence, the liberum that is ascribed to our choice means that it is not the case that we can be overwhelmed by a desire and that our will thereby would be determined outside of our control but rather that we freely let ourselves be determined. According to this understanding of freedom also non-moral (including evil) actions are free even though they obviously cannot be considered autonomous since they are

---

96 This idea perfectly mirrors Rousseau’s view. Cf. the Second Discourse where he says: “Nature commands every animal and the beast obeys. Man feels the same impetus, but he realizes that he is free to acquiesce or resist; and it is above all in the consciousness of his freedom that the spirituality of his soul is shown.” 2D I 26/142.
99 KrV A802/B830. However, this is contrary to what is said in Mrongovius notes: “Freedom is not a property that we learn from experience; for we cannot experience anything negative. Indeed we do many actions by which we appear to act contrary to all stimuli, but we cannot yet infer if from that – for a secret stimulus can still have induced us, e.g., when despite torture and affliction a criminal does not confess then he acts indeed against all stimuli, but still in favour of a stimulus, namely the fear of death. Thus the inference that there is no stimulus because I notice none can accordingly be very uncertain” VMet [Mrongovius] 29:896-7. Indeed, this is very similar to what Kant stresses in the Groundwork, namely that we can (probably) never know the exact cause of an action since the maxims will always remain hidden.
based on private principle rather than on the pure rational will. Depending on our choice of incentive, the *maxim*, and hence also the *action* becomes “formal” or “material”, that is, moral or non-moral.

We will have to return to the notions of *formal*, respective *material*, principles for this distinction makes up a central feature of Kant’s system that also is one of the more problematic ones for those who defend a less rigid reading of Kant’s theory of motivation and moral worth. The crux of the matter in here is to understand the nature and function of the incentives, which, as we will see, is what Kant identifies as the *determining ground* [Bestimmungsgrund] of our will. In order to grasp the nature of this crucial ground we need a deeper understanding of one of the most central psychological faculties – the “faculty of desire” [Begehrungsvermögen]. Before that, we also need to consider Kant’s justification of *transcendental freedom* since it is a prerequisite for the possibility of practical freedom. (*KrV* A534/B562) In the second *Critique*, Kant continues this discussion and adds that without transcendental freedom – which he adds is freedom in its proper sense – no moral law is possible and no imputation in accordance with it.  

Now, although we now enter into areas that neither can, nor need to be examined completely here, it will nevertheless be useful to get a basic idea of what it means to be transcendently free. If we take another look at what is said with respect to the incorporation thesis we will see that practical freedom cannot give us the whole explanation of an action since this thesis does not explain the *origin* of the incentives that we need to adopt in order to create a maxim. While most incentives are elicited by something external, at least one incentive – the moral – has another origin. This need for assuming another kind of freedom is also hinted at by Kant’s use of “absolute spontaneity” at the end of the argument since absolute spontaneity is elsewhere 101

---

100 Admittedly, in his later work, Kant seems to say that freedom of choice cannot be *defined* as the “ability to make a choice for or against the law”. We can, indeed, see – through experience – that the human being is able to choose in opposition to, as well as in conformity with the law but, Kant adds, human freedom as the freedom of an *intelligible* being cannot be defined in this way, since “appearances cannot make any supersensible objects understandable”. Again Kant adds that freedom cannot be located in “a rational subject’s being able to choose in opposition to this (lawgiving) reason […] Only freedom in relation to the internal lawgiving of reason is really an *ability*; the possibility of deviating from it is an *inability*.” (*MS* 6:226) Still, the final sentence obviously focuses on internal law-giving which is at least not *directly* a task for the faculty of choice but rather of the will.

101 *KpV* 5:97. Again we need to recall that Kant’s discussion is not always easy to comprehend and his definitions from various works are not always entirely consistent. For example, in the Canon of the first *Critique* Kant limits his analysis to practical freedom since this is what is needed in order to deal with practical questions like what I *ought to do*. Still, as we saw above, in the second *Critique* Kant obviously believed that morality requires freedom in the *proper* sense of transcendental freedom. However, perhaps Kant’s exclusive focusing on practical freedom in the Canon is due to the fact that he believed that he had already justified the existence of transcendental freedom when solving the third antinomy in the Dialectic where he also, as we know, literally stated that transcendental freedom is a prerequisite for the practical form.
explicitly identified as *transcendental freedom*.\(^{102}\) Perhaps, it could be argued that what Kant says in the incorporation thesis is that our power of choice is free in the practical sense (independence), but that it must be free also in the transcendental sense, since practical freedom entails the freedom to incorporate the lawful principle originating in our will.

By transcendental or cosmological freedom, Kant understands “a faculty of beginning a state *from itself*, the causality of which does not in turn stand under another cause determining it in time in accordance with laws of nature.” (*KrV* A533/B562) Freedom in this sense is a *pure transcendental idea* that, by virtue of being the necessary and proper basis for the imputability of human actions, has nevertheless been a major stumbling-block for philosophy. (*KrV* A448/B476)

The difficulties are due to the very exceptional nature of this idea of an *unconditioned causality*. So, while practical freedom was cognizable through empirical methods this does not hold for the transcendental form. As we know from the analysis of human nature, it is our predisposition to personality that made it possible for us to make the moral law a sufficient incentive for our actions. This predisposition is, as we have seen, also the source of our idea of duty and thus of our intelligible existence. The predisposition to personality, Kant adds, is “freedom and independence from the mechanism of the whole nature” and it is also the capacity to act from *laws of a specific kind* – namely *pure practical laws* given by our own *reason*. (*KpV* 5:86-7)

However, although this may explain the ability to make the law our principle – which is secured through practical freedom – we also need to explain the existence and origin of the law itself; that is, how can we be free to create this ultimate starting point of a motivational process? As Kant had made clear already in the *Groundwork*, practical freedom is unfruitful since it cannot give us any insight into the possibility of the *freedom of an efficient cause*. (*GMS* 4:446, *KpV* 5:94) As was mentioned above, Kant meant that we must assume our will to be *efficient* independently of alien causes. (*GMS* 4:446) Still, in order to proceed from this “merely negative” notion to the positive one, we need to show the origin and possibility of *efficiency itself*. In other words, we must explain the source of the motivating factor by which our will actually – indeed after a free act of incorporation – is determined. Since this factor must be wholly distinct from what Kant calls “natural causality” it must somehow originate in our rational nature. Yet again, the mere fact that the cause springs from reason rather than sensibility does not automatically mean that it is free in the transcendental sense. In the *Canon* Kant says that transcendental freedom demands also an independence of “reason

\(^{102}\) Cf. *KrV* A474/B474, A448/B476, *KpV* 5:48, 5:99. The notion of absolute spontaneity [*spontaneitas absoluta*] figures frequently in the lecture on metaphysics notes, see e.g. *VMet* L1 28:267-9; *VM* [Vigilantius (K3)] 28:1022 as well as in *Reflections*, e.g., Refl. no. 5535 (18:211) and no. 4757 (17:703).
itself (as regards reason’s causality whereby it is able to begin a series of appearances) from all determining causes of the world of sense.” He adds that to this extent, transcendental freedom seems to be contrary to the law of nature, and hence to all possible experience. Since natural laws and causality in accordance with them obviously exists, this indeed seems to be a problem.

2.1.2. Justification of Freedom

Kant’s solution of the apparent contradiction outlined above was spelled out in the Dialectic in the first *Critique* where he presented his resolution of the third antinomy. Put simply, an antinomy is a conflict of reason with itself that arises when speculative reason leaves the realm of possible experience and tries to extend its use to the super-sensible. In entering such extravagant speculations, our reason will sometimes find itself exposed to seemingly incompatible and mutually exclusive conclusions yet forced to accept that both can be said to be true (or false). Since this conclusion is incompatible with our most basic logical axiom (law of non-contradiction) reason has thus produced a dilemma in need of resolution. As to the third antinomy, which deals with the concept of freedom and causality, Kant presents the two alternative positions as follows:

**Thesis:**

Causality in accordance with laws of nature is not the only one from which all the appearances of the world can be derived. It is also necessary to assume another causality through freedom in order to explain them. (*KrV A444/B472*)

**Anti-thesis:**

There is no freedom, but everything in the world happens solely in accordance with laws of nature. (*KrV A445/B473*)

According to ordinary logic, it seems as we have to accept *either* the thesis *or* the anti-thesis: that freedom exists or that it does not exist. Still, Kant argues that both these theses can be validated on the basis of their respective intelligibility. (*KrV A444-51/B 472-79*)

There is a lot that can be said about Kant’s line of argumentation here, but since my purpose in this chapter is not to evaluate the argument itself I will merely sketch the basic features of it as precursor to some very important

---

103 *KrV* A803/B831. Here Kant refers to practical freedom as “one of the natural causes, viz, as a causality of reason that is operative in the determination of the will”. See also *Prolegomena* where practical freedom is said to be “that freedom in which reason has causality in accordance with objective determining grounds.” (Prol. 4:346)
conclusions. So, when Kant later presents his resolution of the dilemma he says:

In respect of what happens, one can think of causality in only two ways: either according to nature or from freedom. The first is the connection of a state with a preceding one in the world of sense upon which that state follows according to a rule. Now since the causality of appearance rests on temporal conditions and the preceding state, if it always existed, could not have produced any effect that first arose in time, the causality of the cause of what happens or arises has also arisen, and according to the principle of understanding it in turn needs a cause. (A532/B560)

Now, if not before, the need for an acquaintance with Kant’s dual perspective of the world becomes inevitable. Apparently, he distinguishes between two kinds of causality of which the one “according to nature” refers to cause and effect in the sensible world or, as he also puts it, causality with respect to appearances. This kind of causality is subject to the conditions of time and, understood in the usual way, the necessity in this causal relation can in no way be united with freedom. Rather, Kant says, these two are directly opposed to each other as contradictory since from this reading follows that every event, or action, is the necessary consequence of effects of the preceding time, that is, things no longer within my control. Given this, no imputation would be possible since the person who committed a crime could not have left his crime undone as his action would have been the necessary consequence of determining grounds in preceding time. Because of this, we must conclude that no agent is ever free at the point of time in which he acts. (KpV 5:94) However, morality requires of all human beings that they refrain from wrongdoing, and thus, they must also be able to. Yet, our ordinary conception of causality according to nature cannot explain this moral demand. Now, if we return to the first Critique, Kant continues:

It is easy to see that if all causality in the world of sense were mere nature, than every occurrence would be determined in time by another in accord with necessary laws, and hence – since appearances, insofar as they determine the power of choice, would have to render every action necessary as their natural consequence – the abolition of transcendental freedom would also simultaneously eliminate all practical freedom. For the latter presupposes that although something has not happened, it nevertheless ought to have happened, and its cause in appearance was thus not so determining that there is not a causality in our power of choice such that, independently of those natural causes and even opposed to their power and influence, it might produce something determined in the temporal order in accord with empirical laws, and hence begin a series of occurrences entirely from itself. (KrV A534-5/B562-3)

Accordingly, practical freedom involves that our faculty of choice must have had the possibility to infringe upon the impelling causes that indeed affect us as a result of previous events in time. Yet, the passage also says that freedom
requires that it must be possible for our choice also to begin a new causal chain entirely from itself; that is, we must not only be able to choose between impelling causes but we also have to be the cause itself.

The explanation of the possibility of this is found in Kant’s well-known theory of the distinction between *phænomenon* and *noumenon*; between the thing that can be an object of experience and the thing in itself of which no direct knowledge is possible. Put very simply, whenever we observe something, we immediately and necessarily ascribe several properties or qualities to it in order to make it comprehensible to a human understanding. These qualities are, of course, Kant’s famous *categories* and *time* and *space* as the two forms of intuition. While the things in the world of appearances are subject to several natural laws and restrictions, there are no such factors in the intelligible world. Having understood this relation properly, we also realize that the same distinction must be applied to ourselves as well. This means that while we in one sense exist as living beings in the physical world (and thus are subject to the necessity of the natural causality), we may at the same time be considered totally free with respect to our noumenal selves. If this division had not been made, and if we took all descriptions of things and events *in one and the same meaning*, we would end up in obvious contradictions when we tried to ascribe freedom to the human soul while, at the same time admitting that it was also subject to natural necessity. (*KrV* B xxvii) If we, on the other hand, when assigning freedom to the actions of human beings, only referred to their *intelligible cause*, we could consistently maintain that there is an additional kind of causality—a causality of freedom—that exists alongside the causality of nature. Accordingly, Kant says: *if appearances would be things in themselves, then freedom cannot be saved.* If, on the other hand, appearances are taken to be nothing more than what they in fact are, that is, nothing more than empirically conditioned representations, then we might still account for a ground for them that is itself not an appearance. In this way we can say that the effect can be regarded as *free in regard to its intelligible cause*, and yet simultaneously, in regard to appearances, as the result of the necessity of nature. (*A536-7/B564-5, KpV* 5:101)

Indeed, Kant was aware of the highly subtle and obscure tone of this distinction but he nevertheless insisted in the enlightening effects of its application. Most of all he makes clear that his aim in these sections was never to *prove* the reality of freedom but only that the antinomy rested on an illusion and that there was thus no real conflict between freedom and nature.104

---

104 *KrV* A537/B565, A558/B586. It may here be mentioned that this transcendental argument is to be found in several *Reflections* during the decade before the publication of the first edition of the *Critique*. See e.g., the note from the mid-1770s saying that: “We cannot prove freedom *a posteriori*, because the absence of the perception of determining grounds provides no proof that nothing of that sort exists. We also cannot cognize its possibility *a priori*, because the possibility of the original ground that is not determined by another cannot be com-
dom is treated here only as a *transcendental idea* and Kant says that “we are fortunate if only we can be sufficiently assured that there is no proof of its impossibility”. (*KpV* 5:94) Such assurance, however, is a purely *practical* undertaking. The reason why this must be a practical task is simply that theoretical, or speculative, reason cannot judge at all about things that are not objects of possible experience without ending up in antinomies. The deficiency becomes particularly problematic when the human being obviously wants to make claims about supersensible objects like *freedom*, *God* and the *future world*. These are concepts that speculative reason “could indeed present as problems but could never solve” (*KpV* 5:132). Yet, as Kant says, for practical purposes it is sufficient that one’s assumptions do not involve any *intrinsic* impossibility (i.e. no contradiction). Since the human being is necessarily connected to the moral law he is also immediately linked to the idea of freedom, and because of this, Kant says, the human being cannot act otherwise than *under the idea of freedom*. Accordingly, the human being is free in the *practical* sense which means that “all laws that are inseparably bound up with freedom hold for him just as if his will had been validly pronounced free also in itself and in theoretical philosophy.” (*GMS* 4:448) In the *Lectures on the Philosophical Doctrine of Religion*, we find the somewhat stronger claim that the fact that we can act as if we were free lets us conclude that *eo ipso we are free*. (*VRT* 28:1068)

Finally, one may consider Kant’s moral argument for the need for postulating freedom. The moral law, as a fact of human reason, provides *a priori* knowledge of freedom, although, he adds, we cannot have any direct *cognition* of this freedom, nor we can claim that we have *insight* into it. (*KpV* 5:4) In this, Kant says, we have finally reached that which Archimedes searched for but never found, namely a *fixed point* which lies there as a sure foundation through the unshakeable moral law. (*Vorschritte* 8:402) The moral law, however, expresses nothing other than the *autonomy* of pure practical reason (i.e. *freedom*) that is also the form of all intellectual causality. Autonomy, in turn is the property of the will of being a law for itself. Thus a free will and a will under moral laws are one and the same. (*KpV* 5: 33, 5:74-5, *GMS* 446-7)

---

2.2. The Faculty of Desire

In the previous sections we have learned that the human being, in virtue of being transcendentally free, can begin a new series of events independent of what has happened in the past. This is, as we have seen, simply a consequence of the dual conception of the human being since if the human being takes account of his nature as *noumenon* there can be nothing antecedent to prehended at all. We thus cannot prove it theoretically at all, but only as a necessary practical *hypothesis.*” (*Refi*. no. 4724, 17: 688)
the determination of his will other than his own reason and the laws that he has himself prescribed. (KpV 5:98) In order to make these insights intelligible with respect to human agency, we need to relate freedom – in both its forms – to our faculty of volition. We have already been faced with the requirements of the incorporation thesis but in order to get a fuller view of how incentives – whether these be moral or non-moral – actually motivate we need to expand the concept of “will”.

In order to understand Kant’s theory of motivation, we need to find out in which way the human mind responds to stimuli. This process is basically outlined in the sub-discipline called empirical psychology. In its basic sense, psychology means *theory of the soul* [Seelenlehre] and can be either rational or empirical. According to Baumgarten, psychology is the theory of the general predicates of the soul. 105 Rational psychology (*psychologia rationalis*) analyzes the soul *a priori* from concepts and not from experience. It is the cognition of objects of inner sense insofar as it is borrowed from pure reason. Its basic focus is the statement “I am”, and the only thing that is taken from experience is the concept that we have a soul. We apply transcendental ontological concepts to it; it is a substance, simple and immortal. 106 The fundamental object for empirical psychology is the “I” as soul and human being. The body must, in this discipline, be thought of as an organ of the soul, or more precise, an organ that is dependent on the soul. However, the body may also be seen as a lodging of the soul and this makes the soul dependent on the body as well. Thus, empirical psychology includes a short anthropology or, as Kant puts it elsewhere, empirical psychology is anthropology. 107

For our purposes in this chapter, we are primarily interested in empirical psychology and the parts of it that cooperate in human motivation. Although Kant repeatedly denied that empirical psychology belongs to metaphysics, his most detailed outline of it is found precisely in the notes to his lectures on metaphysics. Yet this, we are told, is only because of convention for much had been placed in metaphysics simply because we do not really know what metaphysics is. 108 Now, in order to be able to understand Kant’s advanced model of the human will, we must first look at the basic features of

---

105 Baumgarten *Metaphysica* §501.
106 Cf. eg. *VMet* [L1] 28:223, 28:263, In a comment Kant wrote: “The propositions of rational psychology all are grounded on the ‘I am’. For even if time should be added to them, then it would be an object of experience that they were treating of, and everything that would be produced through this would not have to reach any further than to this life.” (AA 23:39).
107 *VMet* [Mrongovius] 29:876. Earlier in the manuscript Mrongovius said that a “psychology of observation” can be called anthropology. *VMet* [Mrongovius 29:757]. In a later price-essay Kant also defines psychology as anthropology; see *Vorschritte* 20:308.
108 *VMet* [L1] 28:223. According to Kant, empirical psychology belongs to metaphysics no more than empirical physics since both concern sensibility. In *Metaphysical Foundations of Natural Science*, Kant says that the empirical doctrine of the soul [empirischen Seelenlehre] is merely to be seen as a historical doctrine of nature and not as a proper science. (*MAN* 4:471)
man’s most elementary psychological faculties. The human mind can, in this case, be divided into three main “parts” or “faculties” that cover all further mental functions. These are the faculty of cognition, the feeling of pleasure and displeasure and the faculty of desire. The faculty of cognition is the only one that may operate independently of the others, but it is connected to the faculty of desire by means of feelings of pleasure or displeasure, for, as Kant says, we cannot desire or shun a thing if feelings of pleasure, or displeasure, are not involved. Furthermore, we cannot feel pleasure or displeasure at the representations of objects that we do not know, or have identified through our understanding; hence desiring requires the faculty of cognition. For our inquiry in this chapter, the most important of these is the faculty of desire.

The faculty of desire is the inner principle that makes us able to act in accordance with representations [Vorstellungen]. It is the faculty that manifests human life and is the faculty by which human beings can be the cause [Ursache] of the objects of our representations. Thus the faculty of desire is the power of pursuing the represented end which in effect motivates us to act in order to satisfy the desire. Human beings have both higher and lower faculties of desire, but it is important to understand that whether a faculty of desire is higher or lower does neither depend on their objects, nor on whether the representation has its origin in the senses or in the understanding, but only on the way by which this faculty is affected. If the determining ground of the action consists in a representation connected with pleasure, it belongs to the lower. A higher faculty of desire is also called will.

Since the will and related notions will be the subject of the next section, I will here concentrate on the lower faculty of desire that involved in non-moral decisions. As we have seen, the human mind is susceptible to feelings and Kant adds that the dependence of the faculty of desire upon feelings or sensations [Empfindungen] is called inclination and this always indicates a

---

109 Ger. Gemüt. The exact difference between the mind and the soul [Seele] is not always clear since empirical psychology [empirischen Seelenlehre] is said to concern the faculties of the mind.

110 Ger. Erkenntnisvermögen, Vermögen der Lust und Unlust, Begehungsvermögen. KU 5:177ff., KU "Erste Einleitung" 20: 207. See also the sections concerning “empirical psychology” in the lectures of metaphysics (all students).

111 At least, we must recognize the object as similar to something we already know. In order to desire it we must regard it as safe or interesting enough to stimulate our curiosity rather than fear or suspicion. Yet, as Kant would say, also the reaction to shun a thing is a response from the faculty of desire, namely “to desire the opposite”. Occasionally, Kant had labeled this counter-reaction of our faculty of desire as a “faculty of aversion” [Verabscheuungsvermögen or facultas aversativa]. VMet [L2] s. 587.

112 Cf. “the faculty of desire is the faculty of a being to be by means of one’s representation the cause of the objects of these representation. The faculty of a being to act in accordance with its representations is called life” (MS 6:212. See also KpV 5: 9n).

113 According to Kant, even sharp-minded [sharfsinnige] men have failed in noting this vital fact. See KpV 22-3. H. F Klemme adds that Kant is here referring to Wolf and Baumgarten. See Klemme p. 227. See also Beck (1960) p. 94n.
need. \((GMS\, 4:413n.)\) He goes on by saying that this dependence on sensations also implies that our interest is here not in the action itself but rather in the object of the action. \((GMS\, 4:413n)\) In other words, in these cases we want, or desire, specific objects or states while a moral action is done without reference to any such ends. The difference between these two kinds of interest is further spelled out in the first part of the second Critique where Kant, in Theorem I states that “all practical principles that presuppose an object (matter) of the faculty of desire as the determining ground of the will are, without exception, empirical and can furnish no practical laws”. \((KpV\, 5:21)\) This is why non-moral principles are material rather than formal.

2.2.1. Inclinations and Other Incentives

In the preceding sections we have been confronted with various arguments against the view that anything that belongs to – or borrows matter from – sensibility can claim to be objective reasons for human actions. Kant uses several names for our sensual impulses, the most common of which are inclinations, incentives, propensities, pathological interests, stimuli and desires.\(^{114}\) Some of these are more or less identical; for instance, Kant explicitly equates inclinations and pathological interests, and inclinations are said to be habitual sensual desires that are developed from our propensity [Hang] to desire the pleasant. An inner pressure on the faculty of desire to attain something one does not yet know is called instinct. \((Rel.\, 6:29n.)\) All these are subjective and sensual, and Kant does not seem to allow for any moral inclinations or desires. Moral impulses are generally called motives [motiva, Bewegungsgrund] but Kant also allows for moral interests and one moral incentive. As we saw above, the difference between a moral interest and a pathological one is that in the case of moral interest, we act because of an interest in the action or simply because we want it to be performed. The sole motivation involved here is the very fact that the action is morally right and hence the interest is in the mere “rightness” of the action without expectations for any possible agreeable objects that might result from it. \((GMS\, 413n.)\) Moral actions are motivated by a special kind of feeling namely a respect for the moral law [Achtung]. Respect is thus the sole moral incentive and the only one that can give an action moral worth. It is derived solely from pure reason and the awareness of duty.

In an attempt to summarize the fundamentals of Kant’s idea of motivation one can say that all human actions originate in our faculty of desire. This faculty, as we have seen, is the faculty that by means of its representations [Vorstellungen] is the cause of the reality of the object of these representations. Consequently, every action requires a representation or mental idea of an end. Moreover, this representation must also be attached with a feeling

\(^{114}\) Ger. Neigung, Triebfeder, Hang, pathologische Interesse, Stimuli, Begierde.
that, in the case of non-moral actions (i.e. actions from inclinations), is a feeling of either pleasure or displeasure. Pleasant representations are thus the foundation for all non-moral acting. The relation between desires and the feeling of pleasure is stated in the axiom: *Nihil appeto nisi quod placet, nihil averto nisi quod displicet,* that is, we desire nothing but what pleases, we avoid nothing but what displeases. From this follows that desires are always connected with pleasure. In other words, that which gives me no pleasure, I simply do not want.

Hence, pleasure and displeasure (*voluptas* and *taedium*), or – as Kant prefers – satisfaction or dissatisfaction (*complacentia* and *displicentia*) are the foundations for all desiring. Yet, we must distinguish between some forms of pleasure for pleasure may be either *pathological* (i.e. sensible/material) or *moral*. Here we find a crucial condition:

– If the law precedes the determination of the faculty of desire (will) and pleasure follows, then, the pleasure is moral; if pleasure precedes the determination of it, then it is pathological.

This axiom, as one might call it, figures frequently in Kant’s argumentation, sometimes in his works, but often as one single strophe in reflections and posthumous writings. It is clear that if a representation of pleasure is present in the mind before the action is performed (and thus affects the direction of it), then the action and the pleasure is non-moral. In contrast, moral pleasure can only follow as an effect upon an action that has already been determined by the law, and this occurs when we feel satisfaction in having acted as morality demands. In these cases the determining incentive, that is, respect

---

115 *VMet.* [Mongrovius] 29:894. This might be a more adequate description than the similar axiom *nihil appetimus, nisi sub ratione boni; nihil aversamur, nisi sub ratione mali,* for we must here take notice of the important ambiguities of the terms *bonum* and *malum.* According to Kant, the German language here can make a distinction not available in Latin; we desire in the subjective terms well and woe [*Wohl, Übel*] and not in the objective (*a priori* knowable) terms of good and evil [*Gut, Böse.*] It can here be mentioned that we often desire the evil and shun the good if these terms are taken in their intellectual sense. See *KpV* 5:59 ff. *VMet* [Mongrovius] 29:896; *VM* [Collins] 27:263.


117 *VMet* [Mongrovius] p. 890, *VMet* [Vigilantius] p. 1013. Although Kant apparently wanted to make certain distinction it might be mentioned that Baumgarten in fact said that *voluptas* corresponds not only to Lust, but also to Gefallen and Vergnügen; in the same way *taedium* refer to Unlust, Missfallen, and Missvergnügen. (*Metaphysica* § 655) Cf. G. Lehman: comments to *VMet.* 29:1126-7.

118 There are numerous similar formulations of this important criterion for the moral status of pleasure in Kant’s corpus. See e.g. *NET* 8:395n: “That pleasure (or displeasure) which must necessarily precede the law, if the act is to take place, is pathological; but that which the law must necessarily precede for this to happen is moral”. See also *MS* 6:378; *KF* 5:41, *Refl.* no 7320 19:316; *OP* 21: 70, 75, 90, 95, 132; 22: 106, 114, 117, *VB/MS* 23: 375, and more.

119 Pleasure may also (as in the case of aesthetic reflecting) be purely *contemplative* and thus not connected to an impulse to realize an object.
for the law, can indeed also be seen as the subjective ground determining the agent to act according to the objective moral law. Although respect is a kind of feeling – the moral feeling – it is by no means a feeling of pleasure but rather the opposite. It is an awareness of something that makes us tremble by exposing the moral law itself to us in its solemn majesty. (KpV 5:77) In this way, my consciousness of the moral law makes me realize the insignificance and unworthiness of my subjective pathological wants, and this insight forces me to abandon all private wants and instead make the objective law the determining ground of my will. In the strict sense, Kant says, respect for the law is not the incentive to morality but morality itself subjectively considered as an incentive. It is the awareness that analogous to an ordinary feeling attaches the representation of the end (in this case the end set by the moral law) and makes me motivated to act in order to realise it. Since such ends are objectively good, a will governed by such principles is a good will.

In view of this, it is clear that inclinations cannot qualify as proper motives of the unqualified and absolutely good will since such a will must derive its goodness from the formal ground of the action rather than from the anticipated aim. Indeed, since we often have plenty of representations present in our mind, we will generally find ourselves as standing at the mental crossroad where our will needs to choose between its formal (a priori) principle and its material (a posteriori) incentive. (GMS 4:400) Quite often, these distinct motivational grounds can lead to the same action. This will be when that an action is in perfect accordance with duty. This is, for example the case with all those persons who, entirely in conformity to duty preserve their lives but do so because of their love of it. It is also the case for the philanthropic man who gladly performs the duty of charity by helping his fellows because he cares for them and rejoices in the happiness he gives them. We may also consider the truthful shopkeeper who treats everybody with the same uprightness and who does not exploit the opportunity to deceive less experienced customers simply because an honest behaviour (and especially his reputation as being honest) is good for his business. It is clear, Kant says, that none of these actions, however lawful and appropriate, can be regarded as worthy of esteem since they all rest on purely subjective maxims, that is, maxims based on inclinations or self-interest. Only a suicidal man, who keeps on living his miserable life solely because his sense of duty forbids him to end it, or the cold-hearted person who does not want to help others but yet does so out of duty, perform actions worthy of moral esteem. The same holds for the merchant who is equally truthful as his just mentioned colleague but who acts simply because he recognizes the principle of hones-

---

120 It is not clear what status Kant assigns to (the feeling of) respect. In the Groundwork he says that it is similar to a feeling [gleich eine Gefühl] and in the Critique that it is a very specific kind of feeling [eigenthümliche Art von Empfindung]. In both works, however, it is clear that it is wholly different from ordinary feelings and that it must have its origin in pure reason and not at all in sensibility. (GMS 4:400 n; KpV ch. III)
ty as his duty. Again, Kant says, we find an example of morally valuable behaviour. (GMS 4: 397ff.) In short, an action from duty must, as Kant puts it:

put aside entirely the influence of inclination and with it every object of the will: hence there is left for the will nothing that can determine it except objectively the law and subjectively the pure respect for this practical law, and so the maxim of complying with such a law even if it infringes upon all my inclinations.121

One may have objections to this way of thinking, and Kant’s supposed antipathy for feelings was criticized and even ridiculed already by his contemporaries.122 Although much of the commonplace criticism – especially of his view on compassion and philanthropy – is based on a misunderstanding of Kant’s intention, it cannot be denied that for Kant, frequent acting from inclinations often indicates a weak and dependent constitution that cannot resist the temptation of these sensual impressions. Yet, it is important to be aware that not even these (indeed ill-reputed) examples tell us anything about Kant’s general ideas of feelings in the human life. Even if the moral law undeniably is a holy directive it does, in fact, not command that we always ought to renounce all claims of private happiness but only that we, when duty is in question, shall take no account of them. (KpV 5:93) Moreover, we learn that happiness is a natural goal for human beings and that the law does not require that we should reject all claims of happiness; indeed, Kant explicitly acknowledges that the promotion of one’s own happiness is also an indirect duty. This is essentially what Kant says in the beginning of Chapter III of the second Critique when he says that the moral law merely represses self-love and tries to restrict it; so far as one’s happiness does not conflict with morality it may be called rational self-love. Indeed, to eliminate all influence of self-love would not only be in vain but also directly contrary to the end of humanity since it is a natural principle that has existed in us far before the moral law.

It is thus important to widen our focus from the notorious passages that most support the stereotypical view of Kant’s thinking. Surely, there is textual evidence where he says that inclinations frequently lead us astray. Likewise, there are others where they are said to be a burden for the right thinking, and Kant even claims that every rational being would wish to be completely free from them.123 However, in the Religion Kant definitely

121 GMS 4:400f. Cf. also “for, in itself duty is nothing other than the limitations of the will to the condition of giving of universal law possible through a maxim adopted, whatever the object of the will or the end may be (thus happiness as well), from which, as well from every end one may have, we here abstract altogether”. TP 8:279f.
122 The most famous example of early mockery is Friedrich Schiller whose poem “scruples of conscience” puts forward a rough caricature of Kant’s thoughts on helpfulness and friendship. Anthr. 7:151; KpV 5:118; GMS 4. 428.
seems to have amended this view since natural inclinations as such are said to be good and that it would be both futile and harmful, or even blameworthy to try to extirpate them. (Rel. 6:58) Moreover, as we have seen, all of our three predispositions – including both animality and humanity (in spite of their possible hazardous consequences) – are, according to Kant, predispositions to good. As long as these natural inclinations do not develop into passions, they may well be acted upon when morality is not the question.\textsuperscript{124} It is thus important to be aware that Kant did not regard inclinations as evil in themselves, and human evil does, in fact, not originate in our sensual nature. The reason why inclinations themselves cannot be the cause of evil is that evil, in its proper sense, must be understood as moral evil and not only in terms of a mere limitation of human nature. (Rel 6:43) As we have seen, we are morally impure beings with both natural and artificial needs. The satisfaction of these needs is secured through our predispositions to animality and humanity – by means of which we preserve ourselves, eat, reproduce and strive for well-being and happiness. Any empirical desire is grounded in our sensibility, but morality must originate in freedom which, as we have seen, always gives us the possibility to refrain from acting against duty.

2.2.2. Feelings and Pleasure in Motivation

The just mentioned criteria for non-moral motivation – that pleasure precedes, and directs, the determination of our will – suggests that Kant fulfil the criteria for being a psychological hedonist with regard to all actions that are not motivated by the sense of respect for the law. Besides the claims we have already seen, Kant repeatedly states that the representation of the desired end is a “foreseeing of what was to come in the future”, and he adds that it is on the basis of such expectations that we choose among present alternatives. (VR [Pölitz] 28:1060) Moreover, in the second Critique we are told:

\begin{quote}
when one inquires for the determining grounds of desire and puts them in the agreeableness expected from something or other, it does not matter at all where the representation of this pleasant object come from but only how much it pleases (KpV 5:23)
\end{quote}

Again we may recall that Kant found it amazing that so many intelligent thinkers still believed that there could be a significant difference whether the representations had their origin in the understanding or in the senses. From

\textsuperscript{124} Again we may recall that for Kant, passions [Leidenschaften] are not just inclinations (sensual desires) but inclinations strong enough to overpower our rational faculty (Rel. 6:29n.). Passions are developed from affects [Affekten] that are excessively strong feelings. For example, fury is an affect while hate is considered a passion. Kant carefully distinguished these two notions since only passions have an influence on the will. See e.g. KU 5:272, Anthr. 7:266f, Rel. 6:29n.
this it follows that for Kant, there was no real difference between the want to read a book, to engage in intellectual conversations, to eat, to visit the cinema or the wish to gain wealth and power or even the desire to help a poor fellow since they are all fundamentally grounded in one and the same basic principle, namely self-love or our own happiness. Accordingly, the only thing that matters when the agent is to choose between alternative non-moral ends is how intense, how long lasting, how easily acquired and how often repeated this agreeableness is. (KpV 5:23) Admittedly, this bears several similarities to a classical hedonistic view. Moreover, one must also admit that Philip Griffiths has a point when he somewhat sarcastically says that it is not unlikely that Jeremy Bentham might have been inspired by this idea of purely quantitative calculation of the amount of pleasure expected from the action.125

Now, while it is undoubtedly true that many (if not most) of our non-moral actions do aim at happiness or pleasure, the very idea that all our seemingly different wants are reducible to one single desire for private pleasure obviously conflicts with our intuitive understanding of human willing. Bernard Williams concludes that Kant apparently identifies all non-moral acting with a determined striving where we, like animals, only act for the sake of pleasure, and Griffiths does not spare the words when he calls Kant’s idea “utterly repugnant, derogatory and degrading.”126 Samuel Kerstein is somewhat more modest in his disapproval but admits that if the plain reading is correct, Kant must be said to defend a suspect theory of motivation.127

Suspect or not, it will become clear that psychological hedonism does indeed follow from Kant’s general arguments but we will also see that this is not as unintelligible as it may initially appear. Because of the clear-cut statement made in Theorem II, the hedonistic conclusion has commonly been taken for granted. As a result of this, Kant has been called an absolute, radical and genuine hedonist and egoist.128 According to Griffiths, we do not even need to perform any “laborious combinations” of Kant’s many arguments in order to verify this conclusion since this statement cannot be read in any other way.129 Similar arguments appear in texts by Terence Irwin and Lewis White Beck, but we also find them in charitable readers like Kerstein

---

125 Griffiths p. 212. Cf. Bentham who famously claimed that pushpin is as good as poetry if it caused at least the same amount of pleasure. Interestingly, he uses almost the same parameters that Kant mentions in the second Critique namely quantity, duration intensity, certainty, closeness, purity. John Stuart Mill latter rejected this and argued for a qualitative hedonistic theory in which it would be better to be an unhappy Socrates than a happy pig. (Mill, Utilitarianism ch. 2.)
126 B. Williams p. 64, Griffiths. p. 212.
127 Kerstein, p. 23.
128 I will return to the discussion of egoism and how it may (or may not) be connected to self-love in section 6.2.2.
129 Griffiths p. 211. See also Edwards p. 419 n. 26.
who regardless of his quite reasonable allowance for modest form of hedonism where not every material action is motivated solely by private pleasure, resigns in favour of the common reading on the basis of it being the most plausible one.\textsuperscript{130} In line with recent discussion I will hereafter call the hedonist/egoist reading of Kant’s theory of non-moral motivation the traditional interpretation.

The inherent objectionable tone of the traditional interpretation has inspired commentators to find alternative readings that they have presented as modest or sophisticated ways of understanding Kant’s hedonism.\textsuperscript{131} While a full analysis of this topic falls beyond the scope of this thesis, there are at least some things that may be mentioned. As will become clear, a substantial part of the disagreements depends simply on a misconception of the very meaning of psychological hedonism. The claim that pleasure plays an important role in acting does, not, after all, tell us very much and, unfortunately, it seems that many commentators have been misled by an oversimplified terminological consensus. Already in the introduction to his alternative non-hedonistic interpretation of Kant, Reath defines psychological hedonism in the following way:

\begin{quote}
Psychological hedonism is a thesis about the objects of our desire which leads to a theory of motivation. It holds that all desires are desires for pleasure in the agent, or for the means thereto, where pleasure is construed as a definite feeling or experience. Alternatively it is the thesis that the only object desired for its own sake is pleasure.\textsuperscript{132}
\end{quote}

This definition infers that all desires can be reduced to one ultimate end: private pleasure. This conception is obviously reductive, and if we assume that Kant was a psychological hedonist in this sense, the wish to modify the reading into a less counter-intuitive one is certainly understandable. While this narrow conception makes it easier to find a reading that is not hedonistic in this particular sense, we need an extended idea in order to account for the entire motivational process that is present in a human mind and precedes the decision to act in a specific way. It is also important to be aware that even if the just mentioned definition of hedonism is the most common, it is not the only one. Iain Morrison, demonstrates his awareness of this fact by quoting the \textit{Cambridge Dictionary of Philosophy}:

\begin{quote}
Psychological hedonism itself admits of a variety of possible forms. One may hold, e.g., that all motivation is based in the prospect of present or future pleasure. More plausibly, some philosophers have held that all choices are
\end{quote}

\textsuperscript{131} See especially Reath (2005b), Kerstein ch. 1 and Herman.
\textsuperscript{132} Reath p. 37.
based on one’s presently taking greater pleasure in the thought of doing one act rather than another.\textsuperscript{133}

As we will see, this second “more plausible” view is more important than it may initially appear. While a wider conception of hedonism can be seen as an obstacle to those who want to remove the hedonistic conclusion from Kant’s psychological theory, it is the key to the “modest” interpretation just mentioned. The reason for this is simply that the expanded view does not require that the pleasure that indeed must figure at some stage of the motivational process is also necessarily the object of the desire, that is, the end that we try to realize when we decide to act in accordance with a specific desire.

In order to see the opportunities of this alternative definition we must recall that Kant explained the faculty of desire as a faculty which by means of its representations motivates the agent to attain the object of those representations. We also saw that the feeling of pleasure that necessarily is present in non-moral motivation attaches to this representation or thought of the end. While it may not be immediately obvious, a closer consideration of this process shows that the latter way of understanding psychological hedonism differs fundamentally from the purely teleological one. While the former is a theory that claims that the agent is motivated solely by the pleasure that he expects to obtain from the attainment of the desired object, the later theory does not imply that we must have a representation of a pleasant object or of any prospective agreeable experiences. In this case, it is the mere representation of the object that has to be pleasant. Admittedly, this way of understanding Kant’s idea of non-moral motivation is not at all unique or revolutionary. On the contrary, some of the most comprehensible and contemporary interpretations of Kant’s hedonism basically explain the process in this way.\textsuperscript{134} Still, there is much disagreement about the implications of this reading and, as I have argued elsewhere, none of the readings that I have seen can be said to capture the essence of Kant’s complex theory of motivation.\textsuperscript{135} While I will not repeat the entire argument here, I will at least say that it is clear that the first and most significant form of motivational hedonism that operates in the moment of choice is the just mentioned pleasure that is tied to the idea of the object. This kind of pleasure has, in contrast to the pleasure that is anticipated (expected) been called “pleasure of anticipation” or “anticipatory pleasure”.\textsuperscript{136} Now, in order to briefly sketch the procedure by which we are affected by feelings I believe that Johnston is correct when he suggests that the motivational process runs as follows:

\textsuperscript{134} See especially Reath (2005b), Johnston and Morrisson.
\textsuperscript{135} I discuss this in Wennersten (2013).
\textsuperscript{136} For the former expression, see Johnston and for the latter see Morrisson ch. 2. Also Reath seems to be aware of this form of pleasure but he does not have any specific name for it. See Reath (2006b), especially the Appendix.
The (proximate) genesis of an action begins with the representation of the action’s end; the agent finds the representation of this end pleasant [...]; this pleasure [...] gives rise to a desire for the end; and the agent freely chooses to act on the desire.137

Using the terminology mentioned above, one can thus say that pleasure of anticipation causes the desire in virtue of the pleasantness of the mere idea of the object. Yet, in order to become an effective motive for acting and thus to make us choose one alternative before another, we need to assume the presence of anticipation of pleasure – the pleasure that we expect to attain when the object is realized. At this stage, we will apply the conditions listed above and decide to do what gives us the greatest share of pleasure for the longest time. In providing this explanation, Johnson argues that he has put the final nail in the coffin that once and for all will rule out all possible interpretations that aim to disprove the traditional interpretation. Yet, this argument fails to live up to his claim, and while the last condition might be regarded as a setback for those who want to disprove the hedonist reading completely, it nevertheless acknowledges that a multitude of possible objects that can be desired also by the non-morally, and hedonistically affected will. As many ethical theorists have pointed out, we may have pleasant thoughts about attaining knowledge, health, or personal improvement in general and we may even think about truly unpleasant objects in a pleasant way but – most importantly – we can also entertain these pleasant or unpleasant feelings in the thought of someone else’s state of affairs.138 This conclusion obviously fits the Kantian scheme; consequently, the process that Johnston refers to does not establish the fact that we only desire pleasure but only that we, in the end, decide on the basis of pleasure.

This is definitely a theory that would deserve to be called both modest and perhaps also sophisticated. As such this suggestion bears several similarities to the system presented by some early critics like Joseph Butler and David Hume who openly denied that self-love could, by itself, direct our will towards any specific objects. According to them, our expectation of gaining pleasure from an end requires an already existing desire for it. In other words, if we did not already have affections towards the end itself, no pleasure would arise. As Butler put it, “pleasure or happiness consists only in the enjoyment of those objects, which are by nature suited to our several particular appetites, passions and affections”.139 Or, as Hume put it, “I feel pleasure in doing good to my friend, because I love him; but I do not love him for the sake of this pleasure”.140

137 Johnston p. 54.
138 For this idea, see e.g. R. B. Brandt p. 312, Schlick p. 128, and Reiner p. 41.
139 Butler, Sermon XI:6, 9.
140 Hume (1741) p. 48.
2.3. Will and Choice

Until now we have been concerned with the way that inclinations affect our faculty of desire but from what was said above, it is clear that Kant could not accept a conception of the human volition that made it directly determinable by an incentive or arbitrary impulse. Everything in nature Kant says, “works in accordance with laws. Only a rational being has the capacity to act in accordance with the representations of laws, that is, in accordance with principles, or has a will.” (GMS 4:412)

Thus is the will a particular form of the faculty of desire, a faculty which, as we know, is a faculty to act in accordance with representations. As we have seen, Kant obviously needed to ascribe freedom to human beings in order to allow for their status as moral beings. Although it is clear that the capacity to make decisions independently from external influences must be attributed to the activity of our will, we need to examine this area more closely for it is it is not without reason that Jens Timmermann calls Kant’s discussion of human volition a “terminological minefield.” Kant’s many definitions of the will must be made mutually compatible and conceivable. Moreover, they must also be consistent with further procedures in the motivational process, like the way desires come about and how these actually affect our will, and how we in the end choose to make a certain incentive the determining ground for acting in line with the requirements of the incorporation thesis

In light of these complexities, it is no exaggeration to say that the human will is one of the most central, but also one of the most difficult notions in Kant’s practical philosophy. The Groundwork begins with his legendary speech about the endless goodness of the good will and ends with the conclusion that the will is free in the sense of being law-giving (i.e. autonomous), and thus only subject to laws that it has set for itself. This is, however, just one of Kant’s many definitions of will. Among his more familiar descriptions we find:

- A capacity to act in accordance with principles. (KpV 5:58-9)
- A capacity to act in accordance with the representations of a rule as a law. (KpV 5:32)
- A capacity to determine itself to acting in accordance with the representation of certain laws. (GMS 4: 412, 4:427)
- A kind of causality of living (and rational) beings. (GMS 4:446)
- Practical reason itself. (MS 6:213, GMS 4:412)
- The moral cast of mind [sittlicher Denkungsart]. (MS 6:387)

We will need to return to these definitions several times in the following chapters. What makes the discussion of Kantian agency complicated is that

---

Kant apparently wants to make distinctions between different aspects of the human faculty of desire. Our ability to act from representations is not only divided into a lower and a higher faculty, but a distinction is also introduced with respect to the process of willing in general. The most familiar is the division between *Wille* and *Willkür* which are usually translated “will” and “faculty of choice” respectively. While the distinction is far from uncomplicated, it is commonly argued that Wille is to be understood as the legislative aspect promoting the executive part – Willkür – to bring about action. This idea gains support from the introduction to the *Metaphysics of Morals* where Kant says:

> Insofar as [the faculty of desire] is joined with one’s consciousness of the ability to bring about its object by one’s action it is called choice [Willkür]. The faculty of desire whose inner determining ground, hence even what pleases it lies within the subject’s reason is called the will [Wille]. (MS 213)

Therefore, Kant adds, the will is the faculty of desire that has less to do with actions than choice has. The role of the will is rather to determine choice to action. In this sense, the will (as Wille) is the faculty that stands in relation to our reason and deliberates about possible ends. This will then tries to convince our choice to choose according to the appropriate principle. Since the Wille-Willkür distinction is commonly presented as highly important, one may wonder why he did not explicitly announce the difference before the above quoted passage in the *Metaphysics of Morals* from 1797. In reply to this, many commentators argue that Kant was well aware of the dual function of the will even before that, and that the distinction is implicit already in the *Groundwork*’s deduction of the moral law.\(^{142}\) It is without doubt true that Kant needed both these functions in order to reach a coherent picture of the way we reason about moral matters, but it is not always clear when we are to understand the term “will” in the sense of Willkür rather than Wille. Moreover, although Willkür figures sparsely in the *Groundwork*, it is not used in the sense of a faculty of choice but rather in the sense of the arbitrary aspects of our willing; those that should be limited or excluded by those things that are ends in themselves and thus objects of respect. (C.f. e.g. *GMS*: 4:428)

In dealing with these matters, it should be made clear that there are several problems with respect to the Wille-Willkür distinction. In spite of the seemingly clear-cut explanation from the *Metaphysics of Morals*, which is cited in almost every discussion as the definition of the terms, there is much more to be said since a closer look at Kant’s theory of the faculty of volition turns much of what may be derived from the quote upside-down. First of all one must be aware that Wille is repeatedly used in a very imprecise sense, seemingly expressing our entire faculty of willing, including wanting, choos-

ing and acting in general. In order to understand Kant’s earlier outline of will and autonomy correctly, commentators claim that one must distinguish between the will [Wille] in its broader sense – which include both the legislative and executive aspect of volition – and will [Wille] with the narrow function involving only legislative faculty. On this reading, will in its broad meaning covers both Wille and Willkür. Besides this, one need only take a look in Kant’s overall writings in order to find several occasions where he defines Willkür in the same way as he generally defines the Wille, namely as the capacity to act from laws or principles.\(^{143}\) This definition is also identical to Kant’s definition of both character in general and character in the more specific sense, namely as our “Denkungsart”. I will return to the relation of will and character in chapter 3.

Although certainly useful and illuminating, I believe that the Wille-Willkür distinction from the *Metaphysics of Morals* should be handled with much care and not imported into earlier works without consideration. After all, it seems as if almost all of the definitions of will listed above are perfectly coherent with the definition of Willkür, and this would mean that Kant is generally using Wille in the broad sense. For example, we know from the incorporation thesis that the human choice [Willkür] is an *arbitrium liberum* that cannot be determined by incentives but only by maxims. This definition perfectly coheres with several of the definitions of the will [Wille] as a “capacity to act from principles” since a maxim is a subjective principle which contains both the form and the matter of the act in prospect. It is this principle that we ought to test against the categorical imperative in order to see whether it can pass as a universal law or not. If we recall the long passage from the *Groundwork* quoted in the introduction,\(^{144}\) we once again see that the worth of an action can be found nowhere else than in the principle of volition without regard for any object, purpose or actual effect. (*GMS* 4:399f) Kant soon deepens this picture by saying that the good principle can be nothing other than the representation of the law itself, a feature that can only be found in rational beings. (*GMS* 4:401; 4:412) Furthermore, he adds that since reason is required for the derivation of actions from laws, the will is nothing but practical reason. The difference between practical and theoretical (speculative) reason is that while the latter only determines the object and the concept of it, the task of practical reason is to make the object real. (*KrV* B x) Again, in the introduction to the second *Critique*, Kant makes clear that theoretical reason deals solely with the object of the cognitive faculty while practical reason is concerned with the determining grounds of our will, and consequently to the realization of those objects that correspond to

\(^{143}\) This is especially common in the *Lectures on Anthropology*.

\(^{144}\) See p. 23 above
our representations.\textsuperscript{145} Hence, since the will is commonly explicitly said to be a capacity \textit{to act} from representations of laws or higher principles, it is thus given the sense of an \textit{executive} power. Clearly, the idea of a \textit{pure} will, “in the narrow sense” of a faculty that provides only the laws themselves, may be inherent in works prior to the “Doctrine of Virtue” and this perspective would, indeed, fill some gaps in Kant’s reasoning. Yet, the minefield-like nature of the area requires, as just said, that we navigate with care. This will be even more obvious in the next chapter where we need to deal with the multifaceted concepts of character and Denkungsart.

2.4. Summary

The purpose of this chapter was to present an overview of some of the most important aspects of Kant’s theory of empirical psychology, namely the faculty of desire and how a decision to act springs from this faculty that, as we have seen, can be either higher or lower. It shall not be denied that the arguments have sometimes been highly complex since they appeal to transcendental ideas of dual “worlds” or “standpoints”. Because of that, it is not easy to see what Kant actually has proved though it is clear that he believes that he has successfully solved the antinomies.\textsuperscript{146}

In order to sum up the most essential lesson from this chapter I would like to begin by referring to what Kant says in the preface to the \textit{Groundwork}. We were told that there are two laws of the world, \textit{nature} and \textit{freedom} corresponding to two fields of science – \textit{physics} and \textit{ethics}. (\textit{GMS} 4:387) Freedom and ethics are thus explicitly connected and it is in this context that the need to assume freedom for moral purposes becomes inescapable. The moral law exists, and since we are limited beings, its demands are imperatives and duties for us. Since “ought implies can”, the very existence of moral demands requires that we must also be able to act in accordance with them regardless of all possible obstacles that our sensible nature raises against our moral conviction. This capacity to be free from determination from impulses can be explained in terms of practical freedom, but, since the existence of this ability is said to require also the transcendental form, further investigations were required. Now, as I have argued, the need for transcendental freedom is obvious simply from the fact that we need to assume a source of the

\textsuperscript{145} \textit{KpV} 5:15, cf. also 5:120 “The interests of \textit{reason’s} speculative use consists in the \textit{cognition} of the object up to the highest a priori principles; that of its practical use consists in the determination of the \textit{will} with respect to the final and complete end”.

\textsuperscript{146} I have intentionally left out the discussion concerning whether Kant’s dualism expresses two \textit{worlds} or merely a two \textit{aspects} or \textit{standpoints}. Both interpretations seem to be faced with problems and since my main discussion within this thesis does not seem to be dependent on this query I will not here take a stand in this debate.
principle that is supposed to begin the new series of events, and this principle, or at least the law for it, must spring from reason itself, as a pure self-activity, otherwise referred to as absolute spontaneity.

Kant’s writings on freedom and will also highlight the need not to confuse practical and mere speculative reason. As Kant says, “the theoretical use of reason we are concerned merely with objects of the cognitive faculty only”. Practical reason, on the contrary, is concerned with “the determining grounds of the will, which is a faculty either of producing objects corresponding to representations or of determining itself to effect such objects [...] that is, determining its causality. (KpV 5:15, emphasis added) From this it is clear that practical reason, our source of the idea of freedom is given the sense of an effective power – a power to bring about objects.

With regard to the debate concerning Kant’s hedonistic position, I have argued that one should not overstate the impact of it. While psychological hedonism may be regarded as a somewhat primitive theory that, as some have put it, reduces our willing to an animal-like striving, the division between non-moral and moral incentives is, after all, a cornerstone in Kant’s theory of agency. For Kant, ethics is not a theory of happiness but a theory about our worthiness of being happy. Still, as we have seen, this division does not have to reduce all our aims to one single demand for private happiness. As I read Kant, he merely aims to say that our desires for an object are caused, or are at least made active, by the feeling of pleasure that we experience in the thought of the object. Actions, in turn, are brought about by anticipation of pleasure but this last criterion is subordinate to the former. Accordingly, the principle of self-love is a general term that signifies our lower faculty of desire as opposed to the higher by differentiating the way by which the faculty is affected without thereby saying anything about the end.

I hope that we now have a better image of how Kant conceived of human actions and their origin. The incorporation thesis can serve as a starting point since it connects some of the most central arguments in this area. It also adds some vital information to Kant’s earlier definition of the arbitrium liberum which was only said to be free in the sense of being independent of sensible impulses. The incorporation thesis seems to take this independence one step further by stating that the faculty of choice cannot be determined by any incentive – “whatever it might be” – unless this incentive has been admitted into a maxim. Using the Wille-Willkür distinction from the Metaphysics of Morals the action could perhaps be described as follows: the will [Wille], in virtue of being autonomous and capable of giving universal laws, provides us with a rational law that is fit for being the incentive of a moral maxim. At the same time, we are constantly affected by pathological incentives that make the volition dependent upon the object of the desire (i.e. the desired end – a material entity). Accordingly, we need to decide whether to express our higher nature and follow the law, or whether we, as the sensual beings

95
that we are, do what we expect to be pleasant. Indeed, we may sometimes want the good and thus reach for the moral incentive. Still, in spite of these good intentions we sometimes fail in the execution of them since the power of the inclination may be stronger. We are weak [Ger. gebreichlich, Lat. fragile] and do not actually do what we would want to do. (Rel. 6:29)

We must now leave the remaining complexities aside but we will return to some of them in the forthcoming discussion where we will apply these theoretical ideas on practical problems, and, in so doing, deepening our knowledge of human nature. Indeed, as was mentioned already in chapter 1, one of Kant’s most interesting definitions of the term “nature” was the “subjective ground of the exercise of the human being’s freedom in general” (Rel. 6:21) One major task will be to explain how freedom actually expresses itself in man’s character or fundamental disposition, that is, the faculty of our nature which is said to “apply to the entire use of freedom universally”. We will also see how Kant searches for the ultimate ground of human actions and thus of both the use and misuse of the power of choice with respect to the moral law.
3. Character

In the two previous chapters we have been looking at some central features of human nature and the active powers of the mind that make rational action possible. We have learned that we have a faculty capable of setting ends (a will) and by making maxims, we then actively choose whether or not to act in order to pursue these ends. This, in turn, presupposes that we are free not only in the negative sense of being independent of the blind causality of nature but also that our will is free in the sense that it can be a kind of causality that is itself not caused by anything. In chapter 2 we saw that the human will has this ability since it, by means of its connection to an intelligible world, is capable of giving laws for itself that are free from all admixture of sensibility. We have also learned that Kant commonly defined the will, and occasionally also the faculty of choice (Willkür), as the agent’s capacity to act from maxims: the subjective laws, rules or principles that he has made for himself.

In this chapter I will deepen the conception of these faculties by focusing on the rules that govern them. As we will see here, and that will be further explained in the next chapter, the rules or principles comprising the foundation of our actions (as their motivational grounds) are not only to be attributed to the maxims themselves, but also to the cause or origin of them. Even if maxims are generally said to spring from our rational will or from our free choice, we still need to explain the reason why our will actually chooses to create just these principles and not others. In order to do that, we must find what Kant calls “the ruling principle” of our volition. This is our character which, like our will, is frequently defined as man’s “capacity to act from maxims”. Once again we will need to enter into a field of Kant’s thinking that, because of its importance and complexity, has caused much trouble for interpreters. Although the real challenge with respect to the moral status of our “inner moral foundation” will be the topic of chapter 4, I will here disclose the difficulties in Kant’s very notion of character and perhaps also resolve some of the misinterpretations that result from incomplete understanding of this term. As will become clear, the definition of character just stated is only one of Kant’s ways of explaining character, and in contrast to the “terminological minefield” that we met in chapter 2 – where we had to handle several different, though related, terms – we need now to identify several meanings of one single term. An important aspect of our inner – and specifically human – nature is the form of character called Denkungsart, but
as we will see, because of the failure to recognize the complexity of this term, some of the problematic implications drawn from Kant’s discussion of character are sometimes overstated.

The chief reason why one needs to be aware of the diversity of Kant’s conception of character is that one of the more troubling claims that Kant made with respect to character in the sense of Denkungsart is that this capacity only rarely figures in human beings.147 This, combined with the general definition of character and Denkungsart as the capacity to act from maxims, seems to imply that most human beings do not, in fact, act from maxims. Still, in the light of the incorporation thesis, we immediately see that this suggestion would contradict Kant’s idea of freedom since the human choice, as an arbitrium liberum, was said to be free from determination by pure impulses. A natural way of escaping the troubling implications of this claim would be to simply admit that we are less sophisticated than we ought to be, and that most of our actions therefore, are not performed from fixed formal principles even though we certainly would wish that they were. This would be consistent with Kant’s claim that we often tend to flatter ourselves by falsely attributing a more noble ground to our actions than there actually is. (GMS 4:407) Indeed, since Kant had once said that “the will in accordance with instinct is temperament; the one in accordance with principles is character” one might perhaps conclude that he believed that people without character would simply act from mere instincts rather than from self-made principles.148 However, given Kant’s general definition of impulses, only a relatively small number of human actions are, in a strict sense, to be regarded as instinctive. The bare fact that a great deal, if not most, of our actions have their origin in our temperament (and thus are done from inclination) does not erase the fact that even such inclination based actions are made from maxims – principles created and chosen by ourselves. Again we may recall the demand of the incorporation thesis: unless the agent has freely chosen to incorporate the incentive – may it be an inclination or respect – into the maxim, his choice would not be determined and there would be no activity of his faculty of desire that could stimulate him to act. Accordingly, if taken literally, the rarity-argument of character will immediately affect several related areas of Kant’s thinking.

This query really reveals the need for a complete understanding of the complexity of the conception of character and Denkungsart. Most of all it exposes the need for abandoning the wish to establish the definition of character or the precise relationship of character and other aspects of human nature, like Gesinnung or sensibility in general. The complexity, I will argue, is

147 Anthr. 7:291-2. Besides this explicit statement about rarity of character, there are comments in the lecture notes that declare that there are many people who do not have character. VA [Friedländer] 25:631.
148 Refl. no. 1117 (15: 499).
present through the entire course of Kant’s philosophy and thus it is not wholly correct to say, as is sometimes done, that there is a certain change in Kant’s conception of character (or Gesinnung) that can be located to a specific time or work. Many of the attempts to make sense of Kant’s “problematic idea of character” in order to make it fit with our intuitive idea of human beings and their moral qualities seem to be less fortunate since these suggestions presuppose that Kant, in fact, is dealing with one basic conception of character, which becomes problematic when one wants to apply it in various contexts. In other words, such ways of tackling the subject actually obscure the real intricacy of Kant’s discussion of human character.

In order to show why the suggestion that Kant changed the definition of character at a certain time is an unsuccessful way of tackling the problem, one must take a close look at a central division that Kant made with respect to this matter. In particular, we need to return to the distinction between the intelligible and the sensible character of which the former is said to be our Denkungsart while the latter is our Sinnesart. Both Henry Allison and Kelly Coble argue that there is a significant change of Kant’s conception of intelligible character between the first and the second Critique though they give different accounts of what this change consists in. Both of them, I will argue, overlook some highly important details and thus neglect the full complexity of Kant’s double-character, which, as we will see, needs to be analyzed in the light of Kant’s overall moral anthropology. Rather than presenting a different notion of intelligible character, the latter work adds another dimension to the term that must be spelled out separately. This new notion, I will show, does by no means affect Kant’s Denkungsart-Sinnesart distinction from the Critique of Pure Reason. Moreover, in an attempt to reach a fuller conception of character in Kant’s thinking I will here identify a number of character pairs – one that figures in the Critiques, one in the Anthropology and another one that is present in the Religion. In spite of their identical names and similar qualities, these conceptions are not equivalent and must not be conflated since they are all needed for Kant’s full idea of the human being. By taking all these descriptions into account, we will see that the rarity-claim with respect to character or Denkungsart can coexist with the incorporation thesis as well as with Kant’s general idea of the human being as an essentially free and rational being.

Accordingly, I will argue that all seemingly conflicting claims about character do not necessarily have to be a sign of problematic inconsistencies with respect to Kant’s entire thinking. Still, there are several knotty passages that need to be tackled. I will here expose the most common problems and deadlocks that all Kant exegetes will meet when they try to make sense of the myriad definitions and explanations of character and its role in human life.

149 Cf. e.g. Allison (1990) p.140 and Coble p. 66 note 17.
3.1. The Need for an Internal Foundation

In the presentation of the morally good will in the opening paragraph of *Groundwork* I, Kant makes clear that none of the many things that we consider to be both useful and valuable can be said to be *unconditionally* good since they could very well be used in an evil way. In order to consider even the free and rational will to be *good* we need to ascribe a certain quality to it, or more precisely, an inner constitution [Beschaffenheit]. This constitution – which, according to Kant is a *peculiar* [eigentümliche] one – is called character and it is there, and thus not in a tender heart, in refined skills or in a friendly temperament that the worth of an action can be said to have its origin. (*GMS* 4:393) In spite of the important role that Kant obviously ascribes to this “inner constitution”, the term character is not given any attention in this work. Besides the just mentioned occasion, Kant refers to character one more time and says that it has incomparably higher worth than a good-natured temperament. It is our character that shines forth when one performs the dutiful action *from duty* and not from inclination. (*GMS* 4:399)

Besides these passages, the *Groundwork* seems to focus on the moral nature of the actions rather than on the nature of the person performing them. We learn a lot about how to deliberate about possible actions and we are told that there is a crucial difference between actions that merely *conform to duty* and actions which are also done *for the sake of duty*. Kant talks carefully about the importance of the *maxims* and of a good will but he does not say much about how these factors actually affect choice and action. According to Allison, one may easily be lead to believe that Kant treated actions as merely “free-floating, isolated decisions”.\(^{150}\) Consequently, Allison says, given only the arguments from the *Groundwork*, Kant’s view on moral psychology must be considered “seriously incomplete”.\(^{151}\) What is missing is a deeper and more genuine understanding of the just mentioned “peculiar constitution” of the will. Such a deeper understanding would provide us with what we need in order to see how our will cooperates with the rest of our mental abilities and how this constitution actually influences our actions. Admittedly, in the *Groundwork*, Kant apparently wanted to restrict his analysis to pure philosophy – to metaphysics of morals – that did not borrow anything from the knowledge of man since this would be mere anthropology. As applied philosophy, anthropology is, as we know, *impure*. Yet, it is clear that a deeper portrait of the *good will*, or the way by which maxims actually are created, would be greatly beneficial also for the understanding of the notion of a moral being.

According to Allison, the incomplete view is due to both the *Groundwork* and the second *Critique*. The missing element, Allison says, is provided in

---


\(^{151}\) Ibid.
the Religion by the introduction of the term Gesinnung (disposition). This term, whose role will be further developed in the next chapter, is tightly connected to the notion of character. Indeed, Allison says, we are here faced with the whole underlying “set of intentions, beliefs, interests and so on which collectively constitute that agent’s disposition [Gesinnung] or character.” Without these, he says, none of our choices or maxims could be imputed much less explained. Thus they would be nothing but “arbitrary expressions of a ‘liberty of indifference’ without any ‘sufficient reason’.” In this respect, Kant’s doctrine of Gesinnung, or character, may thus be assigned the systematic purpose of saving the rationality of our decisions.

I totally agree with Allison that a complete picture of the moral person requires a concept of character. Still, I believe that this, after all, is implicit in the Groundwork through the idea of the good will which lies there as a secure foundation for all moral maxims and thus for all decisions that the human being will take. The need for such an internal foundation is also clearly stated in the above mentioned reference to character in the opening of the Groundwork, where the existence of a character or “constitution” of the will was said to be the most important feature of the moral being.

Accordingly, this need is obviously recognized by Kant already in the Groundwork but it is apparently not spelled out as much as it might have deserved. Therefore, if there is a gap in Kant’s presentation, there are reasons for filling it for the benefits of a fuller view of the human being’s entire constitution cannot be overstated. That character fills this role is made clear in the Lectures on Anthropology, where we are told that character is the main thing in human beings and there is a “confluence in their case of everything toward it”. (VA [Friedländer] 25:648) There are several examples of this view in recent secondary literature. For example, Felicitas Munzel argues in her book Kant’s Conception of Moral Character (1999) that character can be seen as the “critical link” that unifying morality, anthropology and reflective judgment. Her study is extensive and, according to her, the first systematic study of Kant’s conception of moral character aimed at laying out the foundations for a genuine understanding of how this important foundation of human agency works. However, today’s scholars seem to have realized the importance of character, presumably as a result of the general revival of Kantian anthropology. Once one has understood the limitations of a study that merely looks at the particular actions (including the maxims that are supposed to explain them), the importance of an idea of character (inner nature or ultimate foundation) within the anthropological framework becomes obvious since, when making inquiries into the human being, one must, of course, study him and his nature, and also make the relevant distinc-

152 Ibid.
153 Ibid. p.137.
tions touched upon in chapter 1, between what belongs to man as sensual, and rational, being respectively.

In the *Anthropology* Kant says:

In the Characteristic one can, without tautology divide what belongs to a human being’s faculty of desire (what is practical) into what is characteristic in a) his natural aptitude or natural predisposition, b) his temperament or sensibility [Sinnesart], and c) his character purely and simply [Charakter schlechthin] or way of thinking [Denkungsart]. The first two predispositions indicate what can be made of the human being; the last (moral) predisposition indicates what he is prepared to make of himself. (*Anthr.* 7:285)

Here we are confronted with one of the sets of double character and, as we see, the most important form of character, that is, character in the most proper sense of the word; what Kant calls Denkungsart. It is this characteristic expressing our nature as rational beings that denotes a higher nature than mere animality. Denkungsart is, for several reasons, treated as a core concept in Kant’s moral anthropology and Kant frequently stresses that without this predisposition man would be incapable of acting from principles and he would thus be neither good, nor evil. Elsewhere, Kant goes even further and says that without character – the ruling principle that alone defines who one is – man would be a contemptible being. (*Refi.* no. 1113 15:496) The frequent use of character and Denkungsart in his moral and anthropological texts indicates that he regarded it as an important aspect of our being. Since it, as Munzel suggests, seems to infiltrate several related issues of his thinking, the recent interest in Kant’s theory of character is understandable.

### 3.1.1. Problematic Claims with Regard to Character

In its most general meaning, character means merely the “essence” or “nature” of something, and, in this sense, Kant often talks about the character of the species, nation, races and of the sexes in order to describe their various characteristics. Although the character of the species is highly relevant for the general understanding of human nature (and I will return to it in chapter 4), the most important conception with respect to morality concerns the character of the person. This one, however, is also the most problematic. From the passage quoted above we learned that while things belonging to our sensual nature indicated what can be made of the human being, our character specified what the human being is able to make of himself. (*Anthr.* 7:285) From this we see that character is the constitution of the will and can be good or evil. It rules the activity of the higher faculties of our mind and this is why people without character are said to be unprincipled and lacking in behavioural consistency. Character thus indicates strength and resoluteness and even a person of truly evil character, like Sulla, is worthy of esteem since strength of soul is an admirable thing. However, in order to deserve the
epithet “greatness of soul” one needs not only strength but also goodness of
soul. (Anthr. 7:293)

If one delves deeper into Kant’s immense discussion of character, espe-
cially in the lecture notes and Reflexionen on anthropology, one will almost
certainly become confused. From Kant’s overall discussion it is clear that
“character” is never innate but always acquired and this is why character –
good as well as evil – is an admirable thing. Sometimes, we are told that all
of us have an innate predisposition [Anlage] to character but that this re-
quires much work and attention in order to develop into what properly can
be called character.154 Although this argument can explain why some people
(if not most of us) lack genuine character, it seems to contradict Kant’s de-
fence of the claim that an evil character is better than no character at all since
this argument says that an evil one can be transformed into a good one while
a missing one never can be attained.155 If this is true, Kant must mean either
that some people actually lack the very predisposition – or germ – from
which a character is developed (this is explicitly denied on several occa-
sions) or that some men have a defect germ that, in spite of much effort,
ever can be developed. However, none of these alternatives can excuse the
absence of character in a human being for not even the lack of a predisposi-
tion to character is said to make it wholly impossible to acquire a character,
but only that it is extraordinarily difficult to reach this state of firmness of
ones principles. (VA [Menschenkunde] 25:1170) Yet, this single occasion in
which Kant seems to allow for the possibility to lack a predisposition [Anla-
ge] to character conflicts with much of what is said in further texts in which
the germ to good is always and necessarily present in everyone. Moreover,
the first claim – about the possibility of a character transformation from evil
to good – is also impossible to bring into conformity with his claim else-
where that an evil character can never can be transformed into a good one
since the true germ is thereby missing and all such germs must be present
from the beginning.156 To make the discussion of character even more puz-
zling one may consider Collin’s notes where it is said that a human being
cannot get another character than the one he has by nature; he may surely
mitigate his evil but in order to be good, there must be a good germ. (VA
[Collins] 25:228)

Still, Kant generally underlines that that there are no evil germs but only
germs to the good.157 Thus the more one reads about the origin of good, bad
or evil character – or any form of character at all – the more confused one
becomes.

157 Päd 9:448. See also Kant’s Lectures on Rational Theology where it is said that evil do not
have a germs or any positive grounds at all for it is merely the negation of the good. (VR
[Pöliz] 28:1078)
3.1.2. Character and Will

Besides these difficulties, which are all due to mutually incompatible claims about character, we will also have to cope with the fact that Kant’s most general definition of character is a “capacity to act according to principles”. If we bring to mind his definition of both will and choice, we see that they are defined in exactly the same way. Consequently, if we take Kant’s plain definition for granted, the rarity that is ascribed to character might thus also hold for the possession of a will as well as for a faculty of choice. Since the ability to act intentionally, to make principles (maxims) for one’s representations and incentives, was said to constitute life, the idea that there could be people without a faculty of desire – whether higher or lower – appears highly unintelligible. Moreover, this obviously conflicts with the incorporation thesis that, as we saw in the previous chapter, tells us that an incentive can only motivate (i.e., affect our faculty of choice) if we have admitted it as a part of our maxim.

The overlapping definitions need to be taken seriously. Since Kant defines character as the ability of a human being to act from maxims or to create fixed rules for his conduct, one immediately sees that character seems to resemble the will, or faculty of choice. Furthermore, in accord with Kant’s further definitions of our faculty of desire, his conception of character also coheres with the definition of “practical reason”. In the Lecture Notes to anthropology one finds several occasions in which Kant explicitly connects the moral quality of one’s character to the moral quality of one’s will. (See e.g. VA [Friedländer] 25:652) Because of this, it seems that Manfred Kuehn has a point when he says that “almost everything that Kant says about character in anthropology can be translated to what he says about will in his moral philosophy”. He adds that character is the appearance of the will and thus our character reflects the moral quality of the will; a good character is a sign of a good will and an evil character reflects an evil will. This, he says, is the essence of the definition of character as a way of thinking – a Denkungsart.

These conclusions also affect the idea of moral goodness. If we look at Kant’s published works we see that moral character seems to be developed from our “predisposition to the good” considered in its highest form (i.e. to personality). We are also told that this predisposition consists in the “susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” and to this, Kant adds, “a power of choice [Willkür] so constituted is a good character”. (Rel. 6:27, 6:48) Combined with the comments from the Groundwork and second Critique, it seems clear that the good character is intimately connected with the good will, either by being the constitution that provides it with its unlimited moral worth or, as some

suggest, the good will and the good character are essentially the same. For example, Friedländer explicitly says that the good character is the good will.\textsuperscript{159} As we saw in chapter 2, both will and character are said to be the capacity to make principles for one’s actions; that is, maxims that can serve as universal laws. Character is also, like the good will, given an incomparable worth and not merely a relative price. What has a price, Kant says, can be replaced by something else as its equivalent; but the lack of character or good will cannot be compensated for.\textsuperscript{160}

Now, what may still seem problematic with respect to the connection between character and will is that Kant obviously allows for the existence of an evil character which, as we have seen, because of its resoluteness with respect to principles, is an admirable thing. Apparently, character can, because of its firm principles, be evil. Yet, Kant says, an absolutely evil will cannot be attributed to a human being for such a will would be considered diabolic. The addition of absolutely evil [schlechthin böser] with regard to the diabolic will is central for this signifies that evil is chosen qua evil, that is, for its own sake. Elsewhere, however, Kant obviously seems to allow for an evil will in the sense of a will that stands under laws of sensibility, and that basically signifies the will of the homo phaenomenon. (See e.g. GMS 4:455) This will is only evil in the sense that it is not absolutely good. As we will see, the evil will in its trivial sense (i.e. as the will of a human being) is more or less the same as an evil propensity, which coexists with our “predispositions to good” and our “germ of good”. At least the latter term has been defined in terms of the good will by Kant himself. (Refl. no. 1426 15:622) Yet, his “good will”, as the germ that, at least makes acting in accordance with the categorical imperative possible, is not an absolutely good will [schlechthin guten Willen] since such a will is never to be found in limited, but only in holy beings.

Nevertheless the very idea of an “evil character”, which only deserves the epithet “character” because of its firm principles, seems to signify something like an absolutely evil will. The difficulty in understanding this is real and imposes several unintelligible conclusions on Kant. I will return to the discussion of evil and its relation to will, character, Gesinnung and predispositions in the next chapter. For the moment I will continue the investigation of the way in which one can understand what Kant might have meant by character. As mentioned above, we need to abandon the idea of one notion of

\textsuperscript{159}VA [Friedländer] 25:648. Literally, the text reads “der gute Charakter wäre der gute Wille” which in the strict sense means “would be”. Kuehn here replies that he takes the subjunctive mood to indicate the indirect speak that is involved here, namely that Friedländer quotes Kant’s saying that the good character was the same as the good will. (Kuhn 2009 p. 19.n.31)

\textsuperscript{160} Thus there is an apparent analogy between the claims in the anthropology lessons and the Groundwork in that Kant, in both works, distinguishes between a market price and a fancy price which are said to be applicable to gifts of nature as talents and our lower mental powers as sensibility (inclinations). Only character and good will have an inner worth, that is, a dignity. (VA [Menschenkunde] 25:1157, GMS 4:434-5, 4:426 and VA [Mrongovius] 25:1385-6).
character and instead focus on the many conceptions that are provided in Kant’s various texts and find out how this feature actually operates within us.

3.2. Kant’s Dual Conception of Character

In order to breach deeper into Kant’s conception of the human being we need to examine the two terms that, as we have seen, are generally said to explain two distinct aspects of the human being, namely Sinnesart and Denkungsart. Although this “pair”, as already mentioned, is actually several pairs, this distinction brings back to mind the general division of the human being as simultaneously an animal and a rational being. Before taking on the various pairs of character I believe that there are things to say about the translation of the terms since it brings with it some difficulties that need to be taken seriously. The literal meaning of Sinnesart is “way of sensing” and though this definition is far from obvious for the modern reader, it is far less tricky as its intelligible counterpart. In short, one may say that Sinnesart is, in different senses, a description of the natural aspects of the empirically conditioned being. I will return to these senses in the upcoming sections.

With respect to Denkungsart, there are several things that need to be considered. Literally, it means “way of thinking” but in contemporary discussion it is often given a more specific sense of a “conduct of thought”, “mental attitude” or a “cast of mind”.\(^{161}\) It cannot be denied that this expansion of the plain reading is highly beneficial for the discussion of Denkungsart in more practical contexts like moral anthropology. Still this attempt to, as Munzel puts it, “capture the technical sense” of the term may sometimes rather lead us astray since there are several instances in which Denkungsart ought to be taken in its literal and, so to speak, trivial sense rather than as an intelligible, or moral, character.\(^{162}\) It is important here to be aware that this trivial use of the term is not only found in Kant’s more theoretical texts. On the contrary, some oft-quoted occurrences of Denkungsart from the *Groundwork*, as well as from the second and the third *Critique*, should be taken in this trivial sense even if some commentators want to see something more in them. I will return to this shortly but before that, I will present the various forms of character dualism that are present in Kant’s writings.

\(^{161}\) Other alternative translations that have been used in literature are “mode of thought” “manner of thinking”, or simply “thought” or “thinking”. For a detailed discussion of this, see Munzel (1999) pp. xv-xviii.

\(^{162}\) In the sense of a merely “way of thinking”, Denkungsart is more or less equivalent with the terms Denkart and Denkwise.
3.2.1. Denkungsart-Sinnensart in the Third Antinomy

Kant’s most “technical” definition of character appears in the resolution of the third antinomy and is the ultimate outcome of his transcendental thinking. As we saw in the previous chapter, the solution of this apparent conflict required that we look upon the world from two different perspectives and thus as subject to two different laws: nature and freedom. We also learned that we must assume that there is a thing in itself, namely a *noumenal* entity that underlies every sensible appearance as its *intelligible* cause. As Kant says, even if we cannot *cognize* any objects as things in themselves, “we at least must be able to *think* them as things in themselves”. (KrV B xxvii) As we have seen, this argument also actualized the idea of the twofold conception of *human beings* as well as the need of explaining how it can be that everything that happens follows from *laws* but must at the same time be regarded as free. In other words, also the thing in the sensible world, the *appearance*, brings with it a faculty that is itself not an object of intuition but through which it can be the cause of appearances. It is because of this intelligible faculty that one can consider the causality of a being from two standpoints: as intelligible in its actions as a *thing in itself*, and as natural in the effects of that action as an appearance – as happening in the word of sense. Now Kant adds that “every effective cause must have a character, i.e., a *law of its causality* without which it would not be a cause at all”. (KrV A539/B567) He continues:

> And then for a subject of the world of sense we would have first an *empirical character*, through which the actions, as appearances, would stand through and through in connection with other appearances in accordance with constant natural laws, from which, as their conditions, they could be derived; and thus, in combination with these other appearances, they would constitute members of a single series of the natural order.

And

> Yet second, one would also have to allow this subject an *intelligible character*, through which it is indeed the cause of those actions as appearances but which does not stand under any conditions of sensibility and is not itself appearance. The first one could call the character of such a thing in appearance, the second its character as a *thing in itself*.

Although this addition of “characters” was made by Kant in order to clarify the cloudy tone of the argument of the two *perspectives, standpoints* or *aspects* of reality, it is not without reason that this idea of “two I” [zwei Ichs] within the same subject was considered to be “the most obscure and incomprehensible problem in Kant’s critical philosophy” already by his contempo-
rary colleagues. Nevertheless, character is here defined plainly as a law of causality, and these laws are, as we have seen, nature or freedom. Thus the essence of this seems to be that intelligible character – or Denkungsart – is the law of freedom operative in the human being, and so expresses our causality as noumenon. On the other hand, our empirical character, or Sinnesart, is to be seen as the appearance of the intelligible and this means that what happens in time has a noumenal origin. Accordingly, we have Kant’s claim that the causality of reason is not to be regarded as a mere concurrence along with other causes, but as complete in itself, even if certain incentives of sensibility will often work against it. It is because of this completeness of the intelligible cause that every action – at the moment it is performed – is entirely our own work. Hence the blame concerns the person’s intelligible character – his free causality of reason that, regardless of all the empirical conditions just mentioned, could have determined the conduct of the person to be other than it is. (KrV A554-5/B582-3)

Here the practical implication of this dualism becomes clear since even if one believes that actions (e.g. a lie) may be determined by natural causes, one nevertheless blames the person for his deed. This blame, however, does not concern those natural causes themselves, like a bad upbringing, his bad company, the wickedness of his natural temperament, his carelessness or thoughtlessness and so on. Rather, one blames simply because one presupposes that all those things that explained how his life actually was constituted could have been entirely set aside since there is another kind of origin which is wholly unconditioned with respect to all previous states. As Kant puts it in the Groundwork, a rational being must regard himself as intelligence and as such as belonging – not to the world of senses – but to the world of understanding. Accordingly, the human being:

has two standpoints from which he can regard himself and cognize laws for the use of his powers and consequently for all his actions; first, insofar as he belongs to the world of sense, under laws of nature (heteronomy); second, as belonging to the intelligible world, under laws which, being independent of nature, are not empirical but grounded merely in reason. (GMS 4:453)

Although Kant does not make use of the term character here, it is clear that his talk of “two standpoints” relies on the same division that was made in the Critique. Both in the Groundwork and in the Critique of Practical Reason, Kant illustrates this possibility of transferring oneself into another realm in which the person, regardless of the entire course of life he has led up to that point, he could have refrained from his unlawful deed since in this perspective he only stands under the causality of freedom. Since everything that happens takes place in the world of sense, we cannot ask why our reason did

---

not determine itself otherwise. Rather, we must ask why it did not determine its appearances otherwise. Still, no real answer can be provided to this query since we simply have the intelligible character we in fact have, and only a different intelligible character would have given another empirical one.

Here, Kant says, we reach the limits of possible knowledge for while we may go as far as to the intelligible cause, we cannot get beyond it. As a matter of fact, even to ask why just these appearances are given to us does not only surpass the capacity of reason but it also surpasses the authority of reason. However, as Kant says, answering these questions is not essential to the task of resolving whether freedom and natural necessity in one and the same action contradict each other. (KrV A557/B585) Yet, since Kant returns to the discussion of freedom and its coexistence with natural determining grounds in the second Critique I will move to this and see if the applications that are made there can take us any further.

3.2.2. Denkungsart-Sinnesart in the Second Critique

In order to explain why one ought to be cautious in the treatment of character I will here add a subsection dealing with the discussion of character dualism in the second Critique. This is, as previously mentioned, due to the suggestion that there are some significant modifications with respect to character that must not be ignored. Since I do not believe that any of these arguments actually affects the understanding of the fundamental distinction, I will here briefly explain why we, in spite of claims of the opposite, ought to take Kant’s outline here as wholly consistent with the discussion in his earlier work.

First of all, it might be mentioned that Kant himself, although he more or less repeats the essential arguments from the third antinomy, does not use the notion of an “intelligible character” [intelligibele Charakter] in the second Critique. When he outlines the dual conception of the human being in order to explain his existence as freely determinable being (KpV 5:98-106), Kant is talking in terms of “intelligible existence”, “causality as noumenon” or of the human being as “belonging to the intelligible world”. The mere fact that Kant does not explicitly speak of our intelligible and free causality in terms of an intelligible character, however, does, not necessarily mean that he has abandoned this conception. Rather, the second Critique’s description of human nature as susceptible to free causality of pure reason is in all relevant respects fully compatible with the general description that Kant presented in the first Critique. In short, he repeats that we must assume that there is an intelligible ground behind everything that takes place in the sensible world, and if we could have a full insight into this ground, we would be able to predict everything that the agent would do with the same certainty as if it

---

were a solar eclipse. Yet, in spite of all such natural determining factors, man is always conscious of himself as a “thing in itself” and, as such, he does not stand under conditions of time. Thus and in this sense, nothing of the past can be the determining ground of his causality as *noumenon*. (*KpV* 5:98)

Now, in response to the claim that the conception of character in this latter text “differs significantly” I would like to say that whether the definition of character differs or not depends solely on the context and not on the work where it is found. When Allison says that “the notion of intelligible character operative in the second *Critique* cannot be equated with that of the first”, and says that “this modification takes the form of a thickening”, he pays attention to the lines where Kant says that our free intelligible nature consists in the adoption of unchangeable principles. (*KpV* 5:100) This, he says, indicates that Kant did no longer limit the conception of “intelligible character” to be a mere “law of causality” but it is now to be considered as a character also in the “anthropological sense” of a “way of thinking” (Denkungsart).”165 From this it follows that Allison equates this Denkungsart with Gesinnung.

In order not to overstate Allison’s argument, it should be made clear that he indeed says that this change is best understood as a *modification* rather than as an *abandonment* of the earlier one. Nevertheless, although it is true that Kant occasionally uses the term Denkungsart in relation to some claims about moral character in the second *Critique*, I do not think that this should be seen even as a *modification* of the first *Critique*’s conception of intelligible character. Moreover, as we will see, this is not primarily Kant’s *anthropological* notion of intelligible character but it rather resembles the idea of intelligible character that is most notable in the *Religion*.166 In short, the characterization of the intelligible ground at the end of Kant’s review of the arguments of the third antinomy, in terms of a “constitution of the will” that is not “necessary but is instead the consequence of the evil and unchangeable principles freely adopted” (*KpV* 5:100), is actually an additional conception which, together with the other one (a law of one’s causality), both figure in the same discussion. Accordingly, Kant’s use of Denkungsart in the additional senses of a general “moral nature” or a “consequent way of thinking” ought to be treated as a *completely different* notion of Denkungsart that does not indicate a “change” at all since this usage of the term has always been parallel and is to be found in various texts prior to the first *Critique*.167 Moreover, in the very description of character as a “practical, consequent way of thinking according to unchangeable maxims” (*KpV* 5:152) it definitely seems as if “Denkungsart” *as such* is used in its trivial sense, as a way of

---

166 See 3.2.4. below.
167 See the lecture notes to ethics and anthropology, several early *Reflections* as well as the *Beob.* and the *BB.*

110
"thinking" or "reasoning" about maxims and of applying one’s thinking in a resolute (i.e. non-capricious) way. Indeed, as Kant says, in its proper sense, character consists simply in *originality in the way of thinking* [Originalität der Denkungsart]. This is why a character is a certain "way of thinking" where one practices one’s reason in a correct way by thinking for oneself rather than merely imitating for the imitator has no character. This means that "originality" here signifies that one is oneself the *origin* of this "way of thinking", and that one does not, like the epigone, merely adopt other’s "ways of thinking" however original they might be. As Kant says, he who has character derives his conduct from a source that he has opened by himself. (*Anthr. 7:293*)

To Allison’s defense it might be mentioned that his argument for the importance of observing this "significant change" is a part of his argument for the importance of not neglecting the necessity of a moral ground in the human nature: *Gesinnung*. In this, he seems to equate Gesinnung and intelligible character by saying that they express an agent’s “fundamental maxim with respect to the moral law".168 This, in turn, was the only way of escaping the conclusion that Kant regarded actions as “free-floating, isolated decisions”. While I totally agree that Gesinnung plays a central role for Kant’s idea of moral agency, I do not see how this argument requires a general change of the understanding of intelligible character in the second Critique’s outline of human freedom and its possible coexistence with nature. Rather, the notion of Gesinnung raises another dimension of Kant’s thinking and with this it presents an additional idea of our intelligible nature. Again, since Gesinnung and its role for human agency will be the topic of the next chapter I will leave this question for the moment.

The next argument that I will consider here is Coble’s argument for recognizing a “sense-change” of Kant’s usage of Denkungsart between the first and the second Critique. Because of Kant’s abstract theorizing with respect to human nature, it is tempting to find ways of reading him that do not get lost in a non-concrete noumenal spectrum. Coble thus suggests – or indeed proposes a decree – that one would be better off by placing a “strict ban on noumenal talk” with regard to human character, and adds that it is surely possible to discuss Kant’s idea of intelligible and empirical characters in a metaphysically neutral way.169 Although this “metaphysical neutrality” with respect to character is said to hold for Kant’s use of empirical and intelligible character in the *Religion*, the very possibility of this change in understanding character is explicitly sanctioned by Kant himself because of the change that is made between the Critiques. According to Coble, from the *Critique of Practical Reason* and onwards, the intelligible character (Denkungsart) exactly corresponds to the definition of “empirical character” given in the *Cri-

169 Coble p. 66 note 17.
Coble argues for this by drawing attention to the discussion of empirical character given at A549-50/B577-8 and compares it with the discussion at 5:99. It is important, he says, that in the first Critique, the empirical character is “still a character of an agent's rational choice (understood in light of all those factors that make the choice inevitable).”

While Coble does neither explicitly point out, nor cite, the definitions that he refers to in these respective works, he at least seems to pay attention to the fact that in the first Critique, empirical character is given the sense of “a certain causality of [the human being’s] reason that has an effect in appearances”. From these effects, Kant says, one can “derive the rational grounds and the actions themselves according to their kind and degree, and estimate the subjective principles of his power of choice.” (KrV A549/B577) I agree that this passage, if taken in isolation, seems to say that the empirical character is a causality of reason that is expressed in the realm of sense (in experience). Still, we must not ignore the fact that Kant, in the same work, also says that in its empirical character, the being – as appearance – would be “subject to the causal condition, in accordance with all the laws of determination; and to that extent it would be nothing but a part of the world of sense, whose effects, like those of any other appearance would flow inevitably from nature.” (KpV A541/B569, emphasis added)

If we return to the second Critique, we see that if the effects result from nature the determining ground of the action belongs to past time and is thus no longer within the agent’s control. Yet, as we have learned, if we take such determinations in time as determination – not only of appearances – but of things in themselves, freedom cannot be saved. (KrV 5:101) Consequently, there is a need for a character that does not stand under such temporal conditions, and the only one that fulfills this criterion is an intelligible character. Yet again, while Kant does not explicitly use the terminology of an “intelligible character” in the second Critique it is clear that he distinguishes between the determinations of a thing “which stands under conditions of time” and the same subject which is “on the other side conscious of himself as a thing in itself” and, as such, independent of everything antecedent to the determination. We find terms such as “intelligible existence”, “intelligible consciousness of one’s existence”. (KpV 5:98)

It seems, therefore, that we cannot skip the transcendental perspective or the “noumenal talk”, with respect to human character as explanation of human agency since the understanding of our rational nature and the possibility of moral knowledge inevitably requires that we recognize our connection to the intelligible character as a law of the causality of freedom. In other words, morality always entails our recognition of ourselves as noumenon.

---

170 Ibid.
171 Ibid.
3.2.3. Denkungsart-Sinnesart in Anthropology

In addition to the former character-dualism that involved a highly metaphysical tone, Kant also makes a relatively traditional anthropological distinction between the different aspects of our human nature. While the metaphysical distinction from the Critiques separates the mere appearance of the human being (the *homo phaenomenon*) from the purely transcendental idea of a *homo noumenon*, this later division aims to explain the impure nature of the living man, as both rational and sensuous. Although the difference from the former dualism might not be immediately obvious, I will argue that the discussion will benefit from a separation of these perspectives.

The essence of the anthropological division was made explicit in the passage quoted in the beginning of this chapter in which we were told that while our Sinnesart made up our temperament, Denkungsart or *Charakter schlechthin* was said to explain our ability to be our own masters. (*Anthr.* 7:285) If we connect these ideas to Kant’s famous explanation of practical anthropology, as the idea of a number of “subjective conditions” in human nature that either hindered or helped people in fulfilling the laws of the metaphysics of morals, then we will discover some important things.

First, we will see that according to Kant, the common expression that someone has “this or that character” only refers to our empirical nature or temperament, and this can only tell us what we can expect of a human being with regard to his behavior. In the more philosophical sense, we need only to say that he or she has character and this character can only be one or nothing at all. By assigning character to a human being we at the same time ascribe to him a property of his will by which he binds himself to definite practical principles that he has prescribed to himself irrevocably by his own reason. (*Anthr.* 7:292) As was briefly hinted in section 3.1., the very claim that “someone has a character” is to say much of the person in question, indeed it is to praise him. (*Anthr.* 7:291-2) Kant’s conception of character in the proper, or genuine [eigentliche], sense thus explicitly separates it from all empirical aspects of man and, in this sense, it is to be understood as a “character of freedom.” (*VA* [Mrongovius] 25:1384)

Accordingly, when Kant identifies *Charakter schlechthin* or character *simpliciter* with “Denkungsart” one immediately sees that this notion cannot be a description of human nature in general but rather of a very specific quality of this nature. If someone has a character (simpliciter) it means that he or she has acquired a set of binding and definite practical principles, and these principles must have their origin in reason. A central feature of these principles is that they are stable and this means that the person does not “fly off hither and yon, like a swarm of gnats” or, as Kant puts it elsewhere, is not like a piece of soft wax. (*Anthr.* 7:292, *VA* [Friedländer] 25:632) In this sense, Kant says that character is rare and this is the reason why a character, as we have seen, is an admirable thing also when it is evil.
Still, it is difficult to accept that the rarity-claim can be said to hold for our rational nature in general since the main difference between Sinnesart and Denkungsart is that while the former explains the distinguishing mark of the human being as sensuous or natural being, the second is the distinguishing mark of the human being as a rational being endowed with freedom. As we have seen, all human beings are, in virtue of their intelligible character in the sense of the Critiques, subject to the laws of freedom. Accordingly, the general anthropological quality of the human being as a “living being endowed with reason”, must, like the intelligible character as law of causality of freedom, thus be regarded as a universal human characteristic. Here it may be noteworthy that Kant, when he explains the existence of character in the Anthropology, says that this is something that can be found in the Denkungsart of the person and this character has always been acquired. (Anthr. 7:294) Accordingly, while the mere Denkungsart exists – and must exist – in every rational being as the capacity by which we are aware of our existence as free beings, what is rare is the character in this faculty, that is, a fixed and resolute nature of the way that this “way of thinking” (Denkungsart) thinks or reasons about matter.

Moreover, it should be made clear that although the anthropological definition in several respects may be said to be similar to the idea of Denkungsart in the first Critique (as the awareness of our existence as rational and free beings), it cannot be equated with it. There are several reasons for this. First, an important difference between this anthropological interpretation and the former one is that Sinnesart, in the anthropological sense of our temperament, is not the appearance of our Denkungsart taken in its anthropological meaning. For that reason, it is also not the case that the intelligible character as rational nature or “fixed way of thinking”, causes or creates our temperament. Rather, as Kant’s general idea of human nature and practical anthropology shows, we must strengthen our rational self in order to resist the force of inclinations that flow from our temperament. These two anthropological parts of our nature exist so to speak “side by side” and can thus be seen as cooperating or conflicting motivational grounds rather than the two simultaneously present aspects (or standpoints) of the human person. Again we are faced with Kant’s now well-known definition of practical anthropology as the study of the subjective features of human nature that either hindered or helped people in fulfilling the laws of reason.

A second reason for not taking the two characters in the anthropological sense as just another dualism that is equivalent to the former can be found in Kant’s very idea of anthropology. Since anthropology – the study of human nature and its capacities – concerns human being as appearances, one sees that both temperament and rational nature belongs to our empirical character in the sense given in the Critiques. As we have seen, also a being in the world recognizes his connection to another realm, namely that he also has an
intelligible nature in the metaphysical sense. This insight of a second nature is, as we know, recognized even by the villain or scoundrel who, in spite of his many temptations, acknowledges that he does not have to follow them. (GMS 4:454f.) For this reason, it is also important to understand that temperament – Sinnesart or empirical character in the anthropological sense – is, with regard to its constitution, a separate entity that is independent of the rational part of our nature. Still of course, as a moral being, one ought to subject the demands of this empirical part to our rational nature. This relationship is also clearly shown by the fact that a good Denkungsart can coexist with an “evil heart” and an “evil mind”. In a similar manner, a good or tender heart can by no means be sufficient for attributing a good character to anyone. For again, as we were told, without a rule of our will, that conducted the use of our natural gifts, they could be truly harmful or evil.

This reveals the need of education in all its forms, from mere discipline to cultivation, civilization and, finally, moralization. Still, Kant makes clear that the firmness and persistence that is required for a Character schlechthin cannot be acquired gradually but only through an “explosion” that will be the beginning of a new epoch in the person’s life. I will return to the idea of this notable shift, or revolution, in the next section as well as in chapter 4 since this is a central part of Kant’s way of dealing with evil in the Religion.

To further strengthen the uniqueness of such an inner foundation it can be said that, however unforgettable this radical move might have been, not even mere acquisition is enough for actually having a character in this sense. The revolution is only a first step towards the development of a true character, and Kant adds that very few will have performed this revolution before the age of thirty and even fewer will have established any notable firmness by the age of forty. Accordingly, it is no wonder that character, in this sense, is an admirable thing.

However, we will see that Kant cannot dismiss the importance of a gradual reform of our temperament for the progress towards moral goodness. In order to see this, we need to introduce some additional attributes to the moral being and its various characteristics.

3.2.4. Denkungsart-Sinnesart in the Religion

A third version of the Sinnesart-Denkungsart division appears in the Religion. What makes this conception especially interesting is that it seems to encompass essential features from both the metaphysical and the anthropological meaning without being identical to any of them. This text also puts a

---

172 Anthr. 7:294. This resembles what Kant said already in his discussion of Basedow’s school in which we are told that for education we need “not a slow reform, but a swift revolution” and this need also hold for the teachers. (See Aufsätze, das Philanthropin betreffend 2:449 from 1776-7).
new dimension to the empirical character that, in contrast to Sinnesart in the sense of a mere temperament, is also susceptible to moral reasoning and maxim-making. Moreover, the Religion clearly exposes the need for sorting out two different notions of Denkungsart one of which must be found in everyone.

The Religion is, in several respects, a peculiar work that combines various lines in Kant’s overall thinking. However, what readers of today generally associate with this text is Kant’s theory of “radical evil”. This evil expresses a defect of human nature that must be ascribed to every human being as an inescapable “character of his species”. Kant identifies the ground of evil in a first, intelligible choice of an ultimate and universal ground that affects all maxims that the human being will ever create. Since Kant here adopts the rigorist understanding according to which one must be either good or evil (no intermediate or neutral position is possible) we see that all human beings must be considered evil with respect to this first ground.

While there is much to be said about the details of this ground, and especially the choice of it, I will in this section only isolate the forms of intelligible character that are implicit in Kant’s reasoning. Since the first ground that makes up the “character of the species” is a ground for making maxims, it seems to fit the general definition of a Denkungsart and, in this sense, as a first ground – or, as it is also called, as Gesinnung, it must be attributed to every human being. This need to assign such a quality to every human being is further supported by the incorporation thesis that, as we were told, required that every incentive had to be admitted into a maxim in order to be an effective motivational ground of our faculty of choice.

In addition to this general human character as ultimately good or evil, we also have a general human capacity to use this rationality and this can be expressed in various degrees at the individual level. As we saw in chapter 1, this capacity is expressed through our predispositions to humanity and this predisposition develops, for better or for worse, in parallel with the progress of our social skills. Again, we may recall the demand of incorporation and maxim-making for the possibility of determining choice, which means that also in our most general decisions we need to exercise our capacity to make principles. Moreover, as we have seen in relation to the previous definition of intelligible character as the ability to transfer oneself to the intelligible world, this definition is highly relevant also in this context. In the Religion, Kant once again highlights the difference between morality [Moralität] and mere legality [Legalität]: that although something can indeed be in accordance with duty (and thus be legally good or good according to the letter) it is not also necessarily from duty, which is required in order to be also good in the moral sense or from spirit. Now, what is interesting with respect to our

173 Kant explicitly defines the corrupted ground (the propensity to evil), as a perverted [verkehrter] Denkungsart. (Rel. 6:48)
topic here is that Kant calls the former \textit{virtus phaenomenon} or “goodness according to empirical character” that consists in “abiding maxim of lawful actions”, no matter whence one draws the incentives that the power of choice needs for such actions. This virtuous nature comes about when the agent simply has made acting in accordance with duty a \textit{habit} and this is acquired little by little. Most importantly, Kant says that \textit{not a slightest change of heart is necessary for this} but only a new idea of morality and the insight that the laws of it are binding. (Rel. 6:47)

It is clear that also this capacity of the human being, though defined as a mere “empirical character” (which also is said to place obstacles to the intelligible one) unquestionably involves the \textit{ability to act from maxims}. Moreover, these maxims also undeniably involve a certain \textit{firmness} of these principles. Thus I find it legitimate to conclude that an agent with such a strengthened version of empirical character does \textit{not} fly around like the swarm of gnats as the \textit{Anthropology} so illustratively put it. Accordingly, the Religion’s idea of empirical character (or Sinnesart) virtually coheres with the definition of Denkungsart given in the \textit{Anthropology} since in this sense, the only thing that matters is that the principles are firm – not that they are good. This person seems to be, what Kant calls, “a human being of good morals” (\textit{bene moratus}) and while such persons – because of their still impure nature – do \textit{not always} act from the pure principle of duty, nevertheless they never act contrary to the law. (Rel. 6:30)

Still, since the ultimate object of the moral law is the \textit{highest good}, human beings ought to strive to attain it. For this, however, mere legality is not enough, and human beings thus ought to be good also in the moral sense, that is, to be good according to intelligible character (\textit{virtus noumenon}). In order to become what Kant calls “a morally good human being” (\textit{moraliter bonus}), the person must always incorporate the law as the supreme incentive. Now, while “goodness according to empirical character” did not require any radical alterations of our nature (a gradual reform of one’s moral thinking was enough), true morality requires a revolution of one’s intelligible character from which one can, so to speak, perceive a rebirth or the creation of a new man. (Rel 6:47, 6:36; Anthr. 7:294)

Finally, we thus seem to have found something that must be considered as incredibly rare, namely the idea of a disposition \textit{totally reformed in its very ground} through a revolution. While “the human being of good morals” has indeed – through a \textit{gradual reformation} – improved his mores by a refinement of his Sinnesart, he has nevertheless left the \textit{universal root undisturbed}. Vices will thus be a constant threat and, in this sense, this human being still has the traces of evil in his fundamental ground. Yet, even this person has, and must have, a foundation in the sense of a “way of thinking” [Denkungsart]. This ground is indeed still to be regarded perverted but it is nevertheless a ground that is susceptible to refinement. In contrast, “the morally good human being” has not restricted his endeavour to the improvement
of mores but he has also performed the transformation of his “way of thinking” [Denkungsart] and has, through this ground-breaking deed, established a character. This character is a moral character in its purest sense and is not a mere character of the species but of the person. It must also be considered morally good. Yet, as will be developed in the next chapter (especially 4.2.3.) it is unlikely that any human being will ever be able to do this since the being that comes about from this move rather expresses “the ideal of moral perfection” or “the prototype of moral disposition in its entire purity”. (Rel. 6:61)

3.3 Summary: How to Understand Kantian Character

This chapter aimed to outline the basics of Kant’s theory of character and foremost to identify and resolve some of the difficulties that are generally associated with this concept. Some of these considerations also aimed to prepare for the upcoming chapters, especially those dealing with the idea of evil. Moreover, the conclusions we have reached in this chapter will also be useful when we latter will return to Kant’s idea of moral progress.

It is clear, as Allison and others have claimed, that character adds another dimension to Kant’s idea of the moral person by attributing a stable foundation that defines the person. At the same time, Kant’s discussion of character, Denkungsart, Gesinnung, “constitution of will” or whatever we like to call it, has caused interpreters much confusion. However, as I have argued, many of the problems disappear once we can accept that character is used in different ways in different contexts and when we have also learned the conception applicable in each context. As I have argued, the bare fact that character and Denkungsart are used in more than one sense is, by itself, not a problem. Interpretative problems with regard to character are due to the neglect of this diversity. Accordingly, my main intention here was to explain the importance of noticing this diversity but also to relate these forms of character to Kant’s various discussions of the human being and its conduct.

One of the most notable problems that one needs to deal with in discussions of Kant’s idea of character is the argument of rarity of character or Denkungsart. Without the extended understanding of the term one will immediately find oneself faced with conflicting claims that cannot be brought into conformity with Kant’s idea of the moral human being. Fortunately, as we have seen, we do not have to resign ourselves to them since it is possible to make sense of most of his arguments as soon as one has realized that the rarity-argument only holds for a very specific sense of character.

With regard to the dual conception of character as consisting in an empirical and an intelligible part (i.e. in a Sinnesart and a Denkungsart), I have tried to show that also this character-duality is subject to the problem of complexity as well as of diversity. None of these dualisms can be said to
express the correct or accurate idea of character but they must rather be seen as complementary descriptions that are all needed in order to capture the full picture of human nature.\textsuperscript{174} As an attempt to present a schematic overview of the relationship between the many notions of character I would like to say the following: in the third antinomy (as well as in the second Critique’s review of this outline) we are told that we must assume the existence of an intelligible character (Denkungsart) to which our empirical character (Sinnesart) corresponds as its appearance. The fact that Kant uses slightly different terms in these discussions should not be seen as a sign of a change of meaning but rather that the discussion in the latter work is more immediately concerned with moral aspects of the use of freedom than the outline in the first. Everyone has an intelligible character (Denkungsart) in this sense.

The two notions of character that we meet in the Anthropology and which also bear the names of intelligible versus sensible character (Denkungsart and Sinnesart) need, however, to be analyzed in two steps. As purely anthropological and thus descriptive aspects of the human being’s constitution, both of them must be seen as belonging to the empirical sphere and they must thus be considered in relation to our nature as appearances. Still, while the fact that our Sinnesart, as identified with temperament (and, thus a source of feelings, needs and various desires), trivially belongs solely to nature, we have seen that our Denkungsart, as our rational nature, must be somehow connected to the intelligible realm in order to receive the idea of freedom. The reason for this is simply that as a being endowed with reason, man is necessarily aware of a moral law since this law, as we have seen, is present as a fact of reason. This law, in turn, is the ratio cognoscendi of freedom, that is, the way by which we become conscious of our connection to the intelligible world and of a higher form of causality that is the causality of freedom. In this sense, the Denkungsart, or intelligible character in the anthropological sense, encompasses the essence of the Denkungsart in the

\textsuperscript{174} It has been claimed that the anthropological distinction between our two characters is a comparatively late addition to Kant’s anthropology that did not appear until the later lectures. In the editorial preface to the Vorlesungen über Anthropologie we are told that while Denkungsart is a common term that can be found already in the first set of notes (1772) it is not until Pilau’s remarks from the semester of 1777/1778 that we find Denkungsart and Sinnesart discussed together and contrasted with each other. Because of this, Denkungsart is said to be an “older” notion. (see Editor’s preface AA 25:cx and for Pilau’s comments, see 25:821). However, Brandt and Stark here make a mistake when they say that this relation is observable also within Kant’s published works by pointing out that “Denkungsart” appears already in volume 1 of the Akademy edition while the first occurrence of “Sinnesart” is found in volume 3. (ibid) This comment neglects the fact that Denkungsart, as previously mentioned, is often used in the trivial sense – as a simple way of thinking – and this is without doubt the case for the usage in volume 1. Still, one only needs to consult the Reflexionen in order to see that Kant used the term Sinnesart for the empirical aspect of the human being (Ref. no. 904 15:396 dated to the 1770s). In addition it might be mentioned that the entry “Sinnesart” is not to be found in the 1793 edition of Adelungs Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart while Denkungsart is defined in the non-trivial sense as a way – not of thinking – but of acting.
metaphysical sense as the law of free causality. Still, not even in this sense, as a capacity of acting freely (independently of natural impulses) or making maxims, Denkungsart is to be considered as rare. The fact that Kant often adds some qualifications of the character, saying that it can be better or worse, strong or weak depending on the person’s own effort and his constitution in general, indicates that there is at least a kind of “general character” present in us although not all human beings express this character in their actions.

Finally, Kant is also elaborating with an idea of a moral character from which all moral worth of a person is said to have its ground. This kind of character, which is generally identified with the constitution of one’s will, is a continuing theme in Kant’s moral discussion. While indeed present in the Groundwork (though hidden under the general talk about the good will) it is spelled out in length in the Religion. Here we are once again confronted with the now familiar division between Sinnesart and Denkungsart but now we are told that we may, in fact, be considered good with respect to either of them. As we have seen, human beings can be virtuous by strengthening their empirical characters, and thereby, constantly resist vices simply because of the firm and moral constitution that now dominates their souls. That person has a virtus phaenomenon. In contrast, the truly moral being, who possesses virtus noumenon, has undergone a change of heart and is no longer bothered with temptations. Yet, in spite of the still inferior nature of the mere virtus phaenomenon as compared to the virtus noumenon it is clear that also the former one has a character. Still, because of the resoluteness of the person possessing virtus phaenomenon, and his or her commitment to the good principle as well as his compliance with virtue, his actions cannot be reduced to mere instinctive behavior. Accordingly, also the mere virtus phaenomenon must be said to have a character, that is, a Denkungsart, though perhaps not in the most admirable sense.

Before ending this discussion I would only like to add that if the demands of the incorporation thesis are to be taken as a requisite for human actions (which seems to be required for Kant’s idea of freedom) then all actions that are not done from mere instinct are done from maxims. Again, as we saw in chapter 2, even lower faculties of desire can make maxims but not maxims that can hold as practical laws. Yet, not all human beings have the firmness of will required for having a character in the sense that is admirable and rare. They will, indeed, make principles but they may often be influenced by others and their maxims may, therefore, differ from day to day. They may thus fly like gnats and be easily manipulated or simply susceptible to temptations. Such beings are not admirable but as long as they do not act contrary to duty they are not to be regarded as objects of contempt either. Yet, as the next chapter will show, they are all evil because of a universal ground within.
PART II: Good and Evil
Having learned about the motivational process and how our will and character operates, we may now be prepared to take on one of the more controversial and provocative of Kant’s statements concerning human nature. The *Religion within the Limits of Mere Reason* begins by stating with abundant clarity that “the world lieth in evil” and that this has been known from the beginning of history. Once again Kant refers to the hypothetical idea of a paradisiac Golden Age, but as we all know too well, we are no longer there and as fallen beings, we need to cope with the fact that evil is the prevailing principle in the present world. Suddenly, our original happiness and innocence disappeared as if they had been a dream, and surrounded by evil – physical as well as moral – we now seem to merely passing time until the last day. Surely, Kant says, in modern time several heroic attempts have been made in order to alter this idea, but it is clear that experience will not support any of them. This is why he calls the first part of the work “On the radical evil in human nature” and concludes that evil must be ascribed to every human being and that this evil can thus be said to express the character of the human species.

It is no wonder that the *Religion* has inspired philosophers to break through the pessimistic surface in order to find out why Kant found himself forced to make these gloomy claims about the human soul.\(^{175}\) The material available is vast but, as Lawrence Pasternack points out, most studies focus on part I (and occasionally II) at the expense of part III and IV. The fact that very few have studied the *Religion* as a whole exaggerates the moral pessimism that is the focus of the first part.\(^{176}\) Pasternack has a point here and it is true that Kant, in the latter parts, shifts focus from evil and turns towards moral regeneration by outlining his plan for the creation of a kingdom of God on earth. Still, I believe that the imbalance is understandable since Kant’s idea of radical evil deepens the discussion of moral responsibility, maxim-making and freedom that was not explicitly spelled out in the *Groundwork*. Since my own study in this chapter seeks to understand the


\(^{176}\) Pasternack p. 1.
first ground of evil and how such a ground can interact with our anthropological nature in general, it will perhaps partly contribute to this disproportion.

However, far from every reader of the Religion is satisfied with Kant’s defence of the, indeed, quite awkward claims that are made in this work. Besides the universality claim that we are all essentially sinful creatures and that “there is no one righteous”, Kant also argues for a strict rigorism concerning human morality according to which a human being must be either good or evil and thus never indifferent or a mixture of both. Kant returns to this idea in the Anthropology where he says that without this rigoristic claim, the species would have no character. Yet again, Kant declares that any “being endowed with the power of practical reason and consciousness of freedom of his power of choice” must always, even in the “midst of the darkest representations” regard himself as subject to a law of duty and thereby immediately feel respect for this law. (Anthr 7:324) Now, in the light of this, and with Kant’s acceptance of the “ought-implies-can-principle” in mind, the central question then is: Why do rational and morally predisposed beings still regularly willingly adopt unlawful principles?

4.1. Kantian Evil

Before we enter into Kant’s idea of man as tainted with evil we must not ignore the fact that this evil is neither ascribed to our entire nature, nor is it said to affect our entire empirical selves. As we saw in chapter 1, and as Kant states immediately before his analysis of the propensity to evil, man is composed of three predispositions [Anlage] and all these (i.e. animality, humanity and personality) are to be considered predispositions to good. Already in a reflection from the mid-1770s, Kant said that if there were not also a germ of good (a generally good will) inside the naturally evil human being, there would be no hope for bettering. (Refl. no. 1426, 15:622). The notion of a germ of good [Keim des Guten] reappears in the Religion and even if it is not further defined or explained, it is said to be an eternal quality of the human soul that can never be extirpated or corrupted. (Rel. 6:45) In the late text The Conflict of the Faculties, Kant once again made clear that the human being is compounded [amalgamiert] of both good and evil. (SF 7:55)

Kant obviously needed these factors in order to account for his idea of human beings as rational and susceptible to moral laws, as well as for his theory of a possible moral conversion. Still, because of the “multitude of woeful examples” that we see around us every day, Kant apparently found himself forced to declare that evil has taken root in the human soul. Because of this obvious fall of the originally good being into evil, Kant could not accept the naïve belief that moral goodness would be reached if we simply
concentrated on the development of our good germs while ignoring the active enemy within. This mistake, which Kant here ascribe to the Stoics, is due to the misguided belief that the cause of evil could be located in our inclinations, and that this threat could simply be fought by means of discipline and a strengthening of the human soul in order not to be affected by them. The Stoics, as Kant put it, thereby ignored the fact that this enemy – the evil in our nature – actually was to be sought in the highest faculty of the human soul, that is, in the our reason – the faculty that ultimately determines which incentives we will eventually act upon. (Rel. 6:57) The aim of the Religion might thus be said to be the identification of the true cause of this corruption – a corruption that obviously affected our highest part – and in so doing, find out how to reclaim our original germ of good in order to perform the “revolution”, “rebirth” or “change of heart” that is required for our transformation into moral beings.

It is, however, important to be aware that Kant did not say that human beings want evil, or chooses evil qua evil, as the ground for their action since this, as we have seen, would be diabolic. Again, such an “absolutely evil will” is not applicable to the human being. (Rel. 6:35; 6:37) In view of that, it is clear that we do not have any inclinations to evil and “evil” cannot, therefore, be understood as maliciousness or cruelty. Accordingly, as hinted above, we should not be misled by Kant’s strong vocabulary concerning the nature of human evil. By radical Kant simply refers to a quality of our maxim-making faculty that is rooted in the very depth of human nature. Moreover, it is also important to understand that by evil, Kant here refers to any action that is not done from duty, including those actions that are in accordance with it. To say that a human being is evil simply means that he is well aware of the moral law, but frequently acts contrary to it, or from other incentives than from respect for this law. (Rel. 6:32) As Kant puts it:

whenever incentives other than the law itself (e.g. ambition, self-love in general yes, even a kindly instinct such as sympathy) are needed in order to determine the power of choice to lawful actions, it is purely accidental that these actions agree with the law, for the incentives might equally well incite its violation. The maxim, by the goodness of which all the moral worth of the person must be assessed, is therefore still contrary to law, and the human being, despite all his good action is, nevertheless evil. (6:30-1)

Taken in isolation, this statement, as well as some of Kant’s further comments on empirical psychology, may easily lead to the belief that evil actions actually are a result of our inclinations since inclinations are exactly those subjective desires which serve as the determining ground of all non-moral maxims. Furthermore, in an oft-quoted passage from the Groundwork, Kant plainly claims that every rational being would wish to be completely free from these disturbing and morally worthless features. (GMS 4:428) He frequently claims that inclinations – because of their arbitrary and capricious
nature – must be silenced and suppressed since they differ completely from the good principle, that is, the principle revealed to us from the sense of duty. Again, as we saw in the passage quoted above, Kant undeniably says that the reason why an action is considered evil is that a non-moral incentive is allowed to determine our choice, and among these incentives Kant explicitly counts “kindly instincts such as sympathy”. Yet, in spite of all this, it is important to bear in mind (as we saw in chapter 2) that inclinations as such are good and ought to be regarded as a natural and essential aspect of human nature. Evil, in its proper sense, must here be understood as moral evil and hence not simply as the result of an innate limitation of human nature.177

Again we can recall Kant’s criticism of the Stoics who ignored the fact that the enemy was not to be found in anything external but rather in our own rational nature. Recalling again the incorporation thesis, it becomes clear that the inclinations themselves – though caused by external influences – can motivate if (and only if) we have admitted them as decisive parts of our principles. In doing so, we accept the incentive and no one except the agent himself can be seen as the author of it. Therefore, if the maxim turns out to be based on an inclination, then it depends solely on a free, and thus imputable, choice of the human being.

However, once again we must recall Kant’s definition of maxims as the motivational and explanatory principles behind action and bear in mind that the incentives, including inclinations, do not constitute the entire maxim but only its matter. What really is of importance from the moral point of view is the form of this subjective principle, namely which of the incentives that we subordinate to the other.178 As Kant makes clear, an agent may indeed incorporate the law (i.e. he acknowledges its moral significance) but simultaneously includes supplementary incentives that promise happiness. Expressed in Kantian words, an evil human being simply “reverses the moral order” and makes the law dependent on a non-moral drive. In all such cases,

177 Rel. 6:43. This may be compared to what is said in Collins note on anthropology (Winter 1772/73) where we are told: “The question is: where do we locate the source of ill in the human being? We find it in the human being’s animality. In some human beings there are such strong incentives that it is hard for his intelligence to discipline them. And the differences between human beings appear to rest more on their animality than on their spiritual nature.” (VA [Collins] 25:15)

Animality is, as we have seen, expressed by our purely non-rational part that is governed by instincts and inclinations. Two things might be said here: first, the anthropology note predates the Religion by 20 years so Kant may well have changed his mind. Second, in this note, Kant is not using the term evil [Böse] but ill [Übel]. In his moral works Kant often distinguishes between Böse and Übel and treats the former as objective and rational while the latter is a subjective feeling of displeasure or pain which has no moral connotation at all. The positive correspondents are here good [Gut] and well [Wohl] (see e.g. KpV 5:60, cf. also note 115 above). However, the quote obviously seems to say that we need to discipline our incentives in order to prevent ills.

178 Recall here the difference between formal and material principles that was discussed in chapter 2.
the moral law alone does not suffice to motivate us, and this exposes the fact that we are corrupted in the very ground of the adoption of the maxims. Thus, one can say:

in order, then, to call a human being evil, it must be possible to infer a priori from a number of consciously evil actions, or even from a single one, an underlying evil maxim, and, from this, the presence in the subject of a common ground, itself a maxim, of all particular moral evil maxims. (6:20)

Obviously, one single action is enough to account for an evil ground within, that is, something capable of making the otherwise good human being deviate from his original vocation. Accordingly, we have to assume that there is a specific quality of the human being that, in some way or another precedes and affects every single choice and thus every maxim that the agent will ever make for himself. Now, although the mere idea of a corrupted principle in a human being’s mind may be a fairly plausible explanation of his evil actions, a genuine theory of the origin of evil cannot be restricted to particular human beings since such explanations would be based on experience and they would thus be a posteriori. As the passage just quoted makes clear, Kant explicitly says that we must be able to show a priori both that this principle exists and that it is also universal. In other words:

Whenever we say “The human being is by nature good”, or, “He is by nature evil”, this only means that he bears within himself a first ground (to us inscrutable) for the adoption of good or evil (unlawful) maxims, and that he holds this ground qua human, universally – in such a way, therefore, that by his maxims he expresses at the same time the character of his species. (Rel. 6:21)

Accordingly, evil expresses the character of the entire human species and nobody, Kant says, not even the best human being, can thus be exempt from it. It is established as a tendency of the human choice to deviate from the moral path and this is why Kant entitles the section on the source of human evil “Concerning the propensity to evil in human nature”.

4.1.1. Evil as Propensity

Besides the provocative tone of this idea of an evil propensity that exists in every single human being, there are a number of problems with regard to the very definition of it.

According to the general definition, a propensity is “the subjective ground of the possibility of an inclination”. (Rel. 6:29) Kant carefully makes clear that a propensity [Hang] must not be conflated with a predisposition [Anlage] for although they both can be innate, the former need not be thought of as innate but rather as acquired or brought upon the human being by himself.
Moreover, in contrast to the inclination, which is a habitual desire, the propensity has not yet become custom and hence it is to be looked upon as a merely latent feature of the human soul. The term propensity is common in Kant’s vocabulary and there are several propensities that serve as the preliminary stage to inclination to a multitude of things. The propensity to evil, however, is clearly of a special kind and can therefore be regarded a moral propensity, that is, a propensity that affects our attitude with respect to the moral law. In spite of its apparent importance for his theory of evil, Kant’s explicit talk about the meaning of a “propensity” is very short. Besides the plain definition of it as an embryonic inclination he only adds that the term, in its general meaning, refers to the predisposition to desire something that gives pleasure. In this, the propensity is not only distinct from all specific inclinations but also from the instincts that push both men and animals to develop their inborn skills without previous knowledge of the object. In short, one can say that the propensity to evil is not an inclination, or a drive, for evil but rather a distinct feature of our nature, or “heart”, that make us susceptible to certain kinds of things.

Although all evil actions are evil because of the deviation from the moral law, which, in turn, is the result of the evil choice that we once made, Kant allows for three degrees of evil. Of these, weakness (fragilitas) is the least grave since in this case the agent obviously wants to do what he knows to be morally good. The weak person thus incorporates the good principle in his maxim but since he – like all human beings – also has additional desires and these non-moral wants may be stronger, he may, therefore fail to live up to his own high standards.

The second form of transgression of duty that is expressed through the propensity to evil is called impurity (impuritas, improbitas) and depends on the agent’s incapability to make the law alone a sufficient incentive for his action. Even in this case the person knows what is good. Furthermore, he wants it and has also the accurate incentive at hand. Yet, this incentive (i.e., respect for the law) is, in this case, incapable by itself of determining the agent’s will. The action, which indeed is in full agreement with the moral law, is, therefore, good with respect to its object, but because of the admixture of further incentives that were needed in order to motivate the agent to

---

179 The distinction between inclination and propensity corresponds to the distinction between actuality and possibility. *(VA [Friedländer] 25: 583-4)*

180 “Prädisposition zum Begehren eines Genusses” *Rel* 6:29 n. It might indeed seem somewhat surprising that Kant first says that there is an important difference between propensity [Hang] and predisposition [Anlage] and in the same sentence says that propensity is a predisposition. Even if this could be taken as evidence for Kant’s lack of consistency, it should at least be noted that Kant, in this context, uses the term Prädisposition and not Anlage. The actual difference between this Prädisposition and an ordinary “predisposition” [Anlage] (which is said to be distinct from a propensity) is not clear and it should perhaps be noted that the term “Prädisposition” only occurs once in Kant’s own writings.
perform it, the maxim must be regarded as *impure* and the action as not done from duty alone.

While both these stages can account for some evil actions Kant needs much more in order to substantiate his idea of evil as a deeply rooted feature in the human heart. While the former two express merely *unintentional* guilt (*culpa*), true evil consists in *deliberate* guilt (*dolus*), and is characterised by a kind of dishonesty [*Tücke*] of the human heart. This is why he adds the third propensity, *depravity* (*vitiōsitas, pravitas*), that truly expresses why we can be said to be evil also in the more grave Kantian sense. It is because of this depravity that man freely – in his maxim-making faculty – subordinates the law to non-moral incentives. Even in this case, the actions may be in accordance with the law (and thus *legally* good) but, because of the corrupt or perverted nature of the *form* of the maxim, they can by no means be considered good in the true *moral* sense. By this short outline Kant concluded:

> the propensity to evil is here established (as regards to actions) in the human being, even the best; and so it also must be if it is to be proved that the propensity to evil among human beings is universal, or, which amounts to the same thing, that it is woven into human nature. (*Rel.* 6:30)

The rational origin of this disharmony [*Verstimmung*] within our nature will always remain inexplicable to us and it can also never be eradicated [*ausgerottet*]. Still, since the propensity to evil obviously must be *moral*, it can only be attributed to us as a component of the moral faculty of choice and the only thing that can be morally imputable is a *deed* [*Tat*]. Yet Kant says, “by the concept of a propensity is understood a subjective determining ground of the power of choice that precedes every deed, and hence is itself not yet a deed.” (*Rel.* 6:31) It seems, therefore, as if we, in order to avoid a contradiction, must allow for two distinct meanings of a deed of which both, however, must have a relation to freedom. Now, Kant says, in one sense, “deed” may apply to the free adoption of the *supreme maxim* by the power of choice (either in favour of, or against, the law). In another sense, deed refers to the performance of the actions themselves by our active use of the supreme maxim. In light of this, Kant says that the propensity to evil is a deed in the *first* meaning and that it can thus be called a *peccatum originarium*. This *original sin* is, in turn, the formal ground of every further deed in the *second* meaning, thereby of every action in the world. These evil actions may therefore be called *vices* or derived sins (*peccatum derivativum*) since they are all the result of the first evil deed.

The former, Kant says, is an *intelligible deed* “cognizable through reason alone” and is independent of temporal conditions. This is without doubt one of the more perplexing notions of Kant’s idea of the origin of evil for he definitely seems to say that the first ground – *the supreme maxim* – while present within us at birth, is nevertheless *acquired* though “not in time”
[nicht in der Zeit]. It is far from clear how one is to understand this idea of what has been called a “noumenal acquisition” and it is not even clear what it is that is acquired by this choice. Admittedly, one reason why Kant made this ground “innate” and inexplicable is simply to make plain that we cannot find any further cause that may explain why evil has corrupted our highest ground. Moreover, in order to avoid an eternal regress in the search for further causes, the first ground must be inscrutable. (Rel. 6:21n. 6:31) As we will see, however, the difficulties do not only affect the idea of a propensity to evil, but also the notion that is believed to express our ultimate moral ground, namely, Gesinnung. I will return to Kant’s idea of Gesinnung and some problematic implications of this notion in section 4.2.

4.1.2. The Universality Claim and the Missing Proof

Of the many radical arguments that Kant made about evil in the Religion, the universality of the propensity to evil is probably one of the most challenging. However, what perhaps may be said to be even more provoking is Kant’s unwillingness (or perhaps inability) to give any substantial arguments for it. The recent debate with regard to Kant’s so called “missing proof” is based on Kant’s claim that:

We can spare ourselves the formal proof that there must be such a corrupt propensity rooted in the human being, in view of the multitude of woeful examples that the experience of human deeds parades before us. (Rel. 6:32f.)

Besides being annoying for those who want to understand Kant’s grounds for the universality claim, this statement is also perplexing since experience can only provide us with a posteriori knowledge and this method can, of course, never establish any claims of universality. At the same time, Kant does explicitly deny that the judgment “man is evil by nature” is an analytical claim whose truth-value could be derived from the very concept of a human being; for then, Kant says, the propensity would be necessary. (Rel. 6:32).

This argument might sound surprising, especially since Kant repeatedly says that this propensity cannot be eradicated, that it belongs to us universally as a natural propensity, that it is “woven into human nature” and that we thus must accept it as an inescapable feature of the human being. Accordingly, what Kant needs to explain is why the propensity cannot be allowed to be “necessary” even though he underlines that there is no human being without it. Because of Kant’s explicit denying of evil as belonging to the concept of the human being, the general strategy among commentators is to conclude that Kant must mean that the phrase “the human being is evil by nature” is
While the argument here is highly abstract and also involves several obscure implications, I believe that this idea is the most reasonable way of understanding Kant’s intentions. Though a *synthetic a priori* status of human evil will, admittedly, also establish evil as necessarily belonging to the human species, it does not – as the analytical alternative would do – exclude the possibility that man, at the time of his first choice, at least *could* have chosen a good first maxim. Moreover, the fact that the choice was made “outside time” seems to rule out all *a posteriori* explanations of the origin of evil. While the *a priori* status of evil is supported by several passages in the *Religion* (e.g. 6:20; 6:36) and that it seems to be a consequence of the “timeless” choice (Rel. 6:25; 6:43), Kant nevertheless seems to admit that the existence (though not the nature) of this evil ground can be established empirically simply “through experimental demonstrations of the actual resistance in time of the human power of choice against the law” and that, because of the cognition we have of the human being, “he cannot be judged otherwise”. (Rel. 6:35; 6:32) Hence we can at least say that evil is at least *subjectively necessary* in every human being, including the best. (Rel. 6:32)

Commentators generally regard the omission to prove this central thesis as a flagrant blunder in Kant’s line of argumentation. Almost everyone agrees that Kant obviously *needs* a formal proof and at least some readers claim that the very phrase “we can spare ourselves the formal proof” indicates that Kant must have had a proof available, or at least believed that such a proof was possible.\(^{182}\) If this is true one may wonder why he would simply “spare” his reader such evidence of his intellectual capacity. However, it might be the case that Kant actually believed that he had provided what was needed in this case. At the end of section I:iii, which ends with the strong words of the Apostle: “There is no distinction here, they are all under sin – there is none righteous (in the spirit of the law), no, not one”, Kant adds a footnote saying that “The appropriate proof of this sentence of condemnation by reason sitting in moral judgement is contained not in this section but in the previous one.” (Rel. 6:39n.) This comment is generally refuted as mere nonsense since a “formal proof” is nowhere to be found in any of the previous sections and no attempt to derive something like a formal and *a priori* proof from these pages, or indeed from other previous works, has yet been successful.\(^{183}\)

---


\(^{183}\) Muchnik argues that Kant’s missing proof is indeed available but that it is carefully hidden in the preface to the second edition of the *Religion*. See Muchnik (2010). Cf. also Morgan who argues that “seeds of a formal proof” can be found already in the *Groundwork*. See Morgan sec. 3.
Now, since this important claim seems to require more than a mere “empirical generalization” of the evil observable in the world, charitable readers want to help Kant here by providing the proof for him. By escaping behind the obscure idea of a mere “subjective necessity”, it seems as if Kant wants to say that since the human being “cannot be judged otherwise” our reason finds it necessary to assume the existence of an evil propensity in a similar way as the subjective “need of reason” gave us other postulates. Indeed, Allison suggests that the proof could take the form of a postulate. Allison’s main argument for this is based on a reductio that aims to show the impossibility of the existence of a “propensity to good” which, given Kant’s rigorism, seems to be the sole alternative to the propensity to evil. Since a propensity to good is to be conceived as a propensity to subordinate the incentive of self-love to the moral incentive, it would, according to Allison, be incompatible with our nature as finite beings since it would remove the need of a higher demand restraining our will. This argument draws on what Kant had said about finite beings as opposed to holy ones since holy beings will always and necessarily have a will that is in agreement with the moral law. A human and finite will, on the contrary, is sensuously affected and this is why the moral law presents itself to us as a law of duty presenting us with imperatives.

This argument has been criticized by many. Wood calls it “especially un-promising” simply because it makes an, in his view, incorrect link between finitude and unholliness. Others have claimed that not only does it presuppose an inadequate idea of what a propensity to good would be like, it also incorrectly suggests that such a propensity would be logically impossible, which Kant himself never says. As Stephen Palmquist notices, Kant states

184 It should be noted that Wood, who indeed wants to downplay the a priori tone of Kantian evil admits that while he himself had once regarded Kant’s argument as expressing a mere “empirical generalization”, he now admits that this must be considered a naïve view. Wood (1999) p. 287.
185 DO 8:136 Here Kant adds the note: “to orient oneself in thinking in general means: when objective principles of reason are insufficient for holding something to be true, to determine the matter according to a subjective principle”.
186 Allison (1990) p. 155. (2002) part II. On the contrary, Paul Guyer, for example, argues that Kant does not say that the “ubiquitous existence of evil” needs a formal or a priori proof, but what need to be proven is rather that this “ubiquitous evil” that we see all around is radical. This nature of evil can, however, not be proven empirically but must be understood a priori, as the free choice of a first evil maxim. It is the radical evil that represents our “crookedness”. Guyer (2009) p. 147.
188 Wood (1999) pp. 287, 402. Wood seems to base his refutation of Allison’s idea by saying that the finitude of our will only implies that we might not be holy since our inclinations may be in accordance with the law. Still, I believe that Allison is right here and Kant frequently says that a holy will always has the categorical imperative as its principles, and this is obviously incompatible with acting from inclinations however lawful these desires may be. Moreover, a holy being does not only act in accordance with duty but also always from duty. Cf e.g. GMS 4:439 and KpV 5:82, 5:87.
that “a propensity can indeed be innate yet may be represented as not being such: it can rather be thought of (if it is good) as acquired, or (if evil) as brought by the human being upon himself”. Moreover, in the Anthropology, Kant actually talks about a propensity to good [Hang zum Guten], which along with the “favourable natural predisposition” [günstige Naturanlage] must be presupposed if we shall attribute the possibility of a character to human beings. Yet, as we have seen in the previous chapter, Kant certainly wanted to allow for the possibility of developing a character so either we must allow for the possibility of a propensity to good, or we must regard his choice of terms as confused and conclude that we cannot take this passage literally.

The central thing that must be made clear with respect to the possible existence of a propensity to good is what exactly this propensity would be. As we saw, according to Allison, such a propensity would be “based on a maxim, to subordinate the incentive of self-love to the moral incentive.” In other words, Allison says, “a propensity to good would consist in a kind of spontaneous preference for the impersonal requirements of morality over one’s own needs as a rational animal with a built-in desire for happiness.” This propensity would thus always outweigh the inclination. Besides the obvious difference with respect to moral status, this last claim points to a significant dissimilarity between these two possible moral propensities since it is clear that the existence of a propensity to evil does not imply that we will always subordinate the law to our subjective wants. The reason for this is, of course, the simultaneous presence of a predisposition to good working against the corrupted ground by encouraging our choice to make the law our leading incentive.

Now, given these descriptions, it is fairly clear that our predisposition to personality, as the “the susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice” (Rel. 6:27) is at least functionally identical to “a propensity to good” so there is nothing that makes it wholly impossible for us to choose to act from respect for the law. Yet, since there can be only one moral propensity, we cannot have a good one since we would then not have a conflicting tendency that makes us incorporate inclinations rather than the law. As we have seen, the mere presence of subjective ends (inclinations) is not what makes us evil, but we are evil, only

191 Anthr. 7:329. Admittedly, this is the only explicit occasion of “Hang zum Guten” in the Academy Edition. Still, the fact that Kant clearly distinguishes it from the predisposition, one cannot object that he is merely confusing the terms. It could perhaps be noted that for some reason, Louden here chooses to translate “Hang” tendency rather than the more customary “propensity”.
193 One can say that one difference between these two is that the “propensity to good” would be a kind of desire or inclination to good while the “predisposition to good” (as personality) is a capacity to exercise ones rational will in order to make the law one’s chief incentive.
on account of our tendency to subordinate the law to them. Thus, without a propensity to evil no “reversal of the moral order” would ever take place and the human being would, like the holy one, necessarily act from the moral motive.

Accordingly, I believe that Allison is right when he says that in order to keep man’s status as a human (i.e. sensuously affected) being, one cannot ascribe a propensity to good to him. Furthermore, provided that there must be a moral propensity inside us which along with the good predispositions constitute our moral nature, this moral propensity must be a propensity to evil. Whereas this argument is formal enough to fill the gap that Kant leaves by his open omission of the proof can be discussed and while there is much that can be said about Kant’s missing argument I believe that there are alternative ways of tackling this query. It is, in fact, not necessary to assume that Kant’s saying that we “can spare ourselves the proof” actually means that he had one, nor does it indicate that he thought that he needed a formal one. After all, Kant had immediately before these lines explained that according to the empirical knowledge we have of man, “he cannot be judged otherwise” and thus we can spare ourselves of a proof. Moreover, the very way he expresses this argument seems rather to be a rhetorical strategy and who should he paraphrase if not Rousseau? Accordingly, when Kant says that we can “spare ourselves the formal proof that there must be such a corrupt propensity rooted in the human being, in view of the multitude of woeful examples that the experience of human deeds parades before us” he is merely alluding to Rousseau’s so legendary saying that “men are wicked, sad and continuing experience spares the need for proof.” (2D note. 74/202) Since it is almost the same wording, it is hardly a coincidence and in the light of the overall tone of the Religion, particularly its pessimistic description of the present man and his moral fall, one cannot neglect the Rousseauian character of the general discussion. Surely, while this suggestion by no means removes the fact that Kant leaves a critical argument – the strong claim that every human being is without exception infected by an evil ground – without foundation, this explanation could at least explain the origin of the notorious phrase that we may “spare ourselves” the formal proof.

For the moment I will leave possible remaining question marks with regard to the missing proof hanging in the air and continue to a related way of coping with Kant’s idea of evil that also evades the need of a formal proof. This suggestion starts from the idea that evil is simply an effect of the man’s social nature.

4.1.3. Evil as Anthropology

Another way of dealing with the problem is to say that Kant did, in fact, never intend to provide a formal proof and nor he is in need of it. The most prominent defender of this idea is Allen Wood who, instead of creating an-
other “unpromising” reconstruction of the proof, chooses to read Kant’s idea of evil as an anthropological thesis. Regardless of many objections, he persists in his belief that we should take Kant seriously when he says that the ground for the claim that we are permitted to apply the evil-thesis to the entire human species:

can only be demonstrated later on, if it transpires from anthropological research that the grounds that justify us in attributing one of these two characters to a human being as innate are of such a nature that there is no cause for exempting anyone from it, and that the character therefore applies to the species. (Rel. 6:25-6, emphasis added)

This passage appears at the very end of the first section of Part I and precedes the discussions of both the good predispositions and the propensity to evil. It is without doubt true that the Religion is a work that pays much attention to anthropology and especially practical anthropology, which, as we know, is the science that deals with the “subjective conditions” in our nature that helps or hinders us in the fulfilment of moral behaviour (MS 6:217). The fact that the propensity to evil is explicitly located in human nature also strengthens the anthropological tone of Kant’s project. Wood continues by citing the colourful passage from Part III that aims to shed more light on the battle between good and evil that man must endure and cope with in this life. About the human being Kant says:

His needs are but limited, and his state of mind in providing for them moderate and tranquil. He is poor (or considers himself so) only to the extent that he is anxious that other human beings will consider him poor and will despise him for it. Envy, addiction to power, avarice, and the malignant inclinations associated with these, assails his nature, which on its own is undemanding, as soon as he is among human beings. Nor is it necessary to assume that these are sunk into evil and are examples that lead him astray; it suffices that they are there, that they surround him, and that they are human beings, and that they will mutually corrupt each other’s moral disposition and make one another evil. (Rel 6:93-4)

This passage has become Wood’s favourite trump card for his idea that Kant’s theory of radical evil is essentially unsociable sociability. As we have seen, Kant describes this human trait as the propensity of human beings “to enter into society, which, however, is combined with a thoroughgoing resistance that constantly threatens to break up this society.” (IAG 8:20) These conflicting feelings awake the powers that had so far been merely latent in the mind of man.

Wood’s main thesis is thus that the propensity to evil belongs to us as social and historical beings. This conclusion that is spelled out in his com-

---

prehensive book *Kant’s Ethical Thought* (1999), is taken as grounds for his suggestion that Kant and Rousseau, though seemingly different with regard to their ideas of human nature, actually presented one and the same doctrine.\(^{195}\) This argument reappears in his later book *Kantian Ethics* (2008), where we are told that for Kant, men only become evil because of an “invisible enemy” who “hides behind reason” and in this way implants vice in the human mind. This enemy is identified with competiveness, social inequality and the passion for domination over others. Here, Wood adds that this enemy has several names and the most common are self-conceit, unsociable sociability and the propensity to evil. Moreover, Wood adds that all these are the same as Rousseau’s notion of *amour propre* and underlines that this propensity is rational and develops exclusively in a social condition.\(^{196}\) Combined with his further claims we may thus conclude that self-conceit, unsociable sociability and the “radical propensity to evil” are simply “three names for the same reality”.\(^{197}\)

It is no doubt that at least some of these terms – especially *amour propre* and unsociable sociability – have much in common since, as we saw in chapter 1, for Rousseau, it was the sudden closeness to other men that exposed the natural inequality, and made us aware that some people actually were stronger, more talented and/or more attractive. Thus envy was born as well as the shame that followed when we could not succeed in this competition. At the other extreme, the winner who realized his superiority became proud and contemptuous. From these insights about our mutual relations, hate, fury, rivalry and other unwelcome feelings arose and exposed men’s most primitive drives. Again, all these feelings could be put under the general idea of an inner anxiety born from a sense of weakness and fear of not being loved. Accordingly, it seems as if one can justifiably assume that Kant’s idea of the social process is founded on a Rousseauian ground.

Still, before one extends these ideas to the propensity to evil one should take another look at Kant’s idea of the form of *self-love* that according to Kant is the key feature of our predisposition to humanity. In contrast to the predisposition to animality, which was plainly physical and thus mechanical, the predisposition to humanity was characterized by its need for reason. This self-love made comparisons, and this relative outlook gave birth to such harmful social passions as jealousy and rivalry and perhaps also envy, ingratitude and joy in others’ misfortune. Many of the passions mentioned in this section are the same as we found in the outline of unsociable sociability as well as in the beginning of Part III of the *Religion*. Although inclinations are as such not evil, Wood says that in human beings they are essentially

---


corrupted since they only appear in a socially affected (and thus corrupted) person.

This is indeed an interesting way of attacking the dilemma and I see no real problem with the fact that Wood invokes the fundamental historical explanation of human history (and thereby the origin of human evil) since this adds much to the understanding of the main topics presented in the Religion. What makes me somewhat concerned is the fact that he also plainly equates all the various passions and drives that operate in this process. It is clear that he sees the propensity to evil as identical to unsociable sociability since they are merely “different names of the same reality”. Since Kant’s idea of unsociable sociability is the result of cultural and social intercourse, it is practically indistinguishable from the relative self-love explained in the Religion. However, this self-love is our predisposition to humanity and all our predispositions, including humanity, are predispositions to good. Here we may recall what was said in chapter 1.2., namely that social antagonism – by awakening our slumbering abilities – was the “first true step from crudity toward culture” and the cause of everything good and truly human. (IAG 8:21; Anthr. 7:324) Accordingly, since unsociable sociability, as an inner sense of restlessness, is, in all relevant respects identical to relative self-love (and in all relevant aspects also to Rousseau’s idea of amour propre) it seems contradictory to say that this original predisposition to good is just another name of the propensity to evil.

Admittedly, it should be mentioned that Wood himself has not explicitly stated that the predisposition to humanity is identical to unsociable sociability. Yet, in two recent papers he says that the predisposition to humanity turns into unsociable sociability that, in turn, becomes the source of radical evil.198 Still, it seems difficult to separate them for in Kant’s only definition of unsociable sociability we find that he sees it as the “antagonism”, or “resistance”, that awakens the powers of the human being in the same way as relative-self-love does.199 It is caused by our tendency to enter into societies but the predisposition to this process obviously lies already in human nature. Moreover, since it definitely seems that Wood has equated relative-self-love with amour propre, and since he equates amour propre with unsociable sociability it is not easy to navigate in this discussion. I will return to the problem of the cause of evil in Chapter 6 since the remaining queries require a close examination of the various expressions of self-love as well as a better

199 Indeed, if one is to be pedantic, it seems as if unsociable sociability must involve both our two lower predispositions since the desire for entering society is included in our animality and it cannot properly be said to belong to our humanity. This social element of our physical self-love also invalidates the alleged parallel between Kant’s two predispositions and Rousseau’s forms of self-love since amour de soi does not entail any longing for social contact. Still, sociability is obviously present in Kant’s “humanity” and it is also a necessary component of it.
understanding of one of the most powerful of the social passions involved in this process that, according to Wood, is the source of all evil and so essentially the same as unsociable sociability and *amour propre*, namely *ambition*.

Now, presented merely as an explanation of the origin of evil behaviour in the human life, the identification of our propensity to evil with an anthropological and historical tendency may perhaps seem plausible. However, as we will see, this interpretation may not be able to survive the implications that follow from a full study of Kant’s argument in the *Religion*. In order to see the full scope of Kant’s theory of the origin of evil – as well as of good – we need to consider the notion of Gesinnung.

### 4.2. Gesinnung

Kant regarded human beings as evil by nature. At the same time, since we ought to be morally good, there must also be a source of goodness inside for otherwise striving for moral perfection would be nothing but a ridiculous waste of time. Since our good germ exists and can never be lost, we need not *acquire* but rather *reclaim* it and restore it to its full power.\(^{200}\) One of Kant’s main aims with the *Religion* was thus to identify the source of man’s corruption and find out how to reclaim our original germ of good and perform the “revolution” that is required by our duty to strive towards moral holiness and the highest good. Although there are many obscurities with regard to this idea of a moral revolution, it is in any case clear that Kant – at least on one occasion – explicitly declares that this revolution must take place in our *Gesinnung*.\(^{201}\)

---

\(^{200}\) See *Rel.* 6:46. Kant does not here spell out any arguments for the claim that the germ of good cannot be lost. We are only told that *if* we actually *would* lose it, we would never be able to regain it. It might be the case that since the germ of good is here said to consist in the “incentive that consists in the respect for the moral law” the germ of good is simply our predisposition to personality and if we would lose this predisposition, we would no longer be complete human beings since we would then no longer be *moral* beings.

\(^{201}\) *Rel.* 6: 47. *Gesinnung* is a term that has generally caused major problems to English speaking commentators of Kant. Some choose to keep the German form but translators must find a term that is at least acceptable term. No one seems entirely happy with their choice and this is shown by the fact that most authors or translators feel the need to defend their choice in (often extensive) footnotes. The predominant translations are *disposition* and *attitude*. In the preface to Werner Pluhar’s recent translation of the *Religion* (2011), Steven Palmquist, (who among these two favours disposition) admits that Pluhar’s choice of attitude is not entirely inaccurate but suggests that “conviction” would be even better. This choice is further spelled out in his recent article about Gesinnung and his main argument for using conviction is the fact that, because Gesinnung is chosen, it signifies a well thought-of decision and a certainty about one’s choice in the sense of an “Überzeugung”. (See Palmquist 2015). A common objection to the choice of “disposition” is that it is too imprecise. Others point out that “disposition” may easily be confused with *predisposition* [Anlage] which, at least in this context, is something completely different from what Kant means by Gesinnung. Munzel rejects all single termed alternatives and proposes the more complex notion “comportment of mind”. Other suggestions are “moral tenor” or “character”.

138
Along with the highly difficult and perhaps unintelligible notions of a universal human propensity to evil, which together with our natural predispositions to goodness constitutes our moral nature, we must now take on the task of interpreting what Kant may have meant by Gesinnung. It is commonly claimed that the term was first introduced to the German language by Gotthold Ephraim Lessing around 1753 as a translation of the French sentiment.\textsuperscript{202} Now, it seems clear that for Kant, Gesinnung is something entirely different from a mere sentiment. Kant’s conception of Gesinnung obviously involves a normative tone that is not to be found in his general talk about sentiments and feelings for which he normally used the terms Empfindung or Gefühl. Someone who comes much closer to Kant’s view is Baumgarten who in the fourth edition of his Metaphysica (1757) adds some German translations to his Latin vocabulary and suggests “Maxime” or “gewöhnliche Gesinnungen” as equivalents for maxima.\textsuperscript{203} As we have seen, for Kant, maxims reflect the agent’s innermost wants and can, just like the Kantian Gesinnung, be either good or evil. In Kantian contexts, Gesinnung is generally translated simply disposition or attitude. It is – it is said – the ultimate ground of the agent’s maxims and as such, it is also determines the moral worth of every action as well as the agent himself.

The term Gesinnung is used throughout Kant’s entire productions and figures frequently from his earliest notes on anthropology to the moral works of the late 1790s. That this term is important for the understanding of Kant’s idea of the moral person is shown by numerous passages in various works. From Ludwig Collins’ lecture notes we learn that in contrast to juridical matters “ethics refers exclusively to Gesinnung” and moral actions must thus be judged on the basis of the inner goodness of this Gesinnung. (VM [Collins] 27:299) In the Groundwork we are told that moral worth is to be sought solely in the maxim, and accordingly in the Gesinnung of the will. (GMS 4:416) We may also recall Kant’s claim that “anything good that is not grafted on a morally good Gesinnung is nothing but illusion and glittering misery”. (IAG 8:26) In the Critique of Practical Reason, Kant makes the same distinction as he did in the Religion between the one who acts lawfully merely according to the letter and the one who acts according to the spirit and

\textsuperscript{202} E.g. Philosophisches Wörterbuch der Philosophie vol. V p. 107, Eberhard p. 459 (entry 664 where it is said to be the “Äußerung der Sinnesart” or expression of one’s inner life). The identification of Gesinnung with “sentiment” indeed seems to hold for Lessing’s use of the term and it is noteworthy that Lessing, in his early texts, typically uses the plural form (Gesinnungen, i.e. sentiments) which is wholly compatible with the possibility of different or frequently shifting “Gesinnungen” within the mind of one and the same person. Gesinnung is also the term that was used as a translation of the English term “sentiment” in the first German translation of David Hume’s Enquiry, the edition which, according to Warda, was present in Kant’s own library. (Philosophische Versuche über die menschliche Erkenntnis which was included in the collection Vermischte Schriften vol. 2 (1755). A comparison between the English and the German text shows that the term “sentiment” is rendered “Gesinnung” at several occasions).

\textsuperscript{203} Baumgarten, Metaphysica § 699.
equates the latter to one who acts from Gesinnung (KpV 5:72n.) These and numerous other passages show that Kant must have had an idea of a fundamental source in the depth of our nature from which the true self is revealed and from which the morality of the human being can be derived.

This idea of an ultimate and inner moral quality of the human being is not unique, and it is, as such, not even problematic. Indeed, as Allison said, Kant’s introduction of this feature in his moral theory was the missing piece that filled several gaps, and which thus made explicit what was never explained in the Groundwork. As we saw in chapter 3, Allison argued that the Groundwork’s incomplete way of explaining human actions would leave us with the idea of actions as “free-floating, isolated decisions”. Now, by bringing in the idea of Gesinnung, Kant seems to acknowledge the importance of the subjective ground – not only the subjective ground of the actions (i.e. the maxim) – but also of the subjective ground of the maxims themselves, as the basis of our adoption of them. This, in turn, leads us to a deeper and more complete understanding of the moral nature of human beings since this perspective takes into account the underlying intentions or quality of our will and of the faculty of choice. As we have seen, it is this inner ground (nature) that conducts the human being’s use (and misuse) of freedom. (Rel. 6:21)

However, what is problematic for those who want to understand Kant’s true intentions here is that this ground of maxims is also, as we have seen, generally identified with our rational will [Wille] or sometimes with choice [Willkür]. More frequently, this feature – the inner foundation of our moral person – is called moral character or Denkungsart. Furthermore, at least in the Religion, the ultimate ground of the agent’s maxim is identified as a propensity and more precisely with the propensity to evil. Moreover, looking at Kant’s definition of the predisposition to personality one sees that it virtually coheres with the good Gesinnung and this would thus mean that the seemingly important distinction that Kant draws between predisposition and propensities fades away.

I believe that these factors deserve a closer examination. This is not to say that it has not been done before; on the contrary, this topic is widely discussed, mainly because of its central role for the understanding of Kant’s idea of evil. The problems that have attracted most attention are the relationship between Gesinnung and the propensity to evil and the relation between Gesinnung and our moral character. The understanding of these relations may, as we will see, affect the interpretation of the possibility of moral progress. I will begin the following sections by outlining the most common (standard) interpretation of Gesinnung. The reason why this understanding is of special interest is, as we will see, that it gives rise to several of the problems that are generally associated with Kant’s idea of the need of a “first ground of good and evil” in human nature since this interpretation requires that Gesinnung and the Hang coincide. After that, I will point out and briefly discuss some alternative interpretations and finally say what I believe is the
most reasonable way of dealing with Kant’s discussion of the problems that we have met in this chapter.

4.2.1. The Standard Interpretation

Kant admits that the very meaning of Gesinnung is hard to explain. Still, his examples indicate that he uses the term to refer to a deeply embedded aspect of our nature that is also the sole factor that can give moral worth to the human being. The understanding of Gesinnung as the fundamental basis of every choice made by a human being has, combined with some of Kant’s further comments, lead to what I here will label the standard interpretation.

The standard interpretation claims that Gesinnung is to be understood as the agent’s highest (or supreme) maxim [oberste Maxime] and as such it is also the ultimate ground of the adoption of all further maxims. This reading is so established that many commentators do not even bother to argue for it. Those who try to support this view refer to Kant’s claim that Gesinnung is the “first subjective ground of the maxim” (Rel. 6:25) and add that Kant explicitly said that the first ground of the adoption of maxims – in order to be located in the human being’s free power of choice – must be a maxim itself. (Rel. 6:20-21 including the footnote at p. 21.) By looking at Kant’s most prominent works on morals, we see that Gesinnung is ascribed an important role for moral worth. Moral worth originates in our Gesinnung; it is according to this inner disposition that we can judge a being as good or evil. Since the aim of the Religion is to identify the very root of evil (and in this way find the means for a moral revolution) it makes sense to understand this supreme ground in terms of what makes us good or evil. Thus the identification of this first ground with our Gesinnung seems also to be consistent with Kant’s theory as a whole.

The highest maxim-interpretation of Gesinnung requires some new definitions. In order to make room for a so called highest ground, we must make a distinction between two kinds of maxims. From the Groundwork we have learned that a maxim is the “subjective principle of the action” that must be distinguished from the “objective principle” which is the practical law. Such maxims, which thus express the ends and the formal grounds of particular actions, are sometimes called “first-order maxims”. Yet, as we have seen, there must also be another kind of maxims, namely such maxims that serve as the ground of the adoption of all specific maxims that, in analogy with the first kind, are called “second-order maxims”. This division, which

---

204 GMS 4:400 n. In another footnote (GMS 4.421), Kant makes the somewhat misleading claim that the maxim expresses the principle according to which people actually act, whereas the law states the principles according to which one ought to act. This is not the full view, for maxims merely express our intentions or wants. In the Religion Kant makes clear that the very fact that a person has a certain maxim does not mean that he will also act upon it for “between maxim and deed there is a wide gap” (Rel. 6:47).
also can be viewed as a division of maxims into action-maxims versus dispositional-maxims is commonly accepted.\textsuperscript{205} This second-order maxim directs, or determines, which first-order maxims that we choose, or at least tend to choose. What makes second-order or dispositional maxims special is that they, in contrast to our first-order (action) maxims, are said to be purely formal. The Gesinnung, regarded as a second-order maxim, therefore, does not direct the first-order maxim towards any end, but it only affects the form, or moral aspect of the action. This corresponds to the two forms of deed that we met above: intelligible deed – “deed in the first meaning” – through which we adopted the supreme ground for every maxim that we will choose and deeds in the “second meaning”, which are actions in accordance with, or contrary to, the law. This first deed thus explains, through being their supreme ground all further deeds in the “second meaning”.

Although I will not directly reject the maxim-interpretation, I will disclose some difficulties that it has to deal with and I will also show why this reading makes it is impossible to keep the distinction between Gesinnung and Hang. This line of argumentation also reveals the fact that Kant, in fact, does not – or at least not explicitly – say that Gesinnung is our highest maxim. If we take another look at the passages from the \textit{Religion} that are frequently cited in order to verify the standard interpretation we will see:

\begin{quote}
The Gesinnung, i.e. the first subjective ground of the maxim, can only be a single one, and it applies to the entire use of freedom universally. This Gesinnung too, however, must be adopted through the free power of choice, for otherwise it could not be imputed. (Rel. 6:25)
\end{quote}

Obviously, Gesinnung is said to be the \textit{first subjective ground of the maxim}. We also see that this ground must, some way or another, have been adopted, for otherwise we could not properly praise or condemn the agent for having it. Though, there is no reference to a \textit{highest maxim} here we can read some pages earlier that:

\begin{quote}
In order, then, to call a human being evil, it must be possible to infer a priori from a number of consciously evil actions, or even from a single one, an underlying evil maxim, and, from this, the presence in the subject of a common ground, itself a maxim, of all particular moral evil maxims. (Rel. 6:20)
\end{quote}

Here there is, indeed, a reference to a fundamental ground of maxims that is \textit{itself a maxim}. Still, the context shows that Kant is talking in a very broad manner about something that must exist in human nature. In order to specify what he means, he continues:

\textsuperscript{205} See e.g., Schwarz esp. ch. 8, Timmermann (2000), Gressis (2010a) and (2010b), Potter p. 78 and Hills p. 83ff.
by “the nature of a human being” we only understand here the subjective ground – whatever it may lie – of the exercise of the human being’s freedom in general (under objective moral laws) antecedent to every deed that falls within the scope of the senses. (Rel. 6:21)

This reference to a “subjective ground” governing the “use of freedom in general” undoubtedly fits the description of Gesinnung from page 6:25 where Gesinnung was said to apply to “the entire use of freedom universally”. However, if we continue the analysis of page 6:20-21 we learn:

Hence the ground of evil cannot lie in any object determining the power of choice through inclination, not in any natural impulse, but only in a rule that the power of choice itself produces for the exercise of its freedom, i.e., in a maxim. (Rel. 6:21)

and then:

One cannot, however, go on asking what, in a human being, might be the subjective ground of the adoption of this maxim rather than its opposite. For if the ground were ultimately no longer itself a maxim, but merely a natural impulse, the entire exercise of freedom could be traced back to a determination through natural causes – and this would contradict freedom. (Rel. 6:21)

What makes the argument somewhat ambiguous is that Kant, in the first of these two latest quotes says that the ground must lie in a maxim while the second says that the ground is (and must be) a maxim itself. On the other hand, the following lines tell us:

Whenever we therefore say, “The human being is by nature good”, or, “He is by nature evil”, this only means that he bears within himself a first ground (to us inscrutable) for the adoption of good or evil (unlawful) maxims, and that he holds this ground qua human, universally – in such a way, therefore, that by his maxims he expresses at the same time the character of his species. (Rel. 6:21)

It is noteworthy that Kant, in this intricate discussion of the “first ground of all maxims”, which must itself be a maxim, does not make use of the term Gesinnung. In fact, the first occurrence of “Gesinnung” in the Religion is found at the above mentioned passage at page 6:25. Before that, Kant is concerned about explaining the ground of evil in human nature, a ground that he regards as something present in every human being as a characteristic of the whole species.

Now we reach the crux of the matter for the universality of this evil ground is an important observation. When Kant some pages later provides his detailed discussion of the very nature of the propensity to evil he explains this in almost identical terms, as an inner ground that corrupts and distorts
the human power of choice so that we will reverse the internal order of our incentives at the moment we incorporate them into our maxims. As we have seen, the propensity to evil is defined as the rational origin of the “disharmony” of our nature with respect to the way by which the power of choice “incorporates lower incentives in its maxims and makes them supreme”. Accordingly, in my view, there seems to be no apparent difference between Kant’s idea of the nature of our propensity to evil and the “first ground” which is discussed at page 6:20-21. Since the supporters of the maxim-interpretation of Gesinnung must refer back to this discussion, they must thus admit that it is difficult to see exactly how Gesinnung differs from the propensity. Now, since the propensity is unavoidably evil in every human being (even the best) an equating of this propensity with our Gesinnung implies that all human beings must thus have an evil Gesinnung.

Yet, there are several scholars who, while apparently accepting the standard interpretation of Gesinnung as our highest maxim, still deny that our Gesinnung must be necessarily evil. However, in doing so, they need to prove that Gesinnung is not the same as Hang. Whether this is a real problem or not is not obvious but I will present some commonly raised argument against this identity that also stresses the need of separating them.

4.2.2. Gesinnung and Hang

While an identity between Gesinnung and Hang undeniably seems to present a problem for Kant’s theory of the moral person, several prominent Kantians openly accepts this conclusion. However, most readers of the Religion do not treat Gesinnung and propensity as equivalent. While some readers do not even seem to be aware of the difficulty of arguing for the maxim-interpretation in this manner (e.g. by connecting the definition from page 6:25 with the comments from 6:20-21), others struggle to find a reading that avoids this supposed trap. Finally, there are those who frankly state that it is important not to overlook the “clear distinction” between Gesinnung and propensity.

Recently, Pasternack has argued that negligence of the difference of Gesinnung and Hang has major effects on Kant’s idea of moral progress, and he treats the Gesinnung-Hang distinction as one of the most fundamental in the Religion. Without it, he argues, Kant would no longer be able to allow for a change of heart and this, we are told, is “no minor issue”. Once again he refers to the interpreters’ general disregard of the Religion as a whole and

---

207 For the last claim, see e.g. Pamquist p. 288 n.13.
208 Pasternack pp. 100, 117.
claims that once one proceeds beyond Part I, the distinction becomes inescapably necessary. Pasternack sees our moral nature as the result of two grounds. First we have an innate propensity and second a Gesinnung, which is the result of a choice. The propensity thus precedes—logically and causally—the agent’s latter choice of his Gesinnung. Admittedly, Pasternack says, once the choice of Gesinnung is made, the two seem to converge but the fundamental difference does not disappear. According to Pasternack it is also important to grasp that the choice of Gesinnung is not at all “timeless”. The idea of choice “outside time” or “noumenal choice” is, in his view, a very common misreading of Kant’s intention the Religion. In fact, he says, Kant does not say anything about choices made outside time, and those who argue for this understanding do not seem to read Kant’s comments about “intelligible deed” in its context. Rather, the choice takes place in time and Pasternack quotes the passage in which Kant says that “if we want to engage in an explanation of evil with respect to its beginning in time, we must trace the causes of every deliberate transgression in a previous time of our lives, all the way back to the time when the use of reason had not yet developed”. (Rel. 6:42-3). Accordingly, Pasternack argues, we develop evil in early childhood and he adds Kant’s idea that our Gesinnung has been the same from “youth on”. (Rel. 6:25)

Now, while this might seem to be a plausible way of evading all mysterious ideas of timeless choices of evil, this passage cannot be used in this way. First we may notice that Kant here says that this is the way we must proceed if we want to search for an origin in time and he goes on by saying that we must not seek an origin in time of a moral constitution [Beschaffenheit] for which we are to be held accountable. (Rel. 6:43) As we should know by now, the evil in our nature is inscrutable. We may indeed feel the need of such an explanation but we must, in the end, accept that the origin of this original disharmony will forever remain inexplicable for us. Moreover, it cannot be denied that Kant, when referring to the “intelligible deed”, says that this deed is “apart from any temporal condition”. (Rel. 6:31) Moreover, Kant explicitly says that Gesinnung is an innate characteristic which nevertheless has been earned by the agent himself but “not in time”. (Rel. 6:25) Furthermore, Kant continues that we “cannot derive this Gesinnung, or rather its highest ground, from a first act of the power of choice in time” which means that we must thus regard this highest ground as a characteristic of the power of choice that “pertains to the human being by nature.” (Rel. 6:25, emphasis added) Admittedly, if one takes this last claim literally, one could perhaps say that while the highest ground of Gesinnung must be sought outside time, Gesinnung itself need not be. Yet, the highly tangled structure of Kant’s arguments in Part I makes it extremely difficult to make a reliable distinction since Kant alternately speaks of a first (inscrutable) ground of

209 Pasternack pp. 115-6.
maxims (Rel. 6:21), a “first ground of the adoption of our maxims” which cannot be any fact possibly given in experience (Rel. 6:22), a “highest maxim” which is the result of our (timeless) deed. (Rel. 6:32) The list could be made much longer. After all, despite minor varieties in wording, all such statements seem to refer to the same thing, namely to the internal and first ground of the use of our freedom (our way of making maxim) as a ground we cannot explain. This ground is, as we know, the propensity to evil or, as Kant apparently sometimes call it – Gesinnung. Besides this, as we have seen, Kant frequently refers to the “first ground” as innate but still as a deed of freedom. In order to be a “deed of freedom” it must have been chosen, and since it is also, at least in some sense, said to be innate, it must somehow have been chosen before birth. Accordingly, attempts to isolate Gesinnung from Hang by unraveling Kant’s tangled terminology will hardly work. Still in my view, no distinction between Gesinnung and Hang is required in order to allow for the “change of heart” and I will return to this topic in the next section.

Others who have argued for a separation are Pablo Muchnik and Stephen Palmquist. Muchnik admits that the respective choices of propensity and Gesinnung seem to be isomorphic in that they are both the result of an “intelligible deed” which regulates our faculty of choice by being the ground of the adoption of first-order maxims. However, he adds, “isomorphism is not identity”. Muchnik, as well as Palmquist, states that the difference between Gesinnung and propensity is simply that while the propensity expresses the moral outlook of the whole human species, the Gesinnung represent the moral outlook of the individual. It is thus important to notice the “subtle distinction” that Kant makes with respect to these choices. According to Muchnik, when we chose Gesinnung, we establish the ultimate subjective ground of the adoption of our own maxims. (Rel. 6:26) In contrast, the choice of propensity refers to the species’ choice of a subjective ground of the exercise of freedom in general. (Rel. 6:21) This difference, Muchnik says, is crucial for Kant’s attempt to justify the universality of evil. Now, while this could have been a sharp-eyed observation, Muchnik seems to forget this all important distinction quite quickly. In his book (2009) he first (p. 54-5) makes the same claim about a “subtle” but crucial distinction but at a note to the same chapter (n.44 p. 82), he explicitly states that Gesinnung “stands for the total use of freedom an agent makes throughout the various circumstances of her life” (emphasis added) and he states this definition by referring to page 6:20-1, a description which he, as we saw, exclusively assigned to the propensity in order to mark out Kant’s crucial distinction.

---

Nevertheless, Muchnik and Palmquist seem to agree that there is no doubt that Kant meant that all human beings have this propensity to evil, but since Kant must also allow for “worse” and “better” human beings, he must also allow for a faculty that can explain why some people actually act morally while others do not. As Muchnik says, it would be almost ridiculous to read Kant as saying that we all “have chosen an evil Gesinnung”. What Kant means is rather that we all have a propensity to act against the law, but morally good human beings have chosen good Gesinnung which triumphs over this propensity to create corrupt maxims. Although not all authors argue explicitly in this manner (or even points out the exact distinction at all), I think that this view is quite common. In spite of Kant’s explicit claim that the propensity to evil, as being a deed in the first meaning (i.e. as a peccatum originarium) cannot be eradicated, it is commonly claimed that one indeed may have, or at least can acquire a good Gesinnung.

On the surface, this is an interesting way of trying to make sense of Kant’s general idea. Why, some commentators have argued, would Kant use two distinct terms for the same feature of the human being’s moral constitution that also seem to conflict with each other? Furthermore, also the mere possibility of difference between individuals with regard to moral goodness seems to be a relevant point. However, the problem with regard to this supposed “clear distinction” becomes obvious if we return to the passages that we considered above. As we saw, in order to attain the maxim-interpretation one must put together several arguments about the notion of a “first (or highest) ground” from various pages in the Religion. In order to complete the link between them one must thus use also arguments where it is beyond doubt that the “ground” in question is said to be a universal human characteristic (Rel 6:21) and that there are no reasons for exempting anyone from it (Rel. 6:25-6). At least at this last passage, Kant is explicitly referring to Gesinnung. Accordingly, whether one likes it or not, it appears as if Kant actually meant that the Gesinnung, at least originally, is corrupted in every human being in the same way as our propensity to evil. This idea may perhaps explain the query expressed by Bernstein who sees the asymmetry between the two with regard to their moral quality as a quandary which Kant never was able to explain. This asymmetry consists in the fact Kant seemed to say that our Gesinnung may be either good or evil while the propensity is only evil and he emphasises that Kant never talks about a propensity to good. However, if we can accept the idea just mentioned, namely that the

---

213 To this, one may reply that even without any differences with respect to Gesinnung, one might explain the difference between morally “better” and “worse” human beings by saying that better persons have a stronger will (or character) and that they therefore exercise their predisposition to personality or the “germ of good” more often than others. 214 Bernstein p. 26. Again, one should notice that Kant actually once did “talk about” a propensity to good which needs to be supposed in human nature in order to allow for the capacity to bring about a character. (Anthr. 7:329).
Gesinnung of all human beings at least *initially* must be considered evil because of the “first choice” that we made, we may still, perhaps, create a good Gesinnung through the moral revolution that is required of us in order to reach moral goodness. This move, I will argue, does not require that Gesinnung and propensity are regarded as separate things since the revolution that will make us good, that is, the change of heart, must concern our *entire* ground.

I am well aware that this idea may seem to conflict with at least some of Kant’s claims, namely what he says at page 6:31 and 6:51, that the propensity, understood as an intelligible deed, is innate and thus *ineradical*. Still, Kant’s bold talk at the end of Part I of the *Religion* about the need of escaping our corrupted state indicates that this evil, after all, *can* be left behind, or at least that we *must assume the possibility* of a “change of heart” through which we will uproot the evil ground in our nature. That this reading does not conflict with Kant’s general reasoning in the *Religion* will be explained in the following section.

4.2.3. Gesinnung, Hang and the “Change of Heart”

While the *Religion* begins with gloom and resignation, there is glimmer of hope in Kant’s idea of a regeneration through a revolution in the hitherto corrupted human being’s soul by which the good principle, which always is there, finally takes the lead in the human being’s way of thinking. Through this ground-breaking act, we perform a, so called, *change of heart*.

While there is much that can be said with respect to this topic, I will here focus on two main questions: 1) how are we to understand the very idea of a change of heart? and 2) is this change possible for a human being?

First of all, it must be made clear that a moral revolution does not consist in an acquisition of a “lost incentive for the good” for this incentive has always been within us as the germ of good comprising the very ground of our very awareness of a moral law to which we are committed. (*Rel*. 6:44) The need of assuming the necessary existence of this germ is obvious, for without it, we would not be aware of our duty to become “good beings pleasing to God” so there would be no motive for us to try to perform the required change. Yet, as we know, we are also “corrupted in the very ground of our maxims”, that is, we are so disposed that we tend to give priority to the principle of self-love. The solution of the problem of the domination of the evil principle is that “by a single and unalterable decision a human being reverses the supreme ground of his maxims by which he was an evil human being.” (*Rel*. 6:48) Put simply, through this *revolution*, we make a new first choice, and this time we will choose to be good. Still, a closer look shows that there are several obscurities involved in Kant’s outline of this reformation of our evil nature. The foremost quandary concerns where exactly this change is supposed to take place, that is, which part – or quality – of our “nature” is
transformed by this move? If we read the *Religion* we find that the revolution is supposed to affect our “Gesinnung” (*Rel.* 6:47), our “Denkungsart” (*Rel.* 6:48) or the “supreme inner ground of the adoption of all the human being’s maxims.” (*Rel.* 6:51) In the end, however, the fundamental alteration is said to be a “change of heart.”

Because of Kant’s explicit focus on the term *heart* in this context, I believe that one cannot take him at his word when he says that the propensity to evil is inextirpable [unvermittelbar] and cannot be eradicated [ausgerottet]. (*Rel.* 6:51; 6:31) In order to see this, we may recall that the propensity to evil is explicitly explained in terms of a corruption of the human *heart*.*215* It is thus reasonable to assume that by “changing our heart” we will thereby also change the *ground of this heart*, including the propensity that corrupts it. Accordingly, in contrast to the common reading according to which the propensity is *absolutely* inextirpable, I would say that it seems as if the change of heart – if performed – actually uproots this propensity which, as we just saw, is located precisely in our *heart*.

As said above, since it is practically impossible to separate the many notions that, at least in the *Religion* are all connected to (or explicitly identified with) the first or highest ground of our maxims or of our faculty of choice, so we need to assume that the revolution actually can reform us completely also to the extent of uprooting our deep-seated propensity to evil. This is in fact a necessary assumption for, as we will see, nothing but a *total reformation* can lead to the ultimate end of a purely moral being pleasing to God that Kant sets forth in the *Religion*. The question is, though, whether a limited being actually can perform this radical move.

Nevertheless, given Kant’s claim that the human being’s constitution *ought* to agree with holiness (*Rel.* 6:66), we must assume that this path from corruption to goodness is at least conceptually possible. Hence, the human being, who is *corrupted in the very ground of his maxims*, can apparently perform a revolution of his deeply depraved nature which is the locus of his radically grounded propensity through which he is to be considered evil. A change of heart is therefore a duty. As was hinted already in 3.2.4, this move is called “a transition to the maxim of holiness of Gesinnung” through which a *new man* comes about; it can be seen as a *rebirth*, a *new creation* and a *change of heart*. (*Rel.* 6:47) Yet, as we have seen, this *radical* process is merely hypothetical for it is unlikely that any human being will actually perform such a fundamental change during his life. Moral perfection is most of all an *ideal* and this ideal consists in the “*prototype* of a moral Gesinnung in its entire purity”. (*Rel.* 6:61) In order to reach the state of holiness, a mere

---

215 When portraying evil, Kant frequently talks about it in terms of the “corruption of the human heart”. Furthermore, all the three forms of that are presented as degrees of evil (i.e. weakness, impurity or corruption) are said to be imperfections of the human *heart*. *Rel.* 6:29-30.
progression from bad to better is not enough since the deeply rooted depravity within will constantly work against this goal. Thus a total transformation of our Gesinnung is needed and this transformation is said to be the same as a change of the supreme inner ground of the adoption of all the human being’s maxims. Through this act we will get a completely new ground (a new heart) which is now unchangeable. (Rel. 6:51)

Now it is noteworthy that Kant underlines that while there can be no assurance or conviction of this change of Gesinnung that must be posited in the general transformation, we must nevertheless hope that such a change is possible through our own powers. Kant’s line of argumentation here is similar to the way by which he elsewhere establishes the ideal of the highest good by appealing to rational faith and practical postulates. The moral law, Kant says, presents us with the end consisting in “a moral disposition in its complete perfection”. He adds that since this is the ideal of holiness, it is “not attainable by any creature”. Still, we are told, this ideal “is yet the archetype which we should strive to approach and resemble in an uninterrupted but endless progress”. (KpV 5:83-4; 5:122) Accordingly, what seems to be possible for us is exactly the constant approximation to the archetype which we, however, will never resemble completely.

The fact that we will never be able to actually reach the state of holiness is compensated by the fact that Kant obviously seems to allow for a less radical turn consisting just in such an approximation, which also is said to replace the change of heart. As we saw already in chapter 3, we may become better human beings by becoming good merely according to our empirical character (virtus phaenomenon). For this, no change of heart was needed but only a change of one’s commitments to the law’s demands. We are told that the one who really penetrates into his first maxim, which is here called “the intelligible ground of his heart”, will realize that holiness requires an endless progress accompanied with a love of God. Now Kant adds that this bare insight is the same as actually being a good human being and even this can, while not actually being a full revolution, nevertheless be considered a revolution. Although this is far from the radical rebirth necessary for becoming a truly moral being with a flawless heart, this change nonetheless entails that the agent subjects himself to an eternal striving for the better, and hence this is seen as a gradual reformation of the propensity to evil, of the perverted Denkungsart. (Rel. 6:48)

Accordingly, through this mere gradual but eternal transformation of his hitherto corrupt ground, the human being takes the first steps towards the establishment of a character that is the true task for moral education since only this method affects the root of man’s general behaviour and not just (as in the case of the person possessing a virtus noumenon) particular deeds. (Rel. 6:48) We thus cultivate our predisposition to the good by regularly following the good example of morally good people. And so, Kant says, our
good predisposition gradually becomes a Denkungsart and the human being has become steadfast in his commitment to the law. (Rel. 6:48)

In light of this, one may conclude that this is just another expression of what is said all through Kant’s practical theory. We can never escape the fact that we are limited beings who need the rule of a law which, because of the respect it brings with it, makes us aware of the insignificance of subjective wants and makes us want to realize the highest perfection. As Kant said already in the second Critique, no creature can ever reach the purest stage of moral Gesinnung, for this would require that he be able to attain a state in which he no longer felt any temptation to deviate from the law. The only requirement that can be put forward here is to strive towards such a Gesinnung. (KpV 5:83-4) Here Kant also added that while for a perfect being, the moral law is a law of holiness, it must, for a finite rational being, always be considered a law of duty. (KpV 5:82) Accordingly, the conclusion that we all might be endowed with an evil first ground – whatever it is called – does, after all, not express anything else than the fact that we are not angels but humans. And, we must not forget that even without a “revolution” or a genuine “change of heart” we may proceed, little by little, from bad to better. This is, as we have seen, also what Kant generally has put forward as the ultimate vocation of the human species.

4.2.4. Gesinnung, Hang, “Germ of Good” and Personality

Although the moral change that we ought to perform turned out to be less suspect than first thought (since it is merely an idea), there are still some things that may cause troubles for those who want to map out the concrete picture of Kant’s idea of the moral mind. The most obvious question mark concerns the location of this change and now, if not before, the complexity of Kant’s theory of human nature becomes inescapably obvious. As we saw, Kant explicitly said that the moral regeneration consists in a “revolution of our Gesinnung”. (Rel. 6:47) At the same page, however, it is said that this revolution must take place in our Denkungsart and this move must be accompanied by a gradual reformation of the Sinnesart since the latter (as being the expression of our empirical nature), often places obstacles to the former. (Rel 6:47-8)

Obviously, within a few lines, Kant locates the necessary root of morality both to the Gesinnung and to Denkungsart. Moreover, he also equates the “propensity to evil” with the “perverted Denkungsart”. (Rel. 6:48) Furthermore, since Denkungsart is here identified as a propensity, all these three notions seem – at least in this context – to express more or less the same thing. To make the whole thing even more complicated, one may recall the concept of predisposition [Anlage]. As we have seen, Kant denied the existence of evil predispositions in human nature since all three predispositions, including those that concerned our non-rational nature, are not only harmless
but also “predispositions to the good”. As we probably know by now, evil occurs only by our misuse of these predispositions, a misuse caused by the influence of our propensity to evil. As he put it in some early anthropology notes, human evil “is nothing else than animality combined with freedom, to the extent namely, that freedom is not brought under a law”. (VA [Pilau] 25:844) The distinction between predispositions and propensities seems to be important for the predispositions are innate and essentially belonging to us. This is the reason why the way out of our “fallenness” is not a question of acquiring goodness but only of fighting evil.

This shows the importance of the germ of good, which in the Religion was never directly defined but which Kant elsewhere had identified with the good will. (Refl. no. 1426 15:622) In the Religion we are only told that this germ could by no means be self-love since self-love, if adopted as the general principle of all our maxims, is precisely the source of all evil. (Rel. 6:46) If we now recall the definition of our highest predisposition (to personality) we see that it is said to be the “susceptibility to respect for the moral law as of itself a sufficient incentive to the power of choice”. (Rel. 6:27) This signifies that this predisposition is more or less the very opposite of the third form of our evil propensity, depravity, understood as the propensity to subordinate the moral law to inclinations (i.e. self-love). Hence, it seems not too far-fetched to say that the germ of good could consist in this predisposition which, as we have seen, bears striking similarities with the good will.

So far there is no problem with Kant’s idea of the predisposition to morality as something distinct from man’s moral propensity if the latter is understood as the propensity governing (or at least heavily influences) our moral choices with respect to the ordering of incentives in our maxims. However, the last lines of his argument shows that the idea of the moral law, along with awareness of the respect that this law brings with it, cannot – properly speaking – be a predisposition to personality but it is rather personality itself. Then he adds that “the subjective ground of our incorporating this incentive into our maxim” can be seen as an addition to personality and this addition can thus be seen as “a predisposition on behalf of it”. (Rel. 6:28) In other words, the “predisposition to personality”, that is, our “germ of good”, is thus the subjective ground of our incorporating the moral incentive into our maxim. (Rel. 6:28) Such a ground of incorporation is, as we have seen, a “first ground” or “constitution” of our faculty of choice.

Now, if Kant’s discussion of the predisposition to personality as a predisposition to good is to work, he must mean that the “subjective ground” operating in this predisposition, (i.e., the "subjective ground of our incorporating incentives into our maxim") is something that can be good. While this is certainly a good thing (and the necessary condition for moral acting) it becomes hard to make sense of Kant’s apparently important division between predispositions and propensities since they, in the end, actually seem to play the same role in our moral life by making us disposed to choose one incen-
ative at the expense of another and thereby constituting our moral selves. The sole (but crucial) difference between these grounds is that, for some reason, the human being’s moral propensity is said to be unavoidably evil while the predisposition must be considered good in order to make moral acting possible. Moreover, if one adds to this discussion the hypothetical “propensity to good” as outlined above (i.e. 4.1.2) the task of separating these inner “grounds” of human nature becomes even harder since such a propensity to good was said to be the condition of the possibility of a character which, in turn, is the ground of the morality in the human being. In the Religion, it was the predisposition to personality that, through cultivation and repeated acting in accordance with the good principle, turned into a Denkungsart or a character. (Anthr. 7:329, Rel. 6:48)

While it is not easy to understand the true difference between these aspects, or powers, of the human mind, Kant’s wish to separate them is wholly understandable since without this tension within the human soul, man would be either holy or a mere animal. Again, in the light of the proposed rigorism, the moral propensity can only be one and thus we need to attribute an evil one in order to explain why our maxim-making faculty has a characteristic or constitution that sometimes makes us reverse the moral order of incentives.216 It is clear that Kant believed that this characteristic can be evil and so our choices will be evil, or at least they will be evil if we do not exercise our rationality. So, it might be the case that it is the status of a “characteristic” or “constitution” that distinguishes the propensity as well as the Gesinnung from the predisposition to personality since the latter might be thought of as the possibility of breaking through this characteristic and thereby acting contrary to it.

4.2.5. Gesinnung, Character and Denkungsart

The standard interpretation of Gesinnung as the highest “meta-maxim” gains all its support from the Religion. Still, one need only look to Kant’s corpus in order to see that he treated Gesinnung as one of the most important aspects of the moral being long before the publication of this work. We may again recall Collins’ notes were we learn that “ethics refers exclusively to Gesinnung” and moral actions must thus be judged on the basis of the inner goodness (or evil) of this Gesinnung. (VM [Collins] 27:299) Since it seems clear that our Gesinnung can be good, the straightforward identification of Gesinnung with our propensity to evil cannot capture the whole view. Accordingly, we need to see if there are further possibilities.

Throughout Kant’s writings, Gesinnung is explained as something in the depth of the human soul. It is generally qualified in normative terms as vir-

216 See e.g. Rel. 6:25 and 39f. In DiGiovanni’s translation, both terms (characteristic and constitution) correspond to Beschaffenheit.
tuos, vicious, noble, good or evil. Kant repeatedly claimed that morality consists in the Gesinnung – in the inner nature of the being that is also sometimes called character. Moreover, we have also seen that the subjective ground through which man adopts good or evil maxims (i.e. our propensity or Gesinnung) expresses the character of his species. (Rel. 6:21) However, as we saw in chapter 3, the character of the species does not fully reflect the character of the person and one should thus not conclude that Gesinnung, as expressing the character of the species, is something that must necessarily reflect the full scope of the character of the person.

Very often, the terms Gesinnung and character are simply used interchangeably. This is shown by the fact that authors often write “character” and adding in brackets “Gesinnung” and “Denkungsart”. Alternatively, they insert a note in which these terms are said to refer essentially to the same thing. With respect to the general discussion of Kantian agency, there is no real problem with this treatment of the terms since Kant often alternates between them, or assigns them the same function. In Kant’s anthropological works we generally find character defined as “a ruling principle” of the mind and this basically coheres with the technical definition of Gesinnung that we found in the Religion as a kind of second-order maxim that controls every choice and thus serves as the ruling principle of the rational beings free will. Both in the Groundwork and in the second Critique, Kant frequently says that morality must have its root in a good character (or Denkungsart) so both character and Gesinnung seem to be treated as the deeply rooted ground which, in contrast to temperament or a temporary mental state (mood) is essentially steadfast. This description basically corresponds to the lexical one. For example, in Adelung’s Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart (1793) we find Gesinnung defined as “the examined and contemplated resolution on one’s behavior.” This is a definition that describes Gesinnung in terms of a trait of our soul (or personality) of which we are well aware, and as something that is said to be steady in its nature. In other prominent lexicons it is said to be the German correspondent to Latin terms such as animus, mens, ingenium, habitus, indoles and natura – that all relate to the inner disposition, or nature, of a human being. This definitely corresponds to Kant’s usage of the term within his works on moral philosophy. In her book Kant’s Conception of Human Character, Munzel observes that Kant equates Gesinnung with the Latin term animus, and she

argues for the plausibility of this understanding by stressing the difference between *animus* and *anima*. According to her, the difference is that while *anima* is the soul “as a principle of life” and is an entity that in German would correspond to *Seele, or Geist*, *animus* is the soul as a principle of intellection and sensation.\(^{219}\) Still, the appeal to Kant’s insertion of Latin terms is not a reliable method since *animus* is more often used to denote Gemüt and it is also used for Denkungsart, and at least the former is generally distinguished from Gesinnung and treated as a complement to it.\(^{220}\)

This leads us to another possible dilemma for commentators who search for order and consistency within Kant’s terminology. While there seems to be support for the idea that Kant saw Gesinnung as the expression of the moral nature of our character, there are still arguments that can be raised against this interpretation. Critics of the idea that Gesinnung and moral character should express the same quality of human soul commonly draw attention to frequently referred to problem of *rarity* of character. In short, they argue, while Kant’s discussion in the *Religion* (as well as elsewhere) makes clear that we all have (and must have) a Gesinnung, only a few of us have a character. Hence Gesinnung and character, or Denkungsart, cannot possibly be the same. Still, even critics of the identification of Gesinnung with character, or Denkungsart, acknowledge that these notions are at least related. Robert Gressis for example, sees Gesinnung and Denkungsart as two *senses* of character and says that the connection between them is that the moral status of our Gesinnung (which everyone has) determines what kind of Denkungsart we might have, or may acquire. In order to have a Denkungsart at all (whether good or evil) our Gesinnung has to be good.\(^ {221}\)

To this line of argumentation one may reply that (as was shown in chapter 3), Kant’s rarity-claim with respect to Denkungsart only applies to a very specific form of it and hence not to Denkungsart in the sense of a faculty of making maxims or of a way of thinking about moral principles in general. As we saw, all human beings have a Denkungsart in all senses except for the Denkungsart that is the outcome of a total revolution. Because of this, and as a result of Kant’s explicit (as well as implicit) claims in the *Religion*, we cannot escape the fact that the “ground” in which all our maxims have their origin is referred to as both “Gesinnung” and “Denkungsart. Accordingly, once again we have to accept that Kant, after all, did not provide us with the strict scheme that we would need in order to find a consistent way of explaining how he might have understood the powers of the human mind. However, before leaving the exegesis of Gesinnung I will say a few words

\(^{219}\) Munzel (1999) xvii. For Kant’s identification of Gesinnung with *animus*, see *MS* 6:477.

\(^{220}\) For Kant’s use of *animus* as Gemüt, see *Anthr.* 7:131, 7:252; *OP* 22:112, 22:484 and for Denkungsart *Vorarb.* [*MS*] 23:409.

\(^{221}\) Gressis (2013).
about two interpretations that seem to allow for a non-uniform understanding of the term.

4.2.6. Gesinnung and Maxims / Early vs. Late Conception of Gesinnung

In order to make Kant’s discussion of Gesinnung comprehensible and consistent one must conclude that, as in the case of character, Gesinnung is used in more than one sense. Two readers who argue for this idea are Manfred Kuehn and Richard McCarty who both deny the standard interpretation of Gesinnung as the meta-maxim that rules the choice of all specific first-order maxims. Moreover, none of them seem to embrace the idea that Gesinnung is simply the same as character. I will here briefly summarize their alternative attempts to explain what Kant may have meant by Gesinnung.

McCarty begins by pointing out the central role that Gesinnung has for Kant’s theory of evil and adds that the term is usually taken to express something like a “disposition”. However, he says, in Kant’s earlier works, like the *Groundwork* and the second *Critique*, Gesinnung is used as more or less equivalent to “maxims” or “intentions”. Accordingly, he adds, at least in some contexts Gesinnung can be understood simply as “intention” without any significant loss of meaning. In view of that, McCarty says, Kant thus equates the phrases “moral disposition” and “moral intention” and assigns them a “normative, or an honorific sense” and this makes it “oxymoronic” to talk about them as evil.222 However, McCarty says, in later works, and especially in the *Religion*, Gesinnung is given a new and less normative but more descriptive meaning. In this work, Gesinnung is said to be something like the “disposition” that expresses the human being’s attitude towards the moral law. In this sense, McCarty says, the expression “an evil moral disposition” makes sense. Since he does not accept the standard interpretation of Gesinnung as our highest maxim he suggests that Gesinnung should rather be understood as the “ranking” of our natural predispositions (i.e. animality, humanity, and personality) as incentives. A good Gesinnung would simply subordinate the two former to the third (moral) one, whereas an evil Gesinnung would subordinate morality to the incentives of some of the former. Given this, one may consider Gesinnung as the *form* of the highest maxim, which is the “the first subjective ground of the adoption of [subsequent] maxims”.223

Also Manfred Kuehn sees a significant change between an early use of Gesinnung and the way Kant talks about this term in the *Religion*. Like McCarty, he calls attention to passages where Kant obviously seems to treat maxims and Gesinnungen as more or less equivalent, but he adds that Gesin-
nung is be better understood as “the motivational aspect of maxims” or more generally as what is “subjective” in this “subjective principle of volition” or put differently the “internal principle of the maxim”.\textsuperscript{224} Accordingly, he says, Gesinnung is not the same as character since Gesinnung does not add anything new to Kant’s discussion of particular moral agents. As a result of this, all scholars who see character and Gesinnung as being equivalent are mistaken; yet he states, character is the same as Denkungsart.\textsuperscript{225}

In reply to these attempts I would say first that McCarty and Kuehn are not totally mistaken for Kant, at least sometimes, talks about maxims and Gesinnungen in this way. This understanding would also cohere with the definition suggested by Baumgarten of which Kant undoubtedly was aware.\textsuperscript{226} Still, it should be noted that while McCarty gives several references to passages where Gesinnung and maxim are related, none of them directly supports the view that “Gesinnung” actually is (or was) merely another way of denoting “maxim”. At these passages Kant either says “Gesinnungen and maxims” (KpV 5:56 5:327) or “the Gesinnung of the maxim” (KpV 5:84) and “moral worth as a Gesinnung by its maxim” (KpV 5:159). Accordingly, while these passages indeed indicate some kind of close connection between maxims and Gesinnung(en) they do not express a strict identity between them. Admittedly, there is one passage in the Groundwork in which Kant actually says “in the Gesinnungen, that is, in the maxims of the will” (GMS 4:435) but, as we will see, this does not necessarily mean that they have to be regarded as synonymous.\textsuperscript{227} Rather, it can also be taken to mean merely that the Gesinnung somehow belongs to the maxim or, as the standard interpretation says, it is the meta-maxim that caused (or chose) this ordinary action-maxim. Again we may recall that, while the worth of an action depends on the maxim, the worth of the maxim depends on the Gesinnung, and this Gesinnung, finally constitutes the moral worth of the person.

If we continue to the alternative idea of “ranking”, I think that McCarty is more or less correct when he says that Gesinnung is something that governs the process by which our faculty of choice incorporates incentives into our first-order maxims. Yet, I must say that it is hard to see how this alternative idea of Gesinnung actually differs from the standard interpretation. First one must conclude that his rejection of the maxim-interpretation is not only incorrect but also unfair since he suggests that Allison (who he sees as the chief proponent of this interpretation) may have been led to this idea by what Kant says at 6:25, namely that our Gesinnung, in order to be imputable, must

\textsuperscript{225} Kuehn (2001) p 369. See also note 150 pp. 499f.
\textsuperscript{226} See Baumgarten’s Metaphysica §699 where maxima is explained as someone’s “gewöhnliche Gesinnungen”.
\textsuperscript{227} Also Harald Köhl draws attention to the sole occasion in the GMS where Kant identifies Gesinnungen with maxims. Köhl p. 55.
have been adopted. According to McCarty, Allison seems simply to believe that the mere fact that our Gesinnung is adopted implies that is must be a maxim (suggesting that Allison sees maxims as the sole thing that can be adopted by a choice).\textsuperscript{228} McCarty thus ignores that Kant at 6:21 says that if the subjective ground of the adoption of this maxim “were ultimately no longer itself a maxim, but merely a natural impulse, the entire exercise of freedom could be traced back to a determination through natural causes – and this would contradict freedom.” Again, the subtle differences between “subjective ground of the maxim”, “subjective ground of the adoption of the maxim” “subjective ground of the exercise of freedom” and so on cannot be taken as convincing arguments for a separation of them since they in the end, because of the overall argumentative line in the Religion, seem to amount to the same thing, namely the ground in our nature that affects our choices of both maxims and ends.

Now, in order to return to McCarty’s idea, we may recall that he held that the highest, or supreme, maxim (from which all other maxims are derived) is “a maxim that both incorporates and orders multiple incentives”.\textsuperscript{229} Now, the identification of Gesinnung with the form of the maxim and thus with the process of ranking our predispositions – whether one shall give priority to self-love (animality or humanity) or to the law (personality) – is exactly the definition of the propensity to evil. This is clear from the fact that in all human being the first ground of all maxims is necessarily a propensity to evil. From this it would follow that the ranking would generally subordinate personality to the two former and if this rankling is our Gesinnung, then we once again see that Gesinnung and Hang cohere. This conclusion is further supported by the fact that Kant identifies the propensity with the formal ground of every action contrary to law. (Rel 6:31) Moreover, his claim that Gesinnung would be the form of the highest maxim further strengthens the conclusion he wants to escape since, as we have seen, the fundamental ground of maxim-adoption is the “formal ground” of all particular deeds. (Rel. 6:31) For that reason, in spite of his intentions, McCarty’s idea does not seem to be anything but just another – indeed illuminating – way of explaining the standard interpretation.

Finally, it should be said that the idea that the various conceptions of Gesinnung visible in Kant’s corpus can be regarded as a difference between early and late usage of Gesinnung is not supported by Kant’s own texts. Passages that identify Gesinnung with maxims do not only appear in the Groundwork and the second Critique but they are also to be found in later works (e.g. the draft for the Metaphysics of Morals).\textsuperscript{230} Conversely, Kant

\textsuperscript{228} McCarty p. 206 n. 9.
\textsuperscript{229} McCarty p. 206.
\textsuperscript{230} Vorarb.[MS] 23:258. Interestingly, at this page, Kant (perhaps unintentionally) defined Gesinnung both as “maxim” and as “the ground of the maxim”.

158
definitely uses Gesinnung in the supposed novel “Religion-sense” already in the Critique of Practical Reason. Moreover, since the Religion’s definition of Gesinnung is said to be the “the subjective ground of the adoption of maxims” this reading basically agrees with Kuehn’s idea that according to the so called “pre-Religion meaning”, Gesinnung can be understood as “the motivational aspect of maxims” or more generally as what is “subjective” in the “subjective principle of volition”.

4.3. Summary

The Religion is a complex work that involves everything from a version of the theological doctrine of original sin, the idea of the highest possible good as the end of our endeavor, metaphysical questions of timeless existence and, perhaps most notably, the ethical query into the nature and origin of human evil. Besides this, it provides us with a deepened account of the human faculty of choice and, perhaps most importantly, we here find some of Kant’s most genuine inquiries into moral anthropology. It is no wonder that scholars have struggled with the many approaches that are not easy to reconcile. Still, in the preface to the second edition of the text, Kant himself says that the morality given in this work is to be understood according to ordinary morality without any venturing into a critique of practical reason or even less into any critical elaborations of its theoretical use. We only need to grasp the difference between virtus phaenomenon and virtus noumenon, that is, between actions conforming to duty and actions originating in a steadfast Gesinnung directing us towards duty. (Rel. 6:14)

While the Religion is usually not considered to be one of Kant’s foremost moral works, some central definitions are stated here which affect the understanding of his practical philosophy in several respects. Besides the formulation of the incorporation thesis, that, by stressing the requirement of an active choice of an incentive, definitely established the idea of moral responsibility for one’s actions, the Religion is the starting point for several seemingly indeterminable debates. However, I believe that one can usefully highlight at least one main cause to several of the disagreements with regard to Kant’s discussion of the origin of human evil. As was mentioned in chapter 3, and which has been demonstrated above, attempts to establish the definition, or

231 This is clear at 5:84 where Kant talks about Gesinnung as the basis of our maxims.
232 Guyer denies that Kant’s idea of evil as radical in our nature is to be seen as a version of original sin simply because he believes that we could, in fact, have chosen good instead of evil. Our evil or “crookedness” is thus not necessary. (Guyer 2009 p. 147) To this, one may reply that the idea of original sin does not rule out the possibility that we could have made another choice. The central thing is that we in fact did choose evil. We could certainly have refrained from sin and thus have stayed in Eden. Still, we chose evil; for Kant, the first choice was a deed, a peccatum originarium, or original sin.
the relationship, of Kant’s notions applicable in all contexts will rarely turn out to be successful. Kant’s discussions are far too complex for that and his sometimes imprecise, or inconsequent, use of terms further contributes to the impossibility of establishing such indisputable definitions. I will end this chapter by summing up my opinion on the most important disagreements.

Definitions and relations: Upon initial contact with Kant’s theory of evil one is immediately confronted with an exceptional conceptual confusion. While the Religion provides us with the best-known suggestion of how one is to understand the meaning of Gesinnung (as a maxim for maxim-adoption) a careful reading reveals that the adoption of this idea is far from trouble-free. If left unsolved, or simply ignored, these problems, which by no means ought to be regarded as merely terminological issues, will seriously affect Kant’s general moral discussion – not only his argument in the Religion.

As to the Gesinnung-Hang problem I say that if one is to hold on to the standard interpretation of Gesinnung as the supreme maxim, one is also bound to accept that this maxim, the highest ground for the adoption of all further maxims, actually expresses the same entity in our nature that Kant called the propensity to evil. Textual evidence in the Religion makes this conclusion inescapable. Still, I agree with those who insist that it would be much easier to comprehend Kant’s project if they actually were two distinctly separated concepts that operate independently in the human faculty of choice – one evil [Hang] and another – at least possibly – good one, that is, Gesinnung. That said, I certainly do not agree with those who claim that the distinction between Hang and Gesinnung in the Religion is clear or obvious. Moreover, arguments for keeping them separated are generally based on the assumption that a distinction it is necessary for a correct understanding of Kant’s intentions. Yet, such interpretations require one to ignore several of his most central arguments. Fortunately, there are other ways of dealing with the dilemma.

A vital ground for the dissatisfaction with the Gesinnung-Hang identity is that Kant so often equates Gesinnung and character and, as we have seen, character per se is not necessarily an evil ground corrupting our choice. Still, character is a foundation of the human being’s way of thinking; or more precisely, a highly specific quality of one’s way of thinking that the agent has established within himself by making his principles firm. This will at least rule out the two first stages of evil that arise from the propensity since neither weakness nor impurity express any firmness but rather the opposite thereof. Not even depravity, the highest degree of evil, can be said to fully correspond to an evil character since, although it frequently and actively reverses the order of incentives, the principles following from it are generally not firm enough to pertain to a character.

Because of all the obscurities involved in Kant’s discussion, I am sceptic to the idea that it is possible, as Alison Hills puts it in her analysis of these
notions, “to reconstruct the best account of what Gesinnung is and the most plausible explanation of its ethical role, on the basis of Kant’s remarks about it in the Religion and elsewhere.”233 Since Gesinnung is not even uniformly used within the Religion, one must rather interpret it in each particular context and accept, as in the case of character, that one cannot simply establish what Gesinnung is but only what the term denotes in each circumstance. Neither can one reasonably say that there is a shift in the usage between certain works (or at a certain time) since the different varieties tend to appear throughout. In other words, it is equally correct to say that Gesinnung is “the highest ground of maxim-adoption which expresses the character of the species” and to treat it as another synonym to the character of a particular agent constituting his or her personal moral worth. In the latter sense, our Gesinnung can be either good or evil. In light of the just mentioned significance of not blindly believing in the possibility of fixed and universal definitions, I will once again underline the need to be aware of the complexity and ambiguity of terms like character and Denkungsart, and now also add that this caution should also be used with regard to Gesinnung. Gesinnung is undeniably equivalent to character in one sense but not in another. Yet, as long as one is aware of this and does not try to reconcile these different views, this should not be a big problem for the general discussion of Kant and his theory of the origin of good and evil.

**Change of Heart:** With regard to Kant’s outline of a “change of heart” one must admit that his talk is fairly turgid and aims to illustrate the escape from the world that “lieth in evil”. As we have seen, it has been argued that the possibility of such a change requires that we separate the propensity from the Gesinnung (or character) since, because of the strict innateness and inextricability of the propensity, the change must apply only to the Gesinnung or character. Yet, because of what we have learned about the relation between these features of our corrupted heart, this separation seems not to be possible.

However, as I have argued, the inextricability of the propensity to evil is not to be taken as evidence for a need of separating this propensity from Gesinnung, for the change required for moral perfection – if taken in its most radical sense – actually reverses our entire heart (inner nature) and consists in a *total rebirth* or a “new creation” from which an entirely new man emerges who is entirely free of the taint of evil grounds. While this move is, to say the least, radical, Kant says that we must *hope* for the possibility and that we may thereby *posit* is as attainable. Yet, I believe that this change of heart is on the same footing as the claims about our pursuing the *highest good* whose possibility also was secured by the postulates and further encouraged by our faith. What we, as human beings can do, is to endlessly proceed towards it.

---

233 Hills p. 80.
The Missing Proof: My skepticism about reconstructions that simply aim to “save” or “improve” by adding things that are actually missing also applies to the discussion with regard to the “missing proof”. While such suggestions may be both interesting and also push the discussion forward, it is clear that most such attempts are trouble-laden. Moreover, is impossible to know whether Kant would have accepted them as accurate reconstructions of his position. I am well aware that my proposal – that Kant’s claim that we can spare ourselves of the proof is a mere rhetorical statement that originates in his fondness of Rousseau (and almost exactly resembles Rousseau’s saying) – does not replace an actual formal proof but it might nevertheless explain Kant’s intentions. Again, given the Rousseauian tone of this work it is also not a totally implausible suggestion. Kant would probably not have written the Religion in the way he did without Rousseau’s ideas at hand. The very idea of fallen beings who have been thrown out of the happy state parallels the Rousseauian theodicy expressed in the Discourses.

Still remains the proof that would have made Kant’s idea complete or, which at least could have spared commentators much confusion. This will not be provided here but I will return to the topic of evil and Wood’s deduction of it as essentially being brought about by ambition in chapter 6. Again we need to dig deep into Kant’s terminology and, as we will see, there is much more that needs to be understood with respect to the development of the passions and vices that cooperate, or check each other, in the social human being’s mind.

Despite some remaining obscurities with regard to Kant’s idea of the origin of good and evil I believe that we at least have been able to sort out a little of the mess that has hitherto obscured the discussion and made Kant a moral pessimist. Yet, there is still much that need to be considered with respect to his idea of the moral being.
Until now, the discussion has focused on the principles that explained the origin of good and evil in Kant’s system. In this chapter I will devote the greater part of the analysis to a notion that according to Rousseau is claimed to be the ground of all awareness of goodness, justice, truth, virtue, humanity and commiseration as well as sincerity and our connection to the higher order of things. Besides being a controversial aspect of Rousseau’s thinking, this discussion also aims at revealing the intricate relationship between sentiments and reason in Rousseau’s moral psychology. This will be required for the comparative discussion of chapter 6 and 7.

In the First Discourse – a text written as an academic prize essay on the topic: “Has the restoration of the sciences and arts tended to purify [or corrupt] morals”? – Rousseau instigated his analysis of man’s cultural evolution. The text was called Discours sur les sciences et les arts. By adding the alternative “corrupt” he was able to highlight his disapproval of artificial influences. Unsurprisingly, Rousseau denied that the arts and sciences could be regarded as a means to moral refinement since what we learn from them are often products of social prejudices and corrupted ideals. Blinded by the deceptive images of our new inventions that only “spread garlands of flowers over the iron chains with which men are burdened” we learned to love our slavery. Even if culture occasionally might teach us some useful principles, this is not an education that we need. Rousseau’s final words in the thesis read:

O virtue! Sublime science of simple souls, are so many difficulties and preparations needed to know you? Are not your principles engraved in all hearts, and is it not enough in order to learn your laws to return into oneself and listen to the voice of one’s conscience in the silence of the passions? (ID 22/30)

This comment leaves no room for questioning Rousseau’s idea of the natural goodness of the authentic man, an idea that he, in spite of the condemnation of the work, developed in his Second Discourse and Emile. Rousseau’s main thesis was that man understands by nature how he should behave; he only has to learn how to make use of his natural skills and avoid everything that

---

234 ID p. 5/6. Cf. here the famous opening of the Social Contract “Man is/was born free but is everywhere in chains” and the passage where he discusses slavery: “slaves lose everything in his chains, even the desire to get rid of them”. SC II:i.
may lead him astray. In every decision, he must only listen to his true self, for in the very depth of his soul he will find the pure voice of nature which is also the voice of God. This idea reaches its culmination in Emile:

Conscience, conscience! Divine instinct, immortal and celestial voice, certain guide of a being that is ignorant and limited but intelligent and free; infallible judge of good and bad which makes man like unto God; it is you that make the excellence of his nature and the morality of his actions. Without you, I sense nothing in me that raises me above the beasts, other than the sad privilege of leading myself astray from error to error with the aid of an understanding without rule and a reason without principle. (E IV 290/601-2)

The beauty and profoundness of these lines is striking, and it is easy to be impressed by Rousseau’s rhetorical ability. Conscience makes us sociable and humane; it raises us above our brutish nature and talks to us directly from our heart. It is a “sublime emanation of the eternal substance” and a principle that cannot err, since the pure movements of nature are always right. Conscience expresses the love of order; it is a principle of justice and virtue by which we alone are able to judge our own actions as being good or bad. For this, reason itself is an insufficient source; it is conscience that makes us love the good. (E IV 290/600)

Before continuing to the analysis of this central notion it may be beneficial to make some remarks concerning the terminology. First of all, it should be mentioned that the French term conscience means both conscience and consciousness. Yet, in the case of Rousseau, the risk of conflation is generally said to be minimal.235 One reason for this is that Rousseau rarely speaks of the human consciousness simpliciter but normally in terms of our consciousness of something. Hence, when he speaks of conscience it is adequate to assume that he refers to the instinctive voice of our uncorrupted soul that becomes active when we have reached the state of reason and developed some kind of moral thinking. Indeed, as MacLean points out, Rousseau held that the spirituality of the human soul [spiritualité de son âme] is revealed from the consciousness of his freedom [conscience de cette liberté] (2D I 26/142) This, MacLean says indicates that conscience, as being the voice of reason or of God, can be understood in terms of the consciousness of freedom.236 It is clear that conscience requires that we have reached some kind of consciousness of our spirituality, as well as of good or bad or, at least, of right or wrong. Yet, as we will see, conscience is not merely knowledge or awareness of what is good or just but also love of what is good or just. In fact, we are explicitly told the acts of conscience are not judgements but sentiments. [Les actes de la conscience ne sont pas des jugements, mais des

---

235 Cf. Coper p. 83 n. 25 where he says that Rousseau’s general usage of the terms is clear enough to avoid misinterpretations due to this dual meaning to the French word.

236 MacLean pp. 85-6.
sentiments]. (E IV 290/599) As such, it is wholly independent of reason and all derived principles even though it, as we just saw, would not develop without them.

In view of that, we see that there is another term that may be of special interest within this discussion, namely sentiment [Fr. sentiment]. While Rousseau sometimes seems to regard conscience as a principle [Fr. principe] he is generally speaking of it as a sentiment. By calling it a principle, conscience is likened to other mental features that are also called principles of the soul, like pity and amour de soi. All these are said to be innate or, at least, derived from other innate grounds. As principles, these features rule our attentiveness and thinking in general.\(^{237}\) Now, by also assigning conscience the role of a sentiment Rousseau highlights its nature as an immediate response or awareness. As we are told in the Dictionnaire de Jean-Jacques Rousseau, for Rousseau, the term sentiment covers almost everything in our emotional life, like passions, feelings, and various “movements”.\(^{238}\) In all its shapes sentiments are immediately aware or felt – either as reflected or instinctive. As we have seen, the most original feeling that man experiences is the sentiment of his existence [sentiment de son existence] that, in its most basic sense, is stripped from all further affections and earthly impressions.\(^{239}\) As to conscience, we may assume a similar force; our conscience will instantly start to speak to us due to its immediate awareness of the divine order. Yet, there are several objections that can be raised against Rousseau’s way of ascribing this divine and infallible instinct to man and at least some of them will be brought up in this chapter.

5.1. The Problem of Infallibility

Rousseau’s most familiar exposition of conscience is the above quoted passage from the famous section in Émile called “Profession of Faith of a Savoyard Vicar”. (E IV 285/595) This declaration has for several reasons generally been the starting point for discussions of Rousseau’s account of conscience. Although this may be his most powerful description, he insists through his whole philosophy that this feeling plays an essential role for man’s moral thinking. It is clear that, for Rousseau, conscience is man’s

\(^{237}\) It is true that also some acquired traits, e.g. amour propre is said to be a principle, or more precisely, a “principle of wickedness”. (RJJ II 100/789f) Yet, as with the other principles just mentioned, it is also referred to as a feeling, passion or sentiment.

\(^{238}\) Trousson & Eigeldinger p. 854. “Sous le terme sentiment, Rousseau englobe en fait toute la vie affective, les passions, les émotions, les mouvements; ce qui les réunit, c’est leur caractère s’intuition sensible, de présence à la soi”.

\(^{239}\) 2D II 66/193; RSW V 46/1047. For a lengthy discussion of this original “sentiment of existence”, see Gauthier.
most reliable principle; in contrast to reason, conscience would never lead us wrong.

One would, however, be naïve if one maintained that the many objections that have been raised against Rousseau on this subject are totally unfounded, and his outline is certainly controversial. After having quoted Rousseau’s famous definition of conscience (where we learned about its endless goodness, divinity, immortality and infallibility), Arthur Melzer says:

One must agree with Bertrand de Jouvenel regarding the striking novelty as well as the inherent implausibility of this doctrine. “it is to ignore the whole immense problem of the ‘erring conscience’...it is to suppose the infallibility of moral sentiment. Nothing could be more bold: it is to contradict ancient as well as Christian philosophy.” 240

It is obvious that this innate feeling is of greatest importance for Rousseau’s theory of man’s innate goodness. In light of this, one must wonder how human evil could arise. If conscience exists in every human being, if it, as Rousseau expresses it, is for the soul what passions and instinct are for the body, and if it really is an eternal principle directing us towards the good, then – why do so few of us listen to this gorgeous voice of nature? If we really love the good and the just, why do we so often choose the bad? Could it be the case that some people are short of this gift, or that conscience, after all, can err? Is conscience, like many of our other sentiments, corrupted by our social progress? These conclusions would obviously contradict Rousseau’s optimistic view, so I think that there are good reasons to analyse this problem.

As we saw in chapter 1, the general cause of man’s corruption is said to be amour propre. It is true that amour propre plays a role in man’s moral development but as the discussion here (as well as in the next chapter) will show, this passion cannot, by itself, explain any actual evil deed. Furthermore, what makes this particular problem interesting is that we are now confronted with the difficulty of the order of the development of powers and sentiments in the human soul. In the Letter to Beaumont, Rousseau makes clear that conscience would be “null in the man who has compared nothing and who has not seen his relationships”. (LB 28/936) Since it is amour propre that is generally described as the relative sentiment that makes us see our relations, Asher Horowitz concludes that conscience is necessarily, though not solely born from amour propre.241

For those who know merely the basic ideas of Rousseau, this suggestion may be confusing for several reasons. First, it seems to be at odds with his warnings for amour propre and its terrible consequences. Second, there seems to be some paradoxical statements involved already in Rousseau’s

240 Melzer, p. 147 (referring to B. de Jouvenel, Essai sur la politique de Rousseau, 1947).
241 Horowitz p. 142.
explanation of conscience as man’s most original principle. Rousseau’s outline in the Second Discourse seems to leave no room for sentiments other than amour de soi and natural pity and these are said to be the exclusive sources of all further principles of the human soul. These arguments require some comments. Although I agree that the presence of conscience requires that we have entered a state that involves interpersonal comparison in which amour propre probably has developed, I do not believe that this requires that we also must regard amour propre as a necessary condition for conscience. Rather, there is another, far more important sentiment whose role we must not neglect, but which Horowitz nevertheless seems to have overlooked.

I will address these questions in turn and I will begin by analysing the development of moral conscience and how it depends on other features of human nature. This problem is closely related to Rousseau’s idea of “education according to nature” and I believe that an understanding of the order of the development of these sentiments is of greatest importance when we will later evaluate the more serious problems surrounding the corruption, or potential imperfection of this inner judge. In order to meet the problems caused by the corruption hypothesis, we need a clear picture of these relations as well as of Rousseau’s moral psychology in general.

As to the idea of moral corruption, there are some more things that need to be said. According to Gauthier, the transition from solitude to social life is what Rousseau identifies as corruption. Although this step can indeed be seen as a “corruption” or, at least, as an alteration of man’s original nature, it is more common to understand “Rousseauian corruption” in terms of the change of our passions; a change that gave rise to excessive and artificial needs. These needs were born from comparison and were strengthened through the amour propre that followed from this relative view upon ourselves.

In an extensive footnote in his Discourse on the Origin of Inequality, Rousseau makes his famous claim “men are wicked; sad and constant experience spares the need for proof”. (2D 74/202) His immediate addition of “yet, man is naturally good” indicates that evil only applies to social men whose nature has been transformed by cultural prejudices. As Rousseau himself puts it, the theory of man as naturally good is not only an anthropological hypothesis but rather the fundamental principle of all morality. (LB 28/935) Even if one may have objections to Rousseau’s utterly pessimistic view of the effects of the present society, no one can reasonably deny the obvious fact that human beings sometimes adopt evil maxims and that they often do so without any sign of remorse. From this, one question naturally arises: what happened to our original modesty, to our basic love of self and to our conscience? Due to the fact that at least some of our basic sentiments change in an unfortunate way, one may be tempted to conclude that our new

---

242 Gauthier p. 3.
way of living also changed our conscience for, as we have seen, “everything is good when it leaves the hand of the author of things; everything degenerates in the hands of man.”\(^{243}\) In order to understand this moral collapse we must analyse the nature of our original sentiments and find the true cause of man’s degeneration. If it is not conscience that became corrupted, what happened? Which of our dispositions, sentiments or passions changed our nature?

On the one hand, it seems possible to argue in favour of the idea that our conscience cannot, strictly speaking, be corrupted or err (i.e. lead us towards evil or make us love the wicked); Rousseau makes several comments that make this conclusion impossible. On the other hand, the apparently best argument for the rejection of the possibility of conscience corruption seems to involve a problem. As a natural and original principle, conscience is tightly connected to other sentiments and, as we will see, it is the intellectual aspect of *amour de soi*, the passion from which all further sentiments and passions are said to be derived. Since *amour de soi* can apparently be affected in a corruptive way by being transformed into *amour propre*, a critic of the corruption hypothesis must show that this transformation does not affect the entire conception of it. In order to do this we must distinguish conscience from other sentiments and locate the weak link in man’s development towards a social, but still morally good and *authentic* human being.

### 5.2. Internal Order of the Sentiments

The aim of Rousseauian education is essentially to preserve as much as possible of our natural instincts. Yet, in order to proceed from a state of mere peacefulness to a state of genuine moral goodness or virtue, we must, as we have seen, extend the power of reason since only a rational and social being can recognize the rightness or wrongfulness of an action or event. It is from the awareness, the first right or wrong action, that the cry of conscience is raised. \((FM\ 143/225)\) This sentiment is described as the “love of order” that seeks the “well-being of the soul”. When this intellectual love is developed and made active, it bears the name conscience. \((LB\ 28/936)\)

The fact that conscience cannot be developed without reason and consciousness of one’s freedom excludes the possibility that such a moral guide ever existed as an active power in the natural man. The very idea of a natural principle that searches for the “well-being of the soul” certainly seems to disagree with Rousseau’s initial idea that the savage’s needs and wants could be reduced to nourishment, rest and female company. At the same time, conscience is said to be *original*, and so *prior* to all influence of education. It is not easy to see how these conditions could be compatible. If conscience is

\(^{243}\) *E I 37/245*. Cf. 1.2. above.
natural to an uncorrupted but rational being’s mind, rationality must be considered something that we have developed but which is not the effect of social constructions.

These alleged inconsistencies are the reason why Jonathan Marks refuses to accept the romantic idea of men as naturally aware of the principles of justice and he argues that the view expressed in the Profession (and elsewhere) is nothing but a “rhetorical strategy” or a “salutary untruth”. Our mind needs guidance and even the voice of conscience must go through an advanced form of training in order to be that good guide that Rousseau’s definition calls for. Marks’ own arguments are essentially based on the idea that the innate sense of justice will almost certainly bring about violent feelings in an immature and non-rational soul. His main argument is a “surprisingly neglected” passage in which he finds evidence that conscience is not at all infallible. The example illustrates how the little boy, who gets struck by his impatient nurse, responds by showing anger and violent tendencies. (E I 66-7/286-9) Since the reaction is said to be brought about by the child’s first “sentiment of just and unjust” [le sentiment du juste et de l’injuste] and since Rousseau defines conscience as the “first sentiment of justice” [premier sentiment de la justice]. (E IV 275/584) Marks concludes that the fury was caused by the child’s first sense of conscience and, he adds, that “Rousseau’s usage authorizes us to do so”. Accordingly, in light of this example, one must, Marks says, admit that conscience is not – as an original and innate principle – an infallible guide to the good; conscience does not, as we are told in the First Discourse, tell us the principle of virtue unless it has undergone an elaborate cultivation or “voice training”.

I will return to Marks’ example about the child and its reaction in section 5.3. For the moment I will concentrate on a problem that is inherent in this argument. Although I have several objections to this “extension” of the notion of conscience, I will use Marks’ idea since it points to the importance of understanding how and when our various sentiments are born and how they may depend on each other. If we consider Marks’ and Horowitz’ respective views, they both discuss the nature of our moral sentiments and primarily what these sentiments require in order to contribute to our moral training. In both cases we are confronted with the supposed contradiction that conscience is said to be innate and natural, and thus prior to social constructions but yet, at the same time, dependent on reason; a faculty that, as we have been told, was not activated until we saw our relations.

In order to evaluate these two interpretations, as well as the consistency of Rousseau’s own claims, we must show that sociability is not inconsistent with Rousseau’s idea of human nature. If the widespread view of Rousseau’s

---

naturally solitary man is correct, then it is difficult to defend the idea of a pure voice of nature that would guide us through life. As we have seen, Rousseau paints a picture of the isolated being that would be at peace with his existence, but we also saw Rousseau’s allowance for a day when we would realize that we are both weak [faible] and dependent [dépendant] beings. This is what lies behind not only our choice to associate with others but also our need of doing so. Because of our natural limited nature we were more or less predestined to cooperation. As we saw in chapter 1, Rousseau did not eliminate a social vein in human nature. In the *Moral Letters* we see:

> one cannot doubt, man is a sociable animal by his nature, or at least made to become so, he can be so only by means of other innate feelings relative to his species. And it is from the moral system formed by this double relation to oneself and one’s fellows that is born the natural impulse of the conscience.\(^{247}\)

Obviously, nature did not only give us our most basic instincts but also granted us our tender weakness and the need for becoming social beings. As we see, we are affected by other sentiments, such as “innate feelings relative to his species”. Moreover, as the quote tell us, it is from that moral system, “formed by this double relation to himself and his fellows that the natural impulsion of conscience is born”. Horowitz asserts that it is obvious what Rousseau has in mind. The “relative sentiment” in question cannot be anything but *amour propre*, and if conscience is born out of a natural sentiment relative to our species, it must have its origin in *amour propre*.\(^{248}\)

This idea is, however, problematic for at least two reasons. The first is that *amour propre* which, despite being he predominantly dangerous drive that Rousseau has described in terms of “the principle of all wickedness”, is now said to be the necessary source of an infallible moral principle.\(^{249}\) Although *amour propre* is not always bad, it is probably no coincidence that Rousseau’s discussion of these two principles differs significantly. A second and more problematic objection to Horowitz’ suggestion is that *amour propre* is, in fact, not “innate” unless we understand innateness in a very loose way that allows that everything resulting from our natural abilities is innate. It is clear that Rousseau believed that we will all acquire this relative passion, but it is nevertheless difficult to apprehend how something that is explicitly is said to be innate in the human heart, like conscience, should be derived from something that appears later in life, like *amour propre*. Here, it

---

\(^{247}\) *ML* V 196/1109. There is a more or less identical passage in the “Profession” (*E* IV 290/600) but since I follow Horowitz’ discussion, I quote the *ML*. In doing so, we also avoid the problem associated with the reliability of the arguments from the Profession. (Cf. again note 68 above).

\(^{248}\) *Horowitz* p. 142.

\(^{249}\) For the definition of *amour propre* in terms of wickedness [méchanceté], see *E* I 67/28 and *RJJ* II 100/789f.
might be argued that even conscience, in spite of its innateness, needs reason and awareness of external relations. A child or savage would not be able to use such feelings for any moral purposes since these feelings would, as Rousseau says, be null and inactive or, at best, only have negative force.\footnote{LB 28/936. Negative here means that compassion merely hinders us from harming others (this is Rousseau's definition of natural pity).}

The arguments about a “relative sentiment” and a “double relation” are not easy to understand, but I will nevertheless say that Horowitz seems to have overlooked one important detail. It is true that \textit{amour propre} is often defined as \textit{the} relative sentiment, and it is this sentiment that makes us aware of our position within a group of people. In dealing with the problem of human sociability we must, however, be more careful for the “relative sentiment” whose importance Rousseau stresses in the passage above is not necessarily \textit{amour propre} but it might alternatively be \textit{compassion} or \textit{pity}. Pity, Rousseau says, is “\textit{the first relative sentiment} which touches the human heart according to the order of nature.” (\textit{E IV} 222/505) As the being “first relative sentiment”, pity seems thus to be prior to \textit{amour propre} and, in contrast to this latter sentiment, pity is also explicitly said to be innate.

That Rousseau saw a close link between compassion and conscience is uncontroversial but there is great disagreement about the exact relationship as well as the order of their development. In his \textit{Rousseau, Nature and the Problem of the Good Life}, Lawrence D Cooper says: “in \textit{Emile} and with regard to civilized men generally, ‘pity’ stands for only one thing, the sentiment of commiseration. As such it is a manifestation of the principle of soul called conscience.”\footnote{Cooper p. 98.} Conscience is thus given the lead role in the sublimation of human nature and is said to be a necessary means for the possibility of a state of “civilized naturalness”. Still, it is far from clear how this process works. We are told that \textit{amour de soi} cooperates with pity in producing humanity and virtue in the human soul, and in the absence of unnatural passions this would be the only thing that we needed in order to be gentle and sincere.

Now, in contrast to the idea that everything good has its ultimate root in conscience Nicholas Dent says:

Rousseau does not see conscience as the exclusive source of moral understanding and moral motivation in human beings. He thinks that the sentiment of compassion, for example, can direct us towards generosity, clemency and justice without resource to the guidance of conscience (\textit{E IV} 235/522-3); interestingly, these sentiments probably have the same root in our nature as conscience has in Rousseau’s view.\footnote{Dent (1992) p. 60.}
I take Dent’s claim here as saying that conscience – much like generosity, the sense of justice and similar feelings – ultimately stems from our natural pity. This would be consistent with Rousseau’s early claim about man’s two basic principles which, as we have seen, are amour de soi and pity. However, in spite of this apparently plausible idea I believe that Dent’s reference to Emile actually defeats his own purpose. At the passage in question (i.e. E IV 235/522-3) Rousseau explicitly says that he will show how conscience awakes our sense of goodness and justice and that this kind of knowledge cannot arise from reason alone. Natural rightness would be a chimera if it was not founded in our heart. A better reference for Dent’s argument would be some of Rousseau’s extensive declarations in the Second Discourse where he actually speaks of pity as the passion of generosity and clemency without mentioning conscience at all. On the other hand, many of Rousseau’s statements in the Second Discourse are, as we have seen, not intended to explain the mind of the present man but rather the mind of the man who is now nothing but a long lost ideal.

However, Dent’s idea is echoed by Scott who argues likewise; Cooper’s analysis that conscience somehow underlies pity must be erroneous since it contradicts Rousseau’s statement in the Second Discourse, where he claims that there are only two principles anterior to reason: pity and amour de soi. Therefore, Scott says, conscience must be based on pity and not vice versa.253 In some sense one must agree that Scott has a point for, quite surprisingly, the term conscience is nowhere to be found in Second Discourse, which contains Rousseau’s best-known discussions of man and his natural passions and how new needs or feelings develop.254 Nonetheless, in this case, I still want to say that Dent and Scott are mistaken. The assumption that conscience and compassion are two strictly distinct sentiments is not entirely true, and both Dent and Scott seem to overlook the fact that conscience is one aspect of amour de soi, and therefore necessarily involved in the development of all further passions that, as we know, begin to develop as soon as reason sets the mind in motion. Pity, though described as the second of our two basic principles, would never be active without the former, for we feel for others only because we observe a frustration of a true need; a need

253 Scott p. 595. “Cooper argues that conscience is a principle of soul for Rousseau that somehow underlies pity, but aside from the fact that the relationship between conscience and pity is not clear from his presentation, his argument seems to contradict Rousseau's statement in the Discourse on Inequality that there are two "principles" of the human soul anterior to reason: self-love and pity. Conscience seems to me to be based on pity and not vice versa, but at any rate a more systematic account is necessary”.

254 At least it is correct to say that the term conscience is never used in the Discourse itself. Admittedly, Rousseau once appeals to conscience in his “Letter to the republic of Geneva” which he puts as a “pre-preface” to the work. Here he begs all the “magnificent, most honored, and sovereign Lords” to “look deep into your Hearts and consult the secret voice of your conscience”. (2D 7/116)
that we are motivated to satisfy in ourselves because of our proper love of self – *amour de soi*.

The fact that conscience is an aspect of *amour de soi* is clearly stated in the *Letter to Beaumont* where Rousseau says:

Man is not a simple being. He is composed of two substances. While everyone does not agree on that, you and I do, and I have tried to prove it to the others. Once that is proved, the *amour de soi* is no longer a simple passion. But it has two principles, namely the intelligent being and the sensitive being, the well-being of which is not the same. The appetite of the senses conduces to the well-being of the body, and the love of order to that of the soul. The latter love, developed and made active, bears the name of conscience. (*LB* p. 28/936, emphasis added)

Besides this, there are many occasions – especially in *Emile* and in the *Moral Letters* – where Rousseau defines conscience as a principle of the soul prior to reason. In support of Scott, one might agree that conscience must be practiced and developed by means of understanding and comparison. In the natural state, man knows only himself. Yet, even if this might indicate that a developmental process is needed for us to be conscientious, this must also be the case for pity.255 Because of this, I believe that one can say that besides *amour de soi* and pity, both conscience and understanding must have existed even before the formation of a society and thus before the point when *amour de soi* was transformed into *amour propre*. It is thus difficult to say which of them is the source, or origin, of the other. Hence I would say that they exist and develop simultaneously and dependently.

5.3. Early Sentiments of Injustice

We saw above that while conscience itself is independent of social norms, it does not explicitly talk to us unless we have reached a certain state of enlightenment. I will now return to the problem raised by Jonathan Marks that was mentioned in passing. Although it overlaps with some of the problems that we considered upon in the previous section, it goes further and offers an extended understanding of the feeling of justice or injustice which, if correct, might be problematic for the idea of conscience as an infallible guide to good. Moreover, Marks claims, this counts as another argument against the belief that the “Profession of Faith of the Savoyard Vicar” articulates Rousseau’s own idea.

While Marks’ entire project is interesting, I will here only concentrate on the part of it that presents some seemingly problematic themes for anyone who sees conscience as a secure and innate guide to moral behaviour. His

---

255 I will develop the notion of pity or compassion in chapter 7.
argument is essentially based on Rousseau’s description of the violent reaction of a very young boy who apparently had an upsetting experience caused by the strike from his nurse. The hit is not hard and the child is initially simply stunned. Soon, however, something else is rising inside him. Rousseau continues:

The unfortunate was suffocating with anger; he had lost his breath; I saw him become violent. A moment after came sharp screams; all the signs of the resentment, fury, and despair of this age were in his accents. I feared that he would expire in this agitation. If I had doubted that the sentiment of the just and unjust were innate in the heart of man, this example alone would have convinced me. (E I 66/OC 286)

Marks highlights the last line and declares that this innate sentiment of the just and unjust obviously has moral connotations since the expression bears striking similarities to Rousseau’s general definition of conscience as the “innate principle of justice and virtue”.256 He adds that although “one may be tempted to object that this is not the voice of conscience at all” he seems, nevertheless, to be utterly convinced that there is no good reason for refusing to call this child’s “sentiment of the just and unjust” conscience.257 Marks admits, however, that we must bear in mind that this sentiment is still undeveloped in children before the age of reason. What does speak in favour of this reading is that Rousseau, although denying that there can be any morality in the actions of a child, yet adding that “there sometimes is in the sentiment of other’s actions that have a relation to us”. (E I 67/288) According to Marks, this exception allows us to conclude that the fury is “precisely the kind of sentiment the child in the example experiences” and his reaction can thus be called “moral”.258

There are several arguments that can be raised against this suggestion. First, if we look at the Favre Manuscript we will see that “at the first good or bad action, at the first sentiment of good and evil is raised the cry of conscience”. (FM 143/225) Then it follows a forceful rejection of the mistaken ideas of the philosophers who believed that this inner voice was the result of the prejudices or common opinion. It is noteworthy that this discussion is placed in the third part – the so called “age of force” – and not, as with his further discussions about small children’s behavior, in the part about “the age of nature”. This third “age” is preceded by the “age of intelligence”. It is hardly a coincidence that it is in these latter parts that Rousseau begins to speak of the origin of good and bad and introduces conscience as the leading guide. The small child in our example has obviously not reached the age of

256 Interestingly, this definition is from the “profession of faith” i.e., the section that Marks does not accept as a work of Rousseau. Marks admits this but adds that he sees no reason for not using this definition in this context. Marks (2005) p. 174 n. 15.
reason or intelligence and he cannot – strictly speaking – be affected by conscience since the “first cry” of it has not yet been raised. He may surely have some kind of “sentiment of just and unjust” but only as far as he is the one affected by the unwanted effect. Yet, as Marks point out, this feeling seems indeed to be caused by a sentiment that is brought about by an action of someone else who has a relation to him, and thus there can be some kind of morality involved. Still, I would like to argue that there are reasons for doubting that the child’s anger can be said to be a manifestation of conscience although a rudimental form of a “moral sentiment” may, indeed, be present.

According to my reading, Rousseau has something else in mind and the most obvious indication of what he takes to be the cause of the child’s reaction is found at the very same page as he introduced the exception. He makes clear that conscience, although by itself independent of reason, cannot be developed without it, and before the child has reached the age of reason he does good or bad simply without knowing it. It is only in cases when he feels that the actions of others affect him in a negative way that he may react in a way that resembles moral actions. Generally, his frustration will make him violent. He will provoke everyone in order to get attention. He will smash things that come in his way simply because he can and he may kill a small animal without reflection.259 He wants to be seen; he wants to be esteemed as if he were the center of the universe. If this would be the result of conscience, there is no wonder that it must be developed. However, in the very next paragraph Rousseau explains:

Why is that? In the first place, philosophy explains it as being the result of natural vices: pride [l’orgueil], the spirit of domination, amour propre, the wickedness of man and the feeling of his weakness, philosophy could add, makes the child avid to perform acts of strength and to prove his own power to himself. (E I 67/288)

Since Rousseau literally says that these violent deeds are the result of natural vices, he is probably not referring to conscience, not even an embryonic form of it. Rather, since he immediately refers to amour propre we may conclude that some sense of humiliation has been born in the child’s mind. No one wants to be humiliated and this is why the “the spirit of domination” [l’esprit de domination] gains its force in order to make him prove his powers. Accordingly, it seems more plausible to say that the child’s feelings are not the result of conscience, but of early amour propre.

A natural reason for taking the angry, frustrated reaction as an outcome of amour propre is simply that the humiliation that was aroused in the child

---

259 E I 67/288. It might here be mentioned that Rousseau’s description of children and their behavior in book I-IV concerns boys. Small girls who are affected by amore propre rather become vain and want people to think they are pretty (E V 365/703).
was an insight of the actual *inequality* between him and the nurse who, in virtue of her age, size and experience is *superior*. As the quote shows, *amour propre* is the “feeling of his weakness” [*le sentiment de sa faiblesse*] or, more accurately, the feeling of his *relative* weakness. This awareness perfectly corresponds to Rousseau’s depiction of the first appearance of the first insights of natural inequality in the early social state that formed when people gathered around the fire, singing, dancing or showing the results from the day’s hunt. Everyone wanted to be the best, the most handsome, the strongest, the most adroit or the most eloquent since the winner became the most highly considered. (2D II 47/170) While an awareness of one’s superiority generally generated vanity or pride, the inferior, as we know, were overwhelmed with shame or envy. In this way, “the spirit of domination” arose in the human soul and this, Rousseau said, was the beginning of the fermentation of human sensibility which was fatal to happiness and innocence. In my view, the frustration of the boy who feels humiliated and weak is thus a textbook example of an effect of *amour propre* as expressed in the just mentioned passage. (E I 67/288, cf. 2D note 91/219)

In other words, the threat of our supposedly infallible guide to moral truth promoting childish rage seems to have been removed. Furthermore, while I will not discuss Marks’ proposed solution, namely the way by which conscience – through “voice training” and by means of gratitude transforms into what it ought to be – I will nevertheless add that conscience is in no way in need of voice-training, but, as we will see, it is rather *we* who need to practice our *hearing*.

5.4. The Problem of Corruption

The perhaps more difficult query with regard to Rousseau’s account of conscience was its supposed role as an everlasting infallible guide to goodness and justice. Anyone who follows its sublime voice will automatically do the right thing but, because of the “sad and constant experience” of human wickedness that we cannot neglect, we need to find an adequate explanation for the origin of all deceptive principles and evil passions that constantly influence the human soul.

We have already seen that we cannot blame intellectual inventions like arts and sciences as the *original* cause of man’s moral fall since they appeared in a condition in which the seed of vice already had begun to sprout. In other words, they merely expressed the continuance and intensification of the relapse out of natural innocence and they developed all the more destructive ideas.

The problem that we are faced with is how it is possible that mankind expresses these distorted and evil tendencies if there is (as Rousseau obviously seems to say) a divine and infallible principle in every human heart that ac-
tually makes them love the good. We have seen that this principle – the voice of conscience – is an essential part of our most basic instinct (amour de sot) and that it develops alongside with compassion. Yet, since conscience needs reason, and since reason is also the foundation and driving power of amour propre, we ought to be careful, since by making use of skills like reason, comparison and imagination we may, indeed, become gentle and humane, while we also engender our proud and competitive veins. It is, one might say, a question of balance, but why would there be a risk of corruption if conscience really was such a secure and eternal moral guide?

As we saw above, there are several critics of the idea that Rousseau actually embraced the Vicar’s view of man’s inner sentiments even though he quoted them in all their glory. A common objection is that Rousseau did not see our sentiments as eternal and stable since they tend to change as a result of our surroundings. Indeed, this is basically the core idea of the Second Discourse and it is also the reason why Rousseau in Emile insisted on the dangers of exposing the child to the world before he was mature enough to take in and evaluate his impressions in an appropriate way. However, it is obvious that even fully rational and enlightened adults are often affected by new ideas. Joseph Reisert draws attention to the passage in Emile where Rousseau explains how even a perfectly educated person can be led astray if he is exposed to various temptations. The young man who leaves his small village for Paris will probably at first find his new environment scandalous but after a while he will no longer respond in the same way. He will likely retain his sense that the forbidden manners are indecent but eventually he will also find them tempting. In this way, Rousseau says, he will achieve some new opinions, though his heart will remain the same. Eventually, however, we will find a more important transformation in the young man’s mind where also his sentiments will be spoiled by these opinions. Only then, Rousseau says, “he will be truly corrupted”. (E IV 330/658-9)

Reisert refers to this passage and says that Rousseau clearly meant that even our moral sentiments – adding in brackets “our consciences” – become perverted by these new influences. However, as was argued in the previous section, moral sentiment(s) is a broad term that besides conscience also includes pity, benevolence, gratitude, generosity and, most importantly, amour propre. Hence, by paying attention to the multitude of entirely different passions one does not need to locate the source of corruption in conscience since there are several cooperating factors that can explain the person’s way of thinking and thus also the change of his way of thinking.

Yet, one of the best arguments for rejecting the existence of corrupted conscience can be found in Rousseau’s own statements in the fifth and sixth of his Moral Letters as well as in the Dialogues. It is important to be aware that conscience, as an original principle, speaks nature’s language. We, on

---

260 Reisert p. 119.
the contrary, now speak the language of society. Our original self is timid and quiet, but we are now – surrounded by artificial passions – always ambitious and harsh. As the effect of *amour propre*, a multitude of new sentiments have arisen inside us that constantly call for our attention. The noise of these artificial passions not only drowns out the voice of soul but they will also *scare it to silence*. Rousseau continues:

As a result of being dismissed [conscience] finally gives up, it no longer speaks to us, it no longer responds to us, and after such long contempt for it, to recall it costs as much as banishing did.\(^{261}\)

We obviously became deaf to conscience and are forced to create new ways of finding moral truths. It is also important to notice that Rousseau does not, in fact, speak of human immorality in terms of erring, degenerated or corrupted conscience. Indeed, we may revolt against its directives and this inner voice can as Rousseau puts it be “degraded”, “stifled”, “defunct”, “extinct”, “ignored”, “silenced”, “deprived” or “mutinied” but it is never said to be *transformed* in its essence. Consequently, I believe that some commentators make the problem more difficult than it actually is. Timothy O’Hagan, for instance struggles, with this problem by referring to a line from the *Favre Manuscript* where Rousseau says “opinion changes everything, it depraves nature, it *corrupts* conscience”\(^{262}\) Since O’Hagan obviously wants to defend a non-corruptive reading of Rousseau’s idea of conscience, he has to treat this claim as an anomaly, or a badly formed expression. However, the original sentence reads: “*l’opinion change tout, elle deprave la nature, elle altére la conscience*”. \(^{(FM\ 143/225)}\) In fact, *altére* does not directly (and certainly does not necessarily) mean “corrupt” and in the translation included in the *Collected Writings* we find the sentence translated: “opinion changes everything, it depraves nature, it *degrades* the conscience”. To my opinion, this alternative translation better captures what Rousseau might have meant. To *degrade* is not to corrupt; to degrade is to subordinate something in favour of something else. Conscience is weakened or dismissed. This, I believe, is exactly what happens in the depraved man; he *ought* to listen to conscience but instead, he degrades it and follows the voice of his passions.

Given this understanding, it seems as if it is not our conscience that is depraved or corrupted. Seduced by inclinations, we betrayed it. We abandoned its rules for something else, but conscience itself is still inside us, although indeed more or less inactive. That it will never disappear for good is shown by the fact that Rousseau says that the voice of conscience cannot be more stifled in the human heart than reason can be stifled in the understanding and complete moral insensitivity is as rare as madness.\(^{263}\)

---

262 O’Hagan p. 126.
With respect to this, Rousseau’s theory is indeed quite similar to Kant’s. The cause of all morality in human actions (good as well as evil) lies in the use of our freedom. Accordingly, I believe that the arguments for the corruption hypothesis are based on the mistaken belief that the fact that men do evil things is the result of the imperfectness of our conscience. In claiming this one overlooks the fact that conscience itself does not decide how we in fact will act. It is not the case that knowledge of what is good or right is innate in us but only the disposition that makes us love the good. (E IV 290/600) Furthermore, Rousseau repeatedly says: “Did [God] not give me conscience for loving the good, reason for knowing it, and liberty for choosing it? If I do the bad, I have no excuse. I do it because I want to”. (E IV 294/605) We find two almost identical passages in Julie: God has given us “freedom to do good, conscience to will it, and reason to choose it” (Julie 383 III xxi) and later we learn that we have “reason to discern what is good, conscience to love it and freedom to choose it.” (Julie 683 VI vii) What is both interesting and important here is that Rousseau, to the latter claim, adds a footnote saying that “St Preux makes of moral conscience a sentiment and not a judgement which goes against the definitions of the philosophers. I think, however, that in this their putative colleague is right”.264

What these passages tell us is that we can only act in an evil manner if we, through a free choice, decide against conscience and sound reason. A natural question that arises from this is if there is any treatment for this degeneration.

This question brings to mind the idea of the general will. While a detailed examination of this complex and controversial concept falls beyond the scope of this study I will at least mention some problems that one needs to be aware of in order to understand the upcoming discussion. Furthermore even if this concept (or at least the application of it) belongs to Rousseau’s political philosophy rather than to his anthropological idea of human nature, it is based on the idea of man’s moral understanding and on his willingness to give up some claims of self-interest for the common good.

5.5. Conscience, Freedom and the General Will

In book I of the Favre Manuscript we learn that “Civil man is born, lives and dies in slavery”. (FM 9/63, cf. E I 42/253) This is essentially what we are told in the first chapter of Book I of the Social Contract, the famous saying that man is born free, but is nevertheless everywhere in chains. Rousseau frankly admitted that he was incapable of answering the question of how this

---

264 Note to Julie 683 VI vii. This is Rousseau’s way of adding philosophical themes also in a novel. The footnotes therefore ought to be regarded as an expression of his own opinions and not merely as the ideas of the fictive characters.
change had come about but he believed that he could at least explain how we could make this new state legitimate. (SC I:i 131/351)

In the following chapters of the *Social Contract*, Rousseau says that the step into a moral realm requires that man learns to constrain his impulses. As we saw already in chapter 1, it is only when the voice of duty replaces physical drives and right replaces mere appetite that we become truly human beings. (SC I:viii 141/164-5) Accordingly, if based on legitimate principles, a social pact could serve as an antidote to egoism and to the immoral use of freedom, which is the “sad and continual experience” of the first societies. In the scenario that explains the stages of the historical evolution of human perfections, Rousseau argued that the path to a moral state involved that agents, at least partly, must sacrifice their private wills and subjugate themselves to the *general will* of the society. It is also important that the essential principle of the just society is an effect of man’s awareness of freedom that, in turn, is tied to his own will. In becoming subject to the general will man, indeed, gave up the “unlimited right to anything that tempted him and that he could get” but he replaced it with the more restricted right to his own properties. In other words, he gave up *natural freedom* in favour of *civil freedom*. (SC I:viii 141/365) Moreover, as a further step in the evolution of the just moral order man also acquires *moral freedom* by which he finally becomes the master of himself since, as Rousseau famously puts it, to merely follow ones impulse is slavery but *obedience to the law one has prescribed for oneself is freedom*. (SC I:viii 142/365) Not only is this the principle of self-legislation but it is also the condition of the possibility of ascribing morality to human actions. Without freedom, we would forever remain in the state of slavery, either as slaves to passions or as slave under unlawful decrees based on the foolish idea of the right of the stronger. We will always be in chains but we may choose to be bound only with legitimate chains.\(^\text{265}\)

Now, one may wonder what conscience has to do with this. The question is not unfounded. Although Rousseau describes the body politic as a moral being with both intelligence and (general) will, and since he frequently insists on the need of principles that guarantees justice and goodness, he does not refer to conscience in the *Social Contract*, nor does he speak about it in the *Discourse of Political Economy*.\(^\text{266}\) To be sure, it seems as if conscience is here more or less substituted by the general will, which in his political works has been given the role of a celestial voice. This voice, in virtue of being the most just, is in fact the voice of God. The one who follows these “precepts of public reason” will never act inconsistently with himself.\(^\text{267}\) Yet,

\(^{265}\) For an excellent discussion of Rousseau’s allegory about chains and the possibility of “legitimate chains”, see D. L. Williams ch. 5.

\(^{266}\) Besides being absent in the *SC* and *3D*, conscience does not figure in any of Rousseau’s most significant political writings as the *Plan for Perpetual Peace*, *Plan for a Constitution for Corsica* and *Considerations on the Government of Poland*.

\(^{267}\) *3D* 144/246 and 146/248. This similarity is also identified by Melzer (1990) p.148.
it would be inaccurate to say that for the citizen, the general will plays more or less the same role that conscience plays for a properly educated natural man, like Emile. It is important to be aware that Rousseau did not regard the general will as a divine or supernatural command and certainly not as some kind of Kantian good will, which would express a universal rule for the entire mankind. Since the general will is merely the sum of individual wills of a particular society, this particular general will may be totally foreign to members of another community. (3D 143/244)

In this Rousseau’s account of the general will differs significantly from the one presented by Diderot, who described is as an “act of pure understanding, which reasons in the silence of the passions what man can require of his fellow men and what his fellow has the right to require of him”. Diderot’s almost Kantian idea of a general will of the whole species, which could never err, was presented in his article “Natural Right” published in the same volume of the Encyclopédie as Rousseau’s text on Political Economy (generally known as the Third Discourse). This fact may perhaps explain why George Gurwitsch takes also the former text to be a Rousseauian piece and claims that it clearly shows Rousseau’s trust in the general will as a “source of the rational law of nature”. Nothing could be more wrong; as Melzer points out, for Rousseau, the general will rather serves as the replacement of natural law in a society. Given Rousseau’s comments in the Geneva Manuscript that since we can no longer here it, “nature’s gentle voice is no longer an infallible guide for us” and that we need to find another form of “universal motivation” is seems clear that he no longer believed that natural law was a reliable guide for the present man. (GM I ii 77-8/283-4) Again we may recall that since there is no way back to nature, civil laws are the only possible solution for making us moral.

However, Rousseau obviously believed that ethics and politics could not be separated and claimed that “those who want to treat politics and morals separately will never understand anything of either two”. (E IV 235/524) Because of this, it is not clear how one should understand the relationship between his seemingly different agendas in the Emile and the Social Contract. While the former stresses that the good human being must know himself completely and in that way come to know nature, the latter work claims that one cannot be regarded as a moral being unless one sees oneself as part of a community of which one is also subject to its general will. The best answer to this, I believe, is that the former is the ground of the latter. This is supported by Rousseau’s famous idea from Emile, that we must become men

---

268 Quoted by Rousseau in GM (Diderot’s text is reprinted in CW III p.138).
269 “Quelle dieses Vernunftigen Naturrecht” (Gurwitsch pp. 148-9). Gurwitsch’s says that the Natural Law-text was published anonymously but claims that it is without doubt a work of Rousseau.
270 See Melzer (1990) ch. 8.
before we become *citizens*. In other words, in order to get fit for society, we must be good human beings with a soul susceptible to moral ideals.

I am aware that there are commentators who would deny this possibility. For example, Jerome Schneewind argues that since conscience is not active until we understand that we are not isolated beings, we must stimulate this understanding by looking at ourselves in relation to the *general will*. If we do not see these relations, he says, “we will be caught forever in the sad history of corruption revealed in the *Second Discourse*”. Thus, he concludes, conscience, or love of order, is “activated in the individual by the social contract.”²⁷¹ If this reading is correct, only citizens and *not men* can be considered moral beings. However, the belief that our entire development as moral beings corresponds to the *moral use of our freedom* (and hence to the active use of conscience) is, I will claim, based on a common conflation of two distinct steps in Rousseau’s theory of the origin of society. To this one may reply that (as we saw in the previous sections) conscience is made active as soon as we interact with others and when we see our relation to them. This, as we have seen, occurs through compassion and *amour propre*. The knowledge that we exist among others arose at a very early stage in human history – far before all social contracts. This is obvious if we consider the famous passage from the *Second Discourse* that civil society was founded simply by the first man who fenced off a plot of ground and decided to call it “his own”. (2D II 43/164)

Rousseau goes on by saying that even if mankind presumably could have been spared from many disasters if someone had simply hindered this “impostor” it would probably have been too late. The very idea of *property* was the consequence of a long deviation from the path of nature. We had already established social relations and even developed a hierarchic order in which we all wanted to be among the best. It is thus important to be aware that Rousseau’s “state of society” is not identified with “civil society”. What Schneewind calls “the sad history of corruption” was the result of our insight that we were no longer alone, which meant that we saw our fellows as rivals with whom we had to compete. It was from this basic insight that reason, as well as *amour propre* and other social passions developed. This was the cause of corruption as well as the birth of reason and sociability – including the sense of justice. What one can say, however, is that conscience – the sense of justice and the love of order – *reveals the need for such a contract* and thus exposes the need of establishing a *general will*. Furthermore, the belief that Rousseauian moral freedom (i.e. the power of being the master of one’s passions) is acquired in the civic state cannot, after all, be true. It seems quite clear that a Rousseauian social contract can be legitimate only if the citizens have agreed freely to it and this agreement means that they give up natural freedom and the right to everything that tempts them. This choice

requires, of course, reason and moral ideas as well as the power of preferring this state to the former. As in the case of conscience, we see that the social contract does not create any new sentiments or powers but rather provide us with the rules for their applications.

5.6. Summary

There is no doubt that sensibility is a highly complex and central aspect of the theory of the human being and Rousseau is often seen as a vigorous defender of moral sentimentalism. While feelings are sometimes dismissed as “irrational” and merely “fleeting and uncontrollable forces” that need to be restricted, the discussion shows that the idea of sensibility needs to be developed. Clearly, while Rousseau sometimes talks about conscience in terms of a sentiment, conscience is not “merely sensibility”.272

In order to stress the importance of conscience, Rousseau declared that “to know the good is not to love it” and more than plain awareness of what is good is needed to motivate a human will. (E IV 290/600) In a similar manner Kant admits that all actions need both an objective and subjective ground. The first of these is the objective motive [Bewegungsgrund]; the second is the subjective incentive [Triebfeder]. While the former can be purely rational, the latter must, in some sense, affect our faculty of desire. In other words, to know one’s duty may indeed be a motive but without a proper incentive that moves us to bring about the end, no action would be performed. In moral acting the incentive is respect, that is, a feeling self-wrought by means of a rational concept. (GMS 4:401n.) Accordingly, for both Kant and Rousseau, moral motivation involves that we are faced with an involuntary rule – an immediate moral awareness whose impact on our minds we cannot escape. It is a voice of a higher order that fills us with the feeling of respect and wonder which makes us regret that we are the imperfect beings that we are.

This chapter aimed to show that Rousseau, in spite of his outspoken trust in natural goodness and original inclinations still acknowledged that without reason and a social context, our inner judge would never extend its guidance farther than what was required in order to live a solitary life. It was conscience, along with our awareness of our freedom that raised us above the beasts. As the discussion has shown, I take the idea that artificial passions

---

272 It is thus not, as Irving Babbit once claimed, true that Rousseau’s theory of conscience is an expression of sheer romanticism and that it involves an illegitimate transformation of an “inner check to an expansive emotion”. In this, Babbitt says, Rousseau “corrupts conscience in its very meaning”. (See Babbitt (1919) p. 179.) Once one has read Rousseau’s own outline of this sentiment (or principle) one will likely agree with Peter Gay who says that Babbitt “succeeds in misunderstanding Rousseau’s doctrine with astonishing consistency” and that his book is “filled with ill-tempered denunciations of Rousseau, displays to the point of caricature.” (See Gay p. 16.)
corrupted our conscience as a misinterpretation of the theodicy of Rousseau’s theory of human nature. Instead, I have argued that all evil is, as Rousseau himself puts it, the result of *amour propre*. While this passion, or principle, develops simultaneously, and by means of the same social and rational process as conscience does, we must not blame the latter for the evil that is rather the result of *our neglect* of this moral voice. For Rousseau, this internal guide was far more infallible than all books and it never abandons us in need. It would, he said, suffice to lead us innocently, *if we would listen to it*.273

This last condition captures much of what is essential in Rousseau’s (as well as Kant’s) practical or moral psychology. In virtue of being an eternal principle in all men’s souls, conscience becomes almost as incomprehensible as Kant’s transcendental character. Yet it *is* there; we know it is there and we also know that it is giving us the truth. As Rousseau put it in the Fragment about the *State of War*, “The state of society which constrains all our natural inclinations could not, nevertheless, annihilate them; in spite of our prejudices and in spite of ourselves, they still speak at the bottom of our heart and often lead us back to the true which we leave for illusions.” (*CW* XI 63/*OC* III 611)

As we have seen, and which will become even more apparent in the forthcoming discussion, the path to morality is, to a great extent, a matter of *attentiveness* and *receptivity*. Also, Kant declared that our duty with respect to conscience is not to *acquire* it simply since there is no one without it, but rather to sharpen one’s attentiveness to the inner judge and to use every means to “obtain a hearing for it”. Moreover, Kant agreed with Rousseau that an “erring conscience” would be an absurdity.274

Accordingly, conscience does not err but so may our judgement. Again we may recall the maxim that what we must do is to “in the silence of the passions” contemplate the gentle voice of morality.

273 Reply to reactions to the *First Discourse* (in *CW* II 42/*OC* III 42).
274 *MS* 6:401, *Theod.*, 8:268. While Kant’s idea of conscience is not the topic here one may at least mention that for him, conscience is primarily an inner judge or court and not, as for Rousseau, a sentiment of justice and virtue that lead us right and make us love the good.
One of the most urgent tasks for moral anthropology is to explain how and why good and evil actions arise in the human mind. As we have seen, both Kant and Rousseau believed that human beings are by means of reason, respect and conscience, receptive to the good principle. At the same time they are also, because of their imperfect and essentially pleasure-seeking nature, susceptible to evil. Fortunately they both conclude that our rational capacity also makes it possible to choose the good principle, regardless of these constant threats to morality. While the aim of the two previous chapters was to find the locus of good and evil in the human soul, and how our various passions, principles and faculties developed and interacted in the respective systems of these philosophers, I will now try to assess some claims that have been made concerning the relationship between Kant’s description of human nature and the one presented by Rousseau.

As was mentioned in the introduction and that was taken up in chapter 4, one of the leading Kant scholars of our time, Allen Wood argued that although Kant’s doctrine of human nature may seem to be diametrically opposed to the one proposed by Rousseau, they actually express one and the same doctrine. It is no wonder that Wood so strongly holds on to the “one and the same doctrine hypothesis” since it, if true, would show that Kant’s supposed distinction between anthropology and morality is in fact merely a matter of terminology. In chapter 4 we touched upon his identification of the propensity to evil with unsociable sociability. The relevant features are thus the external factors that forced the human being into a new way of living where we had to use his higher skills, like reason and freedom. The short explanation of this process is, as we know, the sudden connection to other human beings and the restless feelings that arose from this intimacy. As Wood says, these factors develop “along with our reason, hence only in the social condition”. So far I have no direct objections but Wood’s further comments are more far-reaching and they deserve a closer examination.

As I said in chapter 4, my reservations do not concern the idea that Kant’s theory of human evil is better understood in the light of human nature and history. Moreover, I will once again stress that it is only Wood’s more far-reaching claims about these factors as being fundamentally identical about which I have reservations. According to Wood, Kant adopted the Rousseaui-
an idea of a perverted form of self-love and gave several names to it, among
them self-conceit, unsociable sociability and the “radical propensity to evil”
which, as we have seen, are simply said to be “three names for the same
reality”. To this list, Wood later also adds “ambition” and (at least implicit-
ly) the rational and foremost relative form of self-love “which is physical
and yet involves comparison”. As we have seen, this self-love, in the Reli-
gion, constitutes our predisposition to humanity. Herein, for simplicity, this
form of self-love will be called “relative self-love”. All these are also explic-
itly connected to Rousseau’s idea of the social self-love that, as we know, is
called *amour propre*. Consequently, according to this analysis, no less than
six notions are said to express the same *decisive cause of man’s evil actions:*
unsociable sociability, ambition, self-conceit, our radical propensity to evil,
relative self-love and *amour propre*.277

This chapter will continue the investigation of our natural passions and
now focus on Kant’s idea of self-regard and ambition and see how these
pertain to evil. The discussion here may exhibit “terminological excess” but
I believe that it will be useful for the understanding of Kant’s very nature of
the human being. The first section will identify the various forms of self-
regard that operate in Kant as well as in Rousseau. As we have seen, it is
common to see a parallel between man’s relative self-love and *amour pro-
pre*. Others have connected Kant’s mechanical self-love to *amour de soi* and
thus claimed that Kant’s division in the *Religion* perfectly maps Rousseau’s
own distinction.278 Comments like these appear repeatedly in discussions of
Kant’s idea of self-love or “the problematic history of man” and, while they
are sometimes made merely in passing, they are, at other times stated as
important arguments for the obvious influence of Rousseau on Kant. How-
ever, it is important to be aware that besides this fundamentally *anthropolog-
ical* dichotomy, Kant also provides a *moral* one. In the *Critique of Practical
Reason* Kant says:

All the inclinations together (even those which can be applied in a tolerable
system and whose fulfilment is called one’s own happiness) constitute self-
regard (*solipsismus*). This is either the love of self [Selbstliebe], a predomi-
ant well-wishing towards oneself (*philautia*) or the satisfaction with oneself
(*arrogantia*). The former is specifically called self-love [*Eigenliebe*], the lat-
ter self-conceit [*Eigendünkel*] (*KpV* 5:73)

Directly after having quoted this passage, Andrews Reath mentions in a
footnote that this division between self-love and self-conceit seems to be

Besides this, translators of the *Religion* generally draw attention to this similarity and add a
reference to Rousseau’s distinction at the definition of comparative self-love at 6:27. (See the
translations of DiGiovanni p. 458 n. 18 and Pluhar p. 39 n. 72.)
derived from Rousseau’s distinction between *amour de soi* and *amour propre*, but that space, unfortunately, does not allow for a detailed investigation of this hypothesis.\(^\text{279}\) Similar claims have been made by others, among them Wood.\(^\text{280}\) Whether Kant had Rousseau in mind when he established these definitions is impossible to know, but in this case one must be cautious especially since the “analogy-claim” apparently has been made with respect to both of Kant’s distinctions. This would imply that the anthropological and the moral distinction are held to be identical. In other words, mechanical self-love (animality) can thus be said to be the same as self-love and our rational and relative self-love (humanity) is essentially self-conceit. At least the last of these analogies is, in my view, obviously problematic.

Admittedly, far from all commentators make bold claims about this supposed identity and in texts on Kant’s *Religion* or the anthropological and historical works, brief notes on their Rousseauian features are the rule rather than exception. Still, I do not think that even the remarks that are made in passing should remain without comments, simply because I believe that far from everyone recognizes the full implications of these seemingly innocent notes. I will here argue that it is of greatest importance to be aware that Kant’s moral distinction between these forms of self-regard (self-love and self-conceit) differs in some significant respects from the anthropological distinction depicted in the *Religion*. The main reason for stressing this difference is that while I accept the similarity of *amour propre* and relative self-love (*humanity*) I do not agree that *amour propre* can be equated with self-conceit. Besides that, I will also say why none of Kant’s forms of self-love (particularly animality or *philautia*) can be said to fully correspond to *amour de soi*, although the analogy in this case is fairly innocent.

The second issue of this chapter returns to the problem of sociability and the passions involved in the social process. Once again I will proceed from Wood’s analysis and this time concentrate on his treatment of ambition, self-conceit and evil. Since Wood’s deduction of their relation is detailed I will quote it in full:

> We can see that ambition is the root of all evil by considering what Kant says about radical evil and the human propensities it involves. Radical evil consists in the maxim of subordinating the laws of reason to my inclinations (*Rel*. 36). This maxim is equivalent to self-conceit, the propensity to make my inclinations legislative for all willing in general (*KpV* 74). But that would be the same as the propensity to ascribe greater value to oneself than to others, preferring one’s own interest to theirs through the delusion that one is better than they are. From this standpoint, the radical propensity to evil is a predisposition to form inclinations that involve delusions of self-conceit. The systematic satisfaction of these inclinations would amount to my “always being on the top”, that is, my being regarded by others as having greater worth...

than they do – which is simply Kant’s definition of ambition (MS 435, 465).\textsuperscript{281}

Since \textit{ambition}, that is, the “\textit{passion for honour}”\textsuperscript{282} is a passion that more or less defines the social man (and thus also the \textit{depraved} man) I believe that there are reasons for taking a closer look at it to find out how it relates to the other notions that were said to express “one and the same reality”. Ambition is one of the three main passions that Kant calls \textit{social passions}. Social passions are those that inevitably appear in the social context, namely in a context where our predisposition to humanity is expressed through relative self-love, unsociable sociability or \textit{amour propre}. Still, this fact is not enough to make them identical. While I accept a great deal of Wood’s idea of the anthropological process that lead to man’s passions – and which keep up and nourish our radical evil tendencies – I will argue that neither \textit{self-conceit} nor \textit{ambition} can live up to his “one and the same reality claim”. Moreover, I will argue that ambition is not the same as self-conceit and above all, I will show the importance of not including self-conceit on this list at all. This inclination is, after all, fundamentally different from all others. The reasons for this are obvious once we understand the full connotation of this notion.

In the third section I will return to the radical propensity to evil and I will begin by recalling what was said in chapter 4. Our radical evil is the subordination of the incentive of duty to inclinations – which is essentially to choose self-love before reason – that, in turn, results in a non-moral action determined by self-love. We have already looked at the possible link between evil and unsociable sociability. Since we have now expanded our vocabulary with some new terms, I will relate them all to this cluster and evaluate the relevance of this explanation of the origin and nature of Kantian evil. However, one cannot say that this propensity is \textit{as such} equivalent with the other notions involved in our investigation, although self-love – in different forms – without doubt figures in all of them.

As the discussion will show, this objection is not a mere priggish one based on simple semantics. Rather, the main reason why I feel that these comments require critique is that this way of putting a number of notions under “one and the same reality” will easily lead the reader to believe that these aspects of our nature are essentially \textit{bad}, especially because of the inclusion of \textit{self-conceit} in this group. Moreover, since this collection of notions – including self-conceit and ambition – are explicitly equated with

\textsuperscript{281} Wood (1999) p. 290.

\textsuperscript{282} Kant’s terms Ehrsuch and Ehrbegierde (and occasionally Ehrgeiz) have traditionally been translated \textit{ambition} and this is the term that Wood uses in this context. I follow this vocabulary here although I find that Louden’s translation “\textit{mania for honour}” better captures the essence of Kant’s passion since a “Sucht” refers to an “addiction” or an \textit{abnormal} passion for something. Still, I find Munzel’s rendering of it as \textit{vainglory} and her depicting of it as a “misanthropic vice” too negative. (Munzel p. 151)
amour propre, this reading inevitably reduces Rousseau’s term to its most negative manifestations. Some of Kant’s terms can undeniably be said to have Rousseauian influences, since, as we have seen, Kant’s description of the social conflict and the problems caused by human interaction, display several similarities to Rousseau’s theory of the degeneration of human nature. Ambition was born, and so were malevolence and all the bloody wars, and the vain and inflated “feelings of superiority” – just to mention some of all the horrible consequences for which Rousseau blames amour propre. Yet, in order to see the shortcomings of this view I will close the chapter by returning to the idea of amour propre and its possible relationship to Kant’s explanation of evil as well as to the idea of the social man in general.

6.1. The Different Forms of Self-Regard

Among the natural impulses of man, self-preservation is probably the most elementary and least controversial, and as most thinkers of that time, both Kant and Rousseau saw this instinct as an expression of self-love. For Rousseau it was evident that we have to love ourselves in order to preserve ourselves. Because of that, it was naturally for him to say that amour de soi was the first and most genuine of all human sentiments. Kant, since he had put all material principles under the general principle of self-love or private happiness, had to consider also man’s most basic drives for preservation and well-being as expressions of self-love. For him, self-preservation was also one of the three drives that made up our animality. In chapter 2 we learned that there is, strictly speaking, only two general incentives, constituting the form of the maxim. These were inclination and respect for the moral law. According to the passage from the second Critique quoted above (KpV 5:73), Kant regarded all inclinations as constituting self-regard since the fulfilment of an inclination always ends up in happiness. This is also, as we have seen, what is expressed in Theorem II: “all material practical principles as such are, without exception, of one and the same kind and come under the general principle of self-love or one’s own happiness.” (KpV 5:22)

Now, if we take a look at the definitions given in the passage from the second Critique quoted above we see that Kant saw self-love (philautia) as a feeling of well-wishing towards oneself while self-conceit involves, what Kant calls, self-satisfaction (arrogantia). In the paragraphs that follow we are told that self-love is a tendency to act from subjective incentives, (inclinations) rather than from objective ones (respect for the law). This is, of course, a sign of a not fully moralised disposition, but actions from self-love are often in accordance with the law. However, because of its unreliable nature, self-love needs to be restricted by moral awareness. As Kant says, pure practical reason infringes [tut Abbruch] upon self-love by restricting it so that it does not conflict with moral demands and this self-love can thus be
called rational. Self-conceit, on the other hand, is not merely the deviation from this law with respect to the incentives involved. In fact, to be self-conceited is to make one's subjective wants into law-giving principles. In this case it is not enough to restrict and modify these principles into conformity to the law; on the contrary, self-conceit must be entirely eliminated or struck down [niederschlagen]. 283 I will return to the idea of self-conceit as law-giving but I will first expand the notion of self-conceit itself.

This sense of self-satisfaction is clearly overestimated and hence unjustified. Because of this, self-conceit can be seen as the direct opposite to humility. If one studies Kant’s overall discussion of self-conceit it becomes clear that it also involves the opinion that others should despise themselves in comparison simply because one has a higher worth than they. In contrast to self-love, which only seeks to be content with one’s perfections, self-conceit expresses an unwarranted pretension to merit. While the former is, indeed, proud of its moral perfections, the latter sees itself as absolutely perfect and in all respects blameless. This is why self-conceit also deserves the name arrogantia. (VM [Collins] 27:357) This identification of self-conceit with the arrogant attitude also shows that it can justifiably be identified as another name for what is elsewhere called haughtiness. [Hochmut] (Anthr. 7:273 MS 6:462) I will return to the notion of haughtiness in section 6.2.

The second dual conception of self-love is expressed by what Kant, in the Religion, considered to be our two first predispositions to good: animality and humanity. The basic functions of these have been explained in previous chapters so I will here merely repeat the main features of them. While the first was called merely physical-mechanical self-love, whose needs were satisfied simply by instinct, the latter relative form of self-love used reason in order to compare and it thus evaluated one’s own position in a social context. What might connect these anthropological forms that the moral distinction from the Critique is that they both seem to propose one – what one might call – innocent form of self-love which is contrasted with a more precarious one. This, combined with Kant’s well-known admiration for Rousseau, makes the common parallel to Rousseau’s two forms of self-love fairly understandable. Muchnik sees in them a perfect mapping, Pasternack says that Kant clearly intends to mirror Rousseau and, Wood includes them within his general – essentially Rousseauian – idea of an anthropological basis for the origin of good and evil with reference to the “one and the same doctrine claim”. 284

As I argued briefly in section 4.1.3. above, I find the analogy of relative self-love and amour propre acceptable since they are both treated as general passions that activate the specific social passions and skills that unavoidably

arise in man when he finds himself within a social context. However, with regard to the parallel between *amour de soi* and mechanical self-love I must say that it is not equally close and I cannot agree with the claims of a “perfect mapping” here. The most obvious mismatch is the social drive attributed to Kant’s mechanical self-love which does not appear at all in Rousseau’s natural state where everyone was guided solely by *amour de soi*. Moreover, it is not at all clear that Kant’s term involves the natural self-respect that is a vital part of *amour de soi*. For Kant, self-respect seems to require knowledge of the moral law and since this law is a fact of human reason it cannot have a natural place in our predisposition to animality. Finally, one may also point out that while Kant’s mechanical self-love constantly operates in all men as the basic instinct that secures our elementary physical needs, Rousseau sees *amour de soi* as a more or less extinct feature of human sensibility since *amour propre* took its place. Nevertheless, I agree that Kant’s anthropological dualism is far closer to the Rousseauian one than it is to the dualism from the *Critique*. Still, as we have seen, many commentators seem to find even these forms to express an analogous relationship and their Rousseauian origin has been said to be everything from a “plausible derivation” to a “direct adoption”.

A key argument here seems to be that self-love, in the *Critique*, is identified as *philautia* which, admittedly, easily leads one to think about something like *amour de soi*, as the wholly justified claim for well-being. Now, according to Wood, it is important to be aware that the self-love that is expressed among human beings is never merely a modest desire for such natural satisfaction. For us, he says, well-being always contains the greater claims that must be expressed by the innate tendency to compare and thus judge one’s state in relation to others. In other words, human beings search for happiness involves what Kant called self-conceit. 285 Although it is entirely – and even trivially – true that Kant held that human beings cannot avoid the comparative-competitive tendency (it is, after all, our predisposition to humanity) I cannot agree with the claim that this corresponds to self-conceit. Our relative self-love is, after all, defined as “the inclination to gain worth in the opinion of others” and this does not necessarily imply an already existing idea of one’s superiority. The very fact that this passion originally was a wish for receiving an equal position seems to disprove the claim that self-love originally or essentially involves such vain ideas. Moreover, the fact that this inclination gradually transformed into the wish to become superior rather indicates the opposite of self-conceit, namely envy and a feeling of being of lesser worth than others. Thus, one may say, rather than involving a feeling that one is superior, this self-love makes us aware one has to strive in order to become this better person. I believe that this is a significant difference which I will use in order to distinguish both ambition and self-conceit

---

from the “one-and-the-same-reality-claim”. As will become clear, it is im-
portant to be aware that self-love is not necessarily a bad thing. It is the prin-
ciple whose purpose is to preserve our life. Applied in the right way, self-
love may be used in order to fulfil the indirect duty to care for our own hap-
piness since unhappy persons are more likely to transgress morality than contented ones do. (GMS 4:399) The most central task here is to find out what distinguishes self-love from self-conceit.

6.1.1. Self-Love and Self-Conceit

As was hinted above, the most distinct difference between self-love and self-
conceit from the moral point of view consists in their respective relation to
the moral law. It is clear that in neither case can the agent make claims of
moral worth since the principle is merely subjective and accordingly, we
have chosen to give in to the inclinations that have forced themselves upon
“our pathologically determinable selves”. (KpV 5:74) Such principles are, of
course, entirely unfit for making universal laws but then we learn:

This propensity [Hang] to make oneself as having subjective determining
grounds of choice [Willkür] into the objective determining ground of the will
[Wille] in general can be called self-love; and if self-love makes itself law-
giving and the unconditional practical principle, it can be called self-conceit
(KpV 5:74)

The difference between these two forms of self-regard thus seems to be that
while self-love merely makes a maxim whose determining ground is based
upon an inclination, self-conceit also makes this maxim lawgiving and un-
conditional. Kant continues by saying that since there is only one objective
principle, the law, this must exclude completely the influence of self-love on
the supreme practical principle and “infringes without end upon self-conceit
which prescribes as laws the subjective conditions of self-love.” (KpV 5:74)
While it is not at all clear what it means to be “lawgiving” or “unconditional
practical principle” here Kant nevertheless seems to make a distinction be-
tween the mere making of the inclination one’s principle (and thus the max-
im one aims to act from) and the making of the inclination a legislative prin-
ciple. Because of this distinction, there seems to be actions made from self-
love which are not self-conceited. If this is the case, Wood and others who
have argued that human acting always involves the non-innocent form of
self-regard must be mistaken. Yet, they could argue that the incorporation
thesis, as well as Kant’s ideas of human agency in general, could be read in a
way that requires that all actions must be the result of one’s legislation of the
maxim. Still, one may reply that the incorporation thesis merely says that the
agent must have made the maxim a universal rule for himself and hence, not
that the maxim is also considered a “lawgiving and unconditional practical
principle”. The former clearly seems to be compatible with the description of self-love above. If all non-moral maxims actually fulfilled the criteria of being self-conceited, then the distinction between self-love and self-conceit would be superfluous. The exact difference is, however, still debated and I will briefly mention some suggestions that have been made with respect to it.

In the original version of his paper “Kant’s Theory of Sensibility” (1989) Reath says that it seems as if self-love involves what one may call “general egoism” which means that the agent finds that his own desires provide sufficient reasons to act upon them. Self-conceit on the other hand, tends toward “first person egoism”, which means that the self-conceited person finds his own desires to be adequate reasons, not only for acting on them, but also to for requiring that other’s should help us to satisfy them.286 According to this reading, while the self-lover does, indeed, give priority to subjective wants at the cost of the moral law, he does not – as the self-conceited seems to do – deny that other people have the same right to do so as well. To this Allison replies that his understanding of self-conceit differs in some respects. He admits that while the definitions mentioned by Reath (i.e. that the self-conceited person finds that his private wants can prescribe laws for others plus that he thinks that he should be regarded as more valuable than they are) are, indeed, important manifestations of self-conceit, they seem not to be the essence of it.287

According to Allison, to be self-conceited is rather to consider one’s acting from subjective concerns as a matter of right, regardless of the effects that these actions may have on others. This implies that one sees other persons as means and this way of thinking apparently violates the categorical imperative. Consequently, Allison says, an agent operating with self-conceited principles denies the legitimacy of the perspectives of others, and this is clearly also incompatible with the categorical imperative in its first form.288 In the appendix to a reprinted version of his paper, Reath now says that he may have overstated the effects of self-conceit and concludes that it is merely a feeling of actually being better than others in virtue of one’s higher personal worth. From this follows that self-conceit is a disposition that ascribes a status to oneself that is necessarily denied to others.289

For my own part, I must admit that the exact difference between these interpretations is not entirely clear and given the present explanations, I do not believe that they are necessarily incompatible. Allison has a point when he says that the self-conceited person is a being who finds his own desires more valuable than those of others. Yet, if these claims, as Allison seemed to say, are to be considered as matters of rights, this would mean that the conceited

---

286 Reath (2006a) p. 15.
288 Allison ibid.
person thinks that he has a right to have his desires satisfied, even if the fulfillment of these desires requires that others should sacrifice their own wants if they conflict with what the self-conceited person wants. This is essentially what is entailed in the “first-person egoism” as depicted by Reath. I will return to this comment in the concluding remarks.

There is much that can be said of the effects of self-regard and its effects within the Kantian framework. In any case, I believe that the important feature of self-conceit is that it involves a genuinely unjustified idea of one’s position in the system of persons, as well as with regard to the moral law. Moreover, an important difference is that, in contrast to self-love, actions from self-conceit are always contrary to the law. I am aware that this may not be entirely obvious and that it may seem to conflict with Kant’s idea that self-conceit is merely the unjustified use of the principle of self-love that, as we know, is the source of all non-moral actions, including the legal ones. We are told about people who are proud and vain since they need no moral commands in order to do what they ought to do. (Cf. e.g. KpV 5:82.) Yet, I believe that the fact that self-conceit need not only to be restricted or moderated, but always struck down, by the law shows that actions that are morally legal are not done from self-conceit. Accordingly, I believe that the law’s moderation of self-love into conformity with moral principles (and thus making them legale) means that in these cases, the self-conceit that indeed may have existed in the agent’s mind is at least not expressed in any action. In other words, actions conforming to law cannot be done from self-conceit. Because of this, the suggestion that basically everything in human life would be an effect of self-conceit must thus be regarded as a misreading of Kant’s intentions. Moreover, not even every action contrary to law seems to fulfil the criteria for being done from self-conceit. For, as Kant says, no human being who is not indifferent to morality can take pleasure in himself if he is aware that his principle does not conform to the law. Indeed, this person would, as Kant puts it, also experience a sense of bitter dislike about himself. (Rel. 6:45-6 n.) Accordingly, most human beings do sometimes act against the moral law but, in contrast to the truly self-conceited person, they will feel guilt and remorse about doing so. Accordingly, I guess that one can say that the self-conceited person is a person who is indifferent to morality since his feeling of self-satisfaction seems not to be affected by the fact that his principle is not lawful. In other words, it is an irrational form of complacency.

---

290 One might say that also a person that fulfils the criteria for being conceited in the sense of being haughty (and thus displays a sense of superiority towards others in general) may not always act from conceit in the sense given in the second Critique, by, in each particular case making the maxim a “law-giving and absolute practical principle”.

194
6.1.2. Self-Love and Egoism

In order to further strengthen the difference between self-love and self-conceit I would like to develop the discussion of egoism. As we saw above, self-love was interpreted in terms of a “general egoism” in the sense that we, when we act from self-love, accept maxims that cannot qualify as universal principles. This understanding, which is fully consistent with Kant’s own definition of self-love, does resemble his description of a moral egoist in the *Anthropology*. A moral egoist puts the determining ground of choice in his own happiness and not in the thought of duty.\(^{291}\) The acceptance of such principles is thus grounded merely in private wants and, as we have seen, for Kant, all inclination can be summed up under the idea of self-regard. ([Selbstsucht].)  

\(KpV\ 5:73\)

The supposed connection between self-love and (at least some kind of) egoism is thus not entirely unfounded. Rather, it may seem a reasonable and logical consequence of much of what Kant says about self-love and actions from inclinations. Again we may recall Theorem II where we were told that the principle of self-love is equivalent to the principle of one’s own happiness. (\(KpV\ 5:22\)) Whereas some of the implications of this Theorem were hinted at in section 2.2.2, we have not yet discussed the possibility of truly unselfish actions within the framework of Kant’s psychological hedonism; a position requiring that the action was somehow necessarily motivated by a feeling of pleasure. The classical, though incomplete definition of psychological hedonism has often led to the belief that it is merely a specific form of psychological egoism, a form that focuses on attaining private pleasure. Accordingly, if one accepts that Kant was a psychological hedonist, then he seems to be committed to psychological egoism as well. Moreover, reading Kant’s often rigid arguments about the sharp distinction between moral and non-moral principles, it is easy to believe that the fact that the moral principle is said to be unselfish (and thus withdraws all egoistic illusions) implies that motives that are derived from the principle of self-love must be selfish.\(^{292}\)

However, while it is without doubt true that any action that we perform is a result of our choices and thus expresses what we, all things considered, want most, this fact seems not to warrant the conclusion that the action must be genuinely selfish with respect to its object or aim. In order to avoid a pointless discussion, we need to make a distinction between purely selfish-egoistic actions and mere acting from private desire. As was shown in chapter 2, while non-moral motivation requires that a feeling of pleasure precede

---

291 *Anthr.* p. 130 (here, Kant calls the moral egoist *eudaemonist*). In *VM* [Collins] we learn that a moral egoist believes himself to be better than others. However, this does not follow from Kant’s definitions of self-love in *VM*, *KpV* or *Rel*.

292 Cf. e.g. *TP* 8:279-80n. and *KpV* 5:82. For a defence of the egoistic reading of Kant’s theory of non-moral motivation, see Johnston. I criticise his view in Wennersten (2013).
and directs one’s choice, this does not imply that this pleasure is also our ultimate aim. The fact that the first and most fundamental stage in the motivational chain is pleasure in the mere representation of the end does not exclude the possibility that we can represent other-regarding ends with pleasure. There are philanthropists who, “without any other motives of vanity of self-interest” find an inner satisfaction in spreading joy around them. The sympathetic instinct is said to be kind. (Rel. 6:30) Furthermore, in the second Critique, just pages from his notorious remarks about expected pleasure as the ground of all non-moral choices Kant says that:

the happiness of other beings can be the object of the will. But if it were the determining ground of the maxim, one would have to presuppose that we find not only a natural satisfaction in the well-being of others but also a need, such as sympathetic sensibility [sympathetische Sinnesart] brings with it in human beings. But I cannot presuppose this need in every rational being (not at all in God). Thus the matter of the maxim can remain but it must not be the condition of the maxim since the maxim would then not be fit for a law. (KpV 5:34)

From this it seems clear that Kant acknowledges that someone else’s happiness can be, not only the object of the action, but also the determining ground. In other words, the incentive that ultimately motivates us has the other person’s happiness as its object. The fact that such a natural instinct is rare does not remove the force of the argument; rather, it strengthens it by explicitly stressing that it is simply because of this rareness that one cannot use this “need” [Bedürfnis] in order to reach a maxim that is based on a categorical imperative. Accordingly, from these lines it follows that actions based on a wish – or a need – for the happiness of others are material and this means that they are, in line with Theorem I and II, done out of self-love. Nothing at all is said about selfishness with regard to this example and the fact that Kant continues by saying that we should rather put the general form of the action in the determining ground, only serves to articulate his wish that it would be better if we relied on an absolute safe ground since inclinations can be capricious.

Accordingly, in line with the categorical imperative, Kant says, we ought not to promote the happiness of others because this happiness is of importance to us, but simply because a maxim that excludes this cannot be included as a universal law in one and the same volition. Using the formula

293 GMS 4:398. I am aware that it is at least possible to take the word “other” [anderen] to mean that the benevolent person does not have any further “egoistic motives” than his present sympathetic inclination that, because of its non-moral status, is necessarily self-regarding. (See e.g. Johnston p. 61 and Kerstein pp. 26-7). Yet, Kant’s characterization of such persons as sympathetic and his claims that their actions are both kind and amiable must mean that the inclination of the sympathetic man is of a specific kind, and, as we have seen, there are reasons for not treating every inclination as fundamentally egoistic since they, as in cases like this, can have someone else’s well-being as their ultimate goal.
of humanity we will reach the same conclusion for this rule says that we ought to make the ends of others our own. (GMS 4: 430, 441) In such cases we would have been acting from duty and, thereby, necessarily unselfishly. Still, this passage shows that the fundamental distinction Kant sets up between material and formal principle does not imply a division between selfish and unselfish deeds. We are only told about two distinct methods of choosing our incentives, which expose two different outlooks with respect to the moral law.

6.2. Ambition and Self-Conceit

When discussing the moral anthropology of Kant and Rousseau, there is one passion whose impact cannot be ignored since it is the passion that, so to speak, defines the social man, namely ambition or the desire for honor. For Kant, ambition is one of man’s three main social passions and it is also the one that is said to be the foundation of the other two. Because of its energetic nature, it bears within it a forceful power and, as a passion [Leidenschaft], it may also make the human being lose control of his behavior. This is why passions, in contrast to ordinary inclinations, ought to be prevented. It is reasonable to believe that it is because of the forceful and potentially overpowering force of ambition that Wood regards it as the definitive root of all evil and thus he places it among unsociable sociability, *amour propre* and self-conceit and also equates it with our propensity to evil. In his deduction of these connections, mentioned at the beginning of this chapter, Wood said that the fact that ambition is the root of all evil actually follows from what Kant says about the propensity to evil. Radical evil, we are told, “consists in the maxim of subordinating the laws of reason to my inclinations.” (Rel. 6:36) Then, Wood continues: “This maxim is equivalent to self-conceit, the propensity to make my inclinations legislative for all willing in general. (KpV 5:74).” And then again:

that would be the same as the propensity to ascribe greater value to oneself than to others, preferring one’s own interest to theirs through the delusion that one is better than they are. From this standpoint, the radical propensity to evil is a predisposition to form inclinations that involve delusions of self-conceit. The systematic satisfaction of these inclinations would amount to my “always being on the top”, that is, my being regarded by others as having greater worth than they do – which is simply Kant’s definition of ambition (MS 435, 465).

---

294 According to Kant, ambition or more precise, desire for honor [Ehırsucht] is the passion that is later followed by desire for power [Herrschsucht] and greed which is the desire for wealth or simply for owning. [Habsucht].

Now, this argumentation needs closer inspection. When Wood discusses the moral effects of ambition, he refers to the passage in the *Metaphysics of Morals*, where Kant distinguishes between the exaggerated and unjustified demand for being loved and being respected by others. (MS 6:462) Wood is here associating these desires with the effects of ambition as the wish to get other peoples’ approval and admiration. In this passage, the former demand (of other people’s love) is said to be an effect of self-love (*philautia*), and the latter (i.e., of other people’s respect) is attributed to self-conceit (*arrogantia*). Even if such a wish may be a fairly normal, and a legitimate preference to some extent, the later claim does not end up far from what Kant calls *haughtiness* [Hochmut] since he defines this passion in terms of *superbia*, *arrogantia* and *ambitio*. As such, there is nothing peculiar with this since high demands or motivations are generally associated with the desire for honor.

Now, the third Latin term that is said to denote haughtiness (i.e. *ambitio*) is also Kant’s Latin equivalent of ambition and this may lead the reader to believe that Kant regarded them as more or less identical. Admittedly, several passages in Kant’s works seem to indicate that haughtiness and ambition have much in common. In Mrongovius’ notes on anthropology we find them both described as hypocrite, stupid chimaeras; they grow because of the disgust for others and so, in contrast to what Kant calls “love of honor” [Ehrliebe], they never exist without society. Moreover, in these notes, we are also told that haughtiness lies as the basis of ambition. Wood’s continued discussion indicates that he takes haughtiness and ambition to be identical, and because of this, he discusses ambition as involving three factors:

- A tendency to think yourself better than others
- A desire that others think of you as better than they are
- A desire to be better than they are

Given this Wood seems to say that ambition consists in the violation of our duties towards others. It is said to involve an illusion of one’s own superiority as well as illegitimate demands for other people’s respect. Furthermore, a feeling of contempt for others seems to be another characteristic that defines this kind of ambitious striving. It is clear that most of these qualities are also to be found in Kant’s discussion of both haughtiness and self-conceit; both

---

296 See *VA* [Mrongovius] 25:1356-7. *VM* [Collins] 27:408-9. In fact, Ehrliebe seems more or less to express the moral aspect of *amour de soi*, namely the sense of basic self-respect and the awareness of what every human being can expect from others. In the Remarks we see that “The origin of love of honor regarding the judgement of physical qualities lies in the means to freedom, preservation of oneself and [one’s own] kind”. *BB*: 20:162. In the Anthropology we are told that it has nothing to do with reputation (see § 85 7:272). In short, love of honor is based solely of the sense of one’s inner worth.

of them are, as we have seen, named arrogantia, and said to be genuinely bad. Now, since ambition and haughtiness are claimed to be identical with respect to their qualities, and since haughtiness in turn seems to correspond to self-conceit, it appears that ambition is the same trait as self-conceit. Yet, I do not think that this is an unassailable relation, and despite all the noticeable similarities between ambition and haughtiness (or self-conceit), I will highlight an, in my view, decisive difference that can also be found in Kant’s own arguments.

It is fairly clear that ambition is a form of striving (or a desire) for honor, or at least a striving for the reputation of honor. This striving, however, may sometimes fail, and failed ambition will generally result in shame and envy. While this is indeed a dimension that makes ambition even more hazardous, it at least signifies that a sense of being superior is not always, and certainly not necessarily, present in the ambitious person’s mind. Self-conceit on the other hand, is like haughtiness rather to be considered the result of successful ambitious striving, without doubt an unwanted result, combined with a false and deceived view of one’s own merits. The arrogant man does no longer have to strive for honor but only to be flattered since he has already “control over the fool”. (Anthr. 7:272) Ambition, I say, is therefore a desire for becoming an honorable and respected person, while the self-conceited already finds himself to be this super-human. It is because if this flawed picture of himself that he is to be regarded haughty, or conceited, and this is the reason why he puts himself above the authority of the moral law. I admit that also the conceited person would like to get constant confirmation of his perfection, but even without it, he will still be supercilious. In other words, the self-conceited has, or at least he think that he has, what the ambitious person wants to get, and the ambitious person’s striving can be defeated or even totally ruined if other people simply ignore him.

Further reasons for a separation of self-conceit and ambition appear in Wood’s own arguments. If we look at the passage quoted above, where he says that Kant defines ambition in terms of the desire for “always being on the top”, we will see that Kant is, in fact, not talking about ambition but about haughtiness, which according to Kant is only said to be a kind of ambition. (MS 6:465). This if further spelled out in the Anthropology where Kant describes haughtiness as the supreme form of ambition, or an inappropriate and mistaken variant of this passion that acts against its own purposes. (Anthr. 7:272) Accordingly, haughtiness is, strictly speaking, rather to be regarded an extreme and disoriented form of ambition. This shows that ambition cannot be essentially the same as self-conceit or haughtiness since it seems possible to be ambitious without thereby necessarily also being truly conceited or haughty.

Moreover, when Wood, in trying to give further evidence for his argument, refers to the opening passage in the Groundwork, where Kant says that our desire for honor, wealth and power might result in pride or arrogance, he
actually defeats his own purpose. Not only is arrogance here explicitly said to be an effect of ambition but merely a potential effect at that. Consequently, even if one undeniably can point out some similar tendencies, this is not a strict synonymy for although it is true that both haughty and self-conceited persons are ambitious, the ambitious may be neither haughty nor conceited. On the contrary, in some early fragments, Kant assigns ambition the same basic drive that he elsewhere ascribes to comparative self-love; a drive for equality and unity. (*BB* 20:165) Moreover, Kant says that, although ambition can be a foolish delusion if it is used as a rule that subordinates other incentives, it may, if limited to being an accompanying drive, be the most excellent [vortrefflich] and the ambitious man can also be polite [höflich].

6.3. Evil, Self-Love and Self-Conceit

I will now continue to the next notion in the “one and the same reality claim” dealing with the relation between self-conceit and radical evil. This claim also makes up the first argument in Wood’s deduction in the long passage quoted above. Radical evil, he said:

> consists in the maxim of subordinating the laws of reason to my inclinations (*Rel.* 36). This maxim is equivalent to self-conceit, the propensity to make my inclinations legislative for all willing in general.

First of all, one may object to the claim that the mere “subordination of the law to inclination” would be the same as self-conceit since the former is, as such, only a description of self-love. Again, while it is not entirely clear what constitutes the exact difference between merely making one’s subjective want “the determining ground of the will in general” (self-love) and making this ground “legislative” or a “lawgiving and the unconditional practical principle” (self-conceit), I believe that one can say that the former does not involve any claims of universal or objective legitimacy of the principle in question. The self-lover will thus agree that the action is unfit for becoming an unconditional and universal law. Not even the fact that the propensity to evil causes the power of choice to create a maxim is enough to conclude that the agent acting on this principle is acting from self-conceit. As we have seen, the incorporation thesis does not say that the agent, by accepting the maxim, makes the maxim a “lawgiving and the unconditional practical

---

298 *Beob.* 2:227; *Anthr.* s. 272; *VA* [Mrokovius] s. 1357. It could perhaps be mentioned that the term self-conceit [Eigendünkel] is frequently used in Kant’s moral writings but is absent from his texts on anthropology. In these texts Kant is typically using “haughtiness” [Hochmut] for the same trait.


300 Cf. the definition from *KpV* 5:74 quoted in 6.1.1. above.
principle” but only that he makes it a rule for himself according to which he will conduct himself. I think that this is an important difference and without it, one had to regard all non-moral actions as the result of self-conceit and the difference between self-love and self-conceit would thus disappear as soon as one acts. However, Kant obviously seems to make room for both kinds of actions and, as we have seen, there seems to be several cases in which the agent indeed is acting from a non-moral maxim but without fulfilling the criteria of being truly self-conceited.

If one considers the full picture of the self-conceited person, one sees that he is, after all, not only acting from inclination, but he does also not reflect on how his action may affect others. Again, the principle of self-love covers all material principles, including every possible sensible incentive including such genuinely benevolent instinct such as compassion. Moreover, as Kant explains it in the Critique, self-love restricts and moderates itself so as to conform to the law. This seems to mean that as long as the action is not contrary to law, it is not the result of self-conceit whose principles, since they cannot be merely restricted, must be eliminated or struck down by means of the law. Accordingly, far from all actions from self-love are immoral enough to fulfill the criteria for self-conceited actions, even though they must invariably be regarded as “evil” in the Kantian sense of “radical evil”. Self-conceited actions involve, as we have seen, more. In addition to arrogance we also see a deep despising of others. We find in the conceited person a self-deception and a moral arrogance that is incompatible with respect, humility, sympathy and other-regarding feelings. This is, as we have seen, not the case for actions originating from our propensity to evil since the subjective principles of our actions may be in accordance with moral motives.

Accordingly, an action caused by our radical propensity to evil is, therefore, just the same as an action from self-love. Yet, as we know, these desires can be, and are often, compatible with duty even though not performed for the sake of duty. It is true that our will is in such cases corrupted and impure since it is determined by inclinations. Self-conceit, on the other hand, transcends all moral constraints; the actions are the result of illegitimate claims on other people’s respect. Again, as the passage from the second Critique puts it, the law must not only moderate it into conformity with the law but altogether strike it down. If such actions are performed, then they are, of course, never done from duty and, because of the explicit need of a total down-striking, they cannot even be said to be “in accordance with duty”. Because of this, self-conceit transcends all forms of evil pictured in the Religion including pravitas, which, in spite of its reversal of the order of incentives in the freely chosen maxim, does not necessarily take the form of a “law-giving and unconditional practical principle.”

It is clear that Kant believed that evil requires rationality and rationality can only be developed in a social state while self-love – at least in some of its guises – is a more basic passion that exists also in uncultivated beings.
her book *Unnecessary Evil*, Sharon Anderson-Gold says that evil can be seen as “a particular transformation of self-love in a social context” \(^{301}\). Indeed, this is similar to Wood’s idea of evil as essentially unsociable sociability, or relative self-love. This is the argument that underlies their respective hypothesis that Kant’s theory of radical evil essentially rests in anthropological premises. Whatever the critics of the anthropological reading say, it cannot be denied that self-love is intimately connected to Kant’s idea of radical evil since radical evil is the propensity to choose inclinations rather than duty as the determining ground of the maxim. Here it is reasonable to recall Kant’s definition of self-love from the *Critique* where he defined it as the “propensity to make oneself as having subjective determining grounds of choice into the objective determining ground of the will in general”. \((KpV 5:74)\) Accordingly, radical evil seems essentially to be the work of self-love. Still, it is important to be aware that self-love itself is not, despite its status as a source of pathological incentives, an anthropological but moral feature. I agree with Anderson-Gold and Wood that Kantian evil can only exist in society where human beings interact. Because of this, radical evil can, at least in some sense, be said to be an effect of our social and anthropological nature. Yet, this does not mean that this evil also consists in this social antagonism, nor does evil fundamentally depend on the incitement that this social antagonism produces. The principal basis for moral evil is the fact that we are rational beings. Hence, it is in our pure practical reason that we find the moral law that we, regretfully, too often subordinate to inclinations. Therefore, it is only because of the free subordination of the law to one’s inclination that the evil expressed in society can be considered moral evil and not merely a simple effect of our limited nature. Anthropology, as occupied merely with our “subjective conditions” cannot account for our responsibility for our use of these conditions. It is solely because of our free power of choice, for which the moral law and the awareness of freedom are always present, that we can be blamed or admired for which incentives we incorporate in our maxims. The ultimate ground of this incorporation – the propensity to evil – is, as we have seen, the result of a timeless choice, and this deed cannot be explained in anthropological terms.

Finally, we must not forget what is meant by “evil” in this context. Even if our radical propensity to evil involves the self-loving act of making one’s own subjective principles the determining ground of ones will, actions originating from this tendency are not always done from a desire for increased power or esteem; fragility and impurity may easily escape this transition to self-conceit and so can actions based on simple ignorance. Even the third form, depravity is commonly legale, which means that it has been moderated into conformity with the law. \((Rel. 6:30)\) Accordingly, we may conclude that while self-conceit may indeed sometimes be expressed through our propensi-

\(^{301}\) Anderson-Gold p. 36.
ty to radical evil, the two are not to be seen simply as “different names of the same reality”.

6.4. Summary: Amour Propre and the Rest

I would like to end this discussion by returning to the supposed link between Kant’s and Rousseau’s respective theories of human nature and the passions that are central for the development of a modern and civilized man. As to Rousseau’s idea of amour propre we have seen that it is generally regarded as the perverted, or at least denaturalized, form of self-love that, like many of the Kantian terms we have met in this chapter, often makes people ascribe greater value to themselves than others. Alternatively, these terms can be said to be the outcome of the dissatisfaction that one may feel with one’s present state as compared to others. Yet, there are strong reasons for making some clarifications concerning the nature of amour propre for there are still several misinterpretations flourishing in the literature that reinforce the negative reading of it. As we have seen, it is undoubtedly true that Rousseau sometimes talks about amour propre in terms of cruelty and malice or even as “the principle of all wickedness”. Nevertheless, he also emphasises that the passions that arise from it can be benevolent and compassionate as well as envious and covetous. (E IV 235/523) Amour propre is a force that can produce almost any kind of passion and – if extended – it can be transformed into a virtue.\(^{302}\) The most correct thing to say about amour propre is that how it will manifest itself depends entirely on the position we want to have among others and how many difficulties we have to overcome in order to achieve it. (E IV 235/524f)

Admittedly, although amour propre, if carefully moderated, may be useful and benign, it cannot be denied that Rousseau underlines that this sentiment is a risky tool that often harms the one who uses it, and that it seldom does good without also doing evil. (E IV 244f/536) This statement, however, must not be taken as saying that amour propre actually can be understood in terms of the dangers that certainly may follow from it although there are uncountable translations that still render amour propre “pride”, “vanity” or even “egoism” that inevitably links it to all evils that we find in our present state. This undoubtedly contributes to the stereotypical view and it is thus important to be aware that Rousseau habitually talks about “amour propre and haughtiness”, or he specifies it as arrogant, excessive, intolerant, inflamed or violent amour propre. His frequent use of other, specific terms,

\(^{302}\) About extended amour propre, see Emile II 92/322 and IV 252-3/547-8. Generally, one can say that the very idea of amour propre in an “extended sense” [sens étendu] is not entirely easy to grasp and, according to O’Hagan, it is merely another name for amour de soi (O’Hagan p. 173).
such as orgueil [pride or conceit] and vanité [vanity], and the fact that he explicitly distinguished between these two, indicates that he did not mean that amour propre was simply synonymous with, or simply reducible to any of them. (E IV 215/494) Moreover, amour propre is also said to transform into pride, vanity or other vices and this seems to imply that amour propre must be something else. Finally, one must not fall in the trap of believing that the fact that amour propre is said to be the source of all evil in the world also implies that everything that comes from amour propre is evil.

It is thus important not to overlook the double meaning that Rousseau ascribes to it. Amour propre is sometimes referred to as a passion or a sentiment, but it is also, like amour de soi (as well as conscience), said to be a principle that is a source of passions. As passion, amour propre is usually associated with the social, or self-regarding, passions and vices mentioned above, but given Rousseau’s overall discussion it seems more plausible to say that these passions are manifestations of amour propre. Accordingly, I will say that envy, pride, greed, vanity and so on simply indicate the presence of amour propre. In other words, such passions simply show that the person in question has reached the age of reason and the moral order, and that numerous passions, or feelings, may now be born in the person’s mind. Again, we must recall that the feelings caused by amour propre are not necessarily a sense of superiority like arrogance and vanity, but they may equally well be feelings of shame or miserableness.

This is the main reason why I believe that this inborn “social restlessness” in man that Kant explained in terms of unsociable sociability, or relative self-love, ought to be kept apart from both ambition and self-conceit. Amour propre is, like Kant’s two notions, an essentially ambitious passion, but ambition seems to fall between cause and effect in this process. The pattern seems to be applicable to Kant as well for it seems clear that, for him, ambition arises from unsociable sociability or relative self-love. That this ambitious and often envious struggle may lead to self-conceit, rivalry and vanity is undeniably true but, as we have seen, this does not have to be the case. According to Kant, these instincts may, due to the inequality they are based upon, lead to much evil but he also points out that they are also the source of everything good.

Accordingly, both Kant and Rousseau would certainly admit that these results – good as well as evil – are indeed identical to specific effects of unsociable sociability, relative self-love or amour propre but these effects cannot, because of the great differences between them, be said to be identical to the three original powers as such. Again, amour propre is, like Kant’s two no-

\[303\] If one reads the fourth proposition in which Kant presents his famous explication of unsociable sociability, one will see that “it is this resistance [i.e. unsociable sociability] that awakens the powers of the human being brings him to overcome his propensity to indolence and, driven by ambition tyranny and greed to obtain for himself a rank among his fellows…” IAG 8:21.
tions, the anxious feeling of being exposed to the opinion of others. It is reinforced by means of the gnawing feeling of dissatisfaction with one’s own self, or of one’s stat as compared to others, and this dissatisfaction can be manifested in various ways of which some are truly beneficial for the human vocation. The fact that ambition is not the same as unsociable sociability but rather follows from it is also, at least implicitly, mentioned by Wood for he opens his chapter on social passions by saying that these passions (i.e. ambition, greed and desire for power) arise from inequality.\textsuperscript{304} What he does not say, however, is that inequality is itself an effect of the social conflict that, in turn, is a pre-condition for the birth of \textit{amour propre}, unsociable sociability and the relative self-love. What one might say, though, is that ambition is the power, or motor, that constantly gives new energy to these competitive forces and that ambition thus is the cause of several of their specific results.

Regarding our propensity to evil it should be clear that it is nothing but the tendency within human beings that tempts them to act from the wrong kind of incentives. As we have seen, if there is any other notion in Kant that can be said to be close enough to this tendency, it is self-love since it is defined as the “propensity to make oneself as having subjective determining grounds of choice into the objective determining ground of the will in general”. This corresponds pretty well to Kant’s account of evil since evil consists in the propensity to give priority to inclinations at the expense of the law. Moreover, both these tendencies are \textit{innate} and the results of them are either in accordance with the law or contrary to it. As we have seen, these incentives may be everything from love and compassion to egoism and malice. This is why I have argued that Kant’s notion of “radical evil” cannot be identified with self-conceit since the latter is \textit{as such} opposed to the law and is incompatible with genuine respect for other human beings. Because of the non-moral way of deliberating about motives and ends, one can easily find links to unsociable sociability, comparative self-love and \textit{amour propre}. Still, it seems as if the propensity to evil \textit{itself} is not the same as any of them since these are merely the forces that generate the desires which affect the “first ground of our faculty of choice” – a ground located in our \textit{reason} – that, in turn, incorporates and orders these incentives in the maxim.

Ambition, as well as unsociable sociability and comparative self-love may, like \textit{amour propre}, be unsuccessful, and the striving person might, end up in a state of self-contempt and bitter enviousness instead reaching a state of satisfaction with the own self. Here, I believe, we have found our most important argument against Wood’s “one and the same reality claim”. Wood’s equating of ambition and self-conceit gains some support from the similarities of Kant’s own definitions, but self-conceit is, after all, a \textit{result}. Self-conceit is thus the \textit{effect} of successful ambitious striving, a sad result that must be considered as truly evil. While also the mere self-lover often

\textsuperscript{304} Wood (1999) p. 259.
seeks reasons for pursuing a higher status, the self-conceited person actually thinks of himself as already being superior to others. For that reason, self-conceit seems to be entirely eliminable from Wood’s list of concepts since, in contrast to all the others, it does not express the fundamental social drive but a single manifestation of it. Because of this I am also sceptic to Allison’s suggestion that the idea of superiority and vanity would be considered as effects of self-conceit rather than its essence. This interpretation would imply that self-conceit could correspond to all Kant’s concepts as well as amour propre and this would both ruin the subtlety of Kant’s system and it would also contradict several of the claims that Allison himself makes with respect to Kant’s theory of the social passions. Rather, I would say that these features are possible effects of ambition. As I see it, only a person who already has a vain sense of being superior would make his subjective wants “lawgiving and unconditional practical principles” and it is only such beings that are self-conceited according to the definition given in the second Critique. Consequently, I find it more appropriate to say that self-conceit is the result of this vain sense of being a better human being. If, however, these two features actually were the effects of self-conceit, they would be necessary and inseparable from any self-conceited person. This is, as we have seen, not the case for any of the other notions that often generate a positive result.

---

7. Cultivating the Conditions of Compassion

We have seen that the development of the human being’s truly human capacities not only involves, but requires, social antagonism. Still, while antagonism and competition might very well constitute what Kant called “the first true step from crudity toward culture”. *(IAG 8:21)* these properties hardly constitute what we generally regard as the ones that explain the moral qualities of the civilized man. On the contrary, when talking about the sophisticated and good human being we rather refer to social virtues like kindness, friendship and to a sympathetic mind without which not even reason or understanding could lead us towards true humanity. Among the sentiments that are most tightly associated with sociability one finds pity or compassion that, according to several thinkers, expresses more or less the essence of human kindness. This fellow-feeling is also the source of all benevolent passions in the human heart for, as Rousseau says:

> what are generosity, clemency, humanity, if not pity applied to the weak, to the guilty, or to the human species in general? Benevolence and even friendship are, rightly understood, the product of a constant pity fixed on a particular object: for is desiring that someone not suffer anything else than that he be happy? *(2D I 37/155)*

As we saw in chapter 5, Rousseau held pity to be “the first relative sentiment that touches the human heart according to the order of nature”. This is why we ought to return to ourselves and, for as long as possible, follow the impetus of natural inclinations. This outspoken defence of natural sentiments as the true guide to human goodness seems to express more or less the opposite of Kant’s idea of the origin of moral worth, an idea that rather highlights the significance of pure reason, and which also requires that all sensible matter must be eliminated from the determining ground of the will. This is why Kant could reach the – so often criticised – conclusions that, *from the purely moral point of view*, it is only the disinterested, or even cold-hearted, agent that can be an object of true respect and whose action can have genuine moral worth. Because of this, indeed, somewhat counter-intuitive idea of human goodness, it is no wonder that few aspects of Kant’s practical philosophy have caused as much debate as some of his statements on compassion or sympathetic beneficence. While critical readers highlight the notorious ar-

---

*306 E IV 287/595-6, LM V 193/1106.*
arguments so commonly attributed to Kant with respect to this matter, others have developed alternative readings that aim to show that the established reading especially of *Groundwork* 1, is a vulgarization that overlooks several subtleties in the text.

7.1. A Novel Kantian View of Compassion?

It is easy to find passages where Kant praises “Stoic apathy” and it is difficult to read the so oft-quoted passage about moral worth from the first part of the *Groundwork* in any other way than that it requires us to eliminate all influences from arbitrary and feeling-based principles and subject ourselves solely to the objectively grounded command of the purely rational moral law. This section also contains the (in)famous story about the cheerful philanthropist who gladly helps his fellows but whose actions are said to lack true moral worth, simply because they are not done from duty, but from inclination. Although an action of this kind may indeed be *amiable* and deserves praise and encouragement, it does not qualify for esteem. According to Kant, compassion and beneficence are, after all, on the same footing as other inclinations (e.g. for honour) and such actions will only accidentally conform to duty. They are also, as we know, like any action from inclinations, expressions of self-love. However, if the philanthropic man has sorrows of his own, which have made him unable to care about the happiness of others, but he still manages to tear himself off from this state of dejection and helps others without any pleasure leading him to this action, then his action has true moral worth. The same holds for the man who is by nature cold and indifferent to the sufferings of others but who still helps them simply because duty requires him to do so. Consequently, Kant says, this emotionless man has something that is of far higher worth than a good-hearted temperament, namely a disposition to act from duty and not from inclination. (*GMS* 4:398ff.)

Naturally, this plain reading does not sit well with those scholars who want to show that Kant was, after all, not the arch-rationalist that is so commonly claimed. Even if it is true that Kant’s statement here is far from unambiguous (and that critics of Kant often overstate his antipathy for feelings) it is, again, difficult to convincingly neglect all arguments that indicate that he meant that we ought to set aside subjective concerns in favour of objective rules of pure practical reason. The importance of moral purity, as the absence of sensible matter influencing the principle, is, after all, emphasized throughout Kant’s writings and, as we learned in the *Religion*, impurity of maxims was one of the expressions of our propensity to evil. Again, any non-moral maxim is, in spite of its kind or useful intentions, to be considered evil. (6:30-1) Moreover, also the second *Critique* teaches us that sympathetic benevolence expresses a proud sense of being independent of duty. (*KpV*
5:82) Later in the same work he adds that compassion is burdensome to right-thinking and that a rational being would like to get rid of this enslavement to one’s emotional nature. (KpV 5:118)

However, because of the difficulties involved in the simply glossing over the seemingly clear-cut arguments that evidently figure in at least some of Kant’s writings, a popular approach to the problem is to discuss an alleged change with regard to this matter. This is commonly done by comparing his earlier arguments with those expressed in the later work The Metaphysics of Morals. Since the aforesaid passage from the Groundwork is the one that is generally believed to express Kant’s main attitude to compassion in motivation, it is not easy to grasp what is going on when Kant in §35 of the “Doctrine of Virtue” says:

But while it is not in itself a duty to share the sufferings (as well as the joys) of others, it is a duty to sympathize actively in their fate; and to this end there is therefore an indirect duty to cultivate these natural (aesthetic) feelings in us, and to make use of them as so many means to sympathy [Teilnehmung] from moral principles and the feelings appropriate to them. It is therefore a duty, not to avoid the places where the poor who lacks the most basic necessities are to be found but rather to seek them out, and not to shun sick-rooms or debtor’s prisons and so forth in order to avoid sharing the painful sympathetic feelings one may not be able to resist. For this is still one of the impulses that nature has implanted in us to do what the representation of duty alone would not accomplish.307

Compared to the account expressed in the Groundwork and the Critique, the claim that we have a duty to cultivate compassionate feelings must, to say the very least, be considered unexpected. Kant’s description of compassion in terms of a natural impulse that makes us do “what the representation of duty alone would not accomplish” is even more perplexing, especially if we recall Kant’s statement from the Religion that our predisposition to personality guarantees that we can always make respect for the law our sole and sufficient incentive. Moreover, we have also seen that any admixture of non-moral incentives would turn the maxim impure and thus morally worthless.

Now, §35 is indeed an interesting passage which, because of its seemingly novel approach, has given hope to sympathetic readers of Kant since he now, in apparently sharp contrast to the Groundwork, seems to say that compassion may play at least an indirect role in moral motivation in that it can be a supporting ground of our dutiful insights. Still, since Kant’s outspoken argument for the duty of compassion-cultivation is limited to these lines, one needs to be careful in not drawing too far-reaching conclusions.

from them for the lion's share of the “Doctrine of Virtue” still argues in favour of an ethics based on duties and rational commands.

The passage from §35 has been analysed from several viewpoints and much has been suggested with respect to the nature of sympathetic feelings in general and how one is to understand Kant’s claim that we, after all, need them as supporting elements for moral acting. Although Kant’s apparent emphasis on the role of compassionate feelings is, in many respects, highly beneficial for the general understanding of Kant’s idea of motivation, I believe that one should widen the perspective and not focus exclusively on the particular feeling of compassion itself, nor should one begin by discussing the possibility, or methods, of cultivating it. As I will argue, the main query here seems not to be the moral status of sympathetic feelings. Rather, the passage calls for an anthropological analysis of the very being who is supposed to cultivate these feelings. In other words, there are several conditions, or fundamental requirements, of the human being’s natural constitution that must be met before we can even begin the cultivation that is called for in §35. As we will see, an excellent way of exposing the full complexity of this problem is to discuss it from a Rousseauian perspective.

7.2. Compassion and Its Conditions

In line with what was said above, my main concern in this chapter is not to discuss whether §35 and similar passages from the Metaphysics of Morals indicate a major change of Kant’s attitude towards compassion or sensibility in general. The reason why I will use §35 as the starting point of my analysis is that it discloses some central aspects, not of human morality, but of human nature. The method will be as follows: instead of analysing how and to what extent we are supposed to use these cultivated feelings of compassion, I would like to go “backwards” and examine the very possibility of this cultivation. Indeed, the process explained in § 35 can, as some commentators have suggested, strengthen our receptiveness to other people’s condition which, in turn, will make us more attentive with respect to our imperfect duty of beneficence by, as Allison so elegantly puts it, “lengthening our moral antennae”.

As such, however, the mere process of simply visiting the sick or the poor can – at the best – only provide us with cognition of a

---

308 In order to mention just some of the many discussions of Kantian sympathy (and especially §35) one may mention Baron (1993) and (1995) ch. 6, Allison (1996) 121-3 and (2001) ch. 9, (esp. pp. 229-35), Baxley ch. 4.4-4.5, as well as the texts of Sherman, Borges and Cagle. If one really wants to learn about this matter I highly recommend the paper by Melissa Seymour-Fahmy which provides a close reading of this passage and it is the first I have seen that really analyses what Kant may have meant by “tätiges Teilnehmung” (i.e. active sympathetic participation) and how this activity is related to the two forms of sympathy that are discussed in §§34-35 See Fahmy (2009).

factual state. Recalling the process that moves the faculty of desire, we see that motivation begins with the faculty of cognition registering a state of affairs and produces a representation. At the next stage via the faculty of pleasure and displeasure a feeling connects to the representation, and this representation may then become an incentive capable of motivating the faculty of desire to perform an action. Accordingly, if an insight of this kind is to affect my faculty of desire in the way that §35 seems to suggest, it must be accompanied with feelings of displeasure, or distress, in the sight of the person in need. From the structure of §35, it is clear that Kant believed that the sight of misery would, more or less, automatically generate compassion, and he thus seems to have neglected the fact that a colder person might rather feel contempt, or disgust, and hence, instead of strengthening his compassion, he would rather cultivate his own self-conceit. Accordingly, in order to develop the appropriate aspects of our sensibility by social communication one cannot start from compassion. A closer look will show that without some underlying feelings, or capacities of human nature, compassion would, in spite of its naturalness, never arise there at all.

In order to see the complexity of this subject we may once again recall Kant’s fundamental definition of practical, or moral, anthropology as a doctrine of the “subjective conditions in human nature which either hinder or help people in fulfilling the laws of the metaphysics of morals”. Here we also need to consider the conditions of some of the feelings Kant himself had acknowledged as “sensible preconditions [Ästhetische Vorbegriffe] for our susceptibility to duty”, namely respect, conscience, the love of human beings and the moral feeling. (MS 6:399-404) Among the qualities in question we find the power of imagination, communication, identification, fear, weakness and, most importantly, sociability and humility.

By studying these capacities, feelings and traits we will see that Kant’s idea here not only mirrors several of Rousseau’s most central anthropological principles but it will also become clear that Kant must admit that these qualities, although purely anthropological and treated as being of only minor importance for ethics, are actually conditions, not only of compassion and humaneness, but also for our awareness of duty and for the possibility of respect for the moral law. In other words, in order to be consistent, Kant needs to admit that moral awareness actually requires more than a good character. He must admit that before even beginning the education of this moral character, man needs a well-formed or good-hearted temperament. Moral formation, accordingly, requires a cultivation of a far broader range of our sensibility than the cultivation of compassion and the “Ästhetische Vorbegriffe” just mentioned. This extended formation of our sensibility that will be discussed here, requires that we begin from our most primitive nature and walk through the hazardous comparative-competitive stages that tend to bring about so many vices. Nevertheless, one must not believe that this involves a setback for Kant’s ethical project. On the contrary, it highlights the
role of respect for humanity and the importance of true humility – two concepts whose importance Kant never denied. To be sensitive is, indeed, in some sense a sign of weakness, but weakness is not only an expression of imperfectness, but also a necessary means to sociability and humanity. In order to show how these new insights can lead us to a more complete understanding of Kant’s theory of human nature and morality, I would like to evaluate the argument of § 35 in the light of Rousseau’s moral maxims.

7.2.1. The Naturalness of Compassion

Besides the vast admiration for human reason, the idea of man as a sensitive and emotional being was probably one of the more wide-spread themes of the enlightenment. Adam Smith mentioned that it is evident that there is a principle in human nature – indeed a very lively one – that makes him interested in the fortune or misfortunes of others and that not even the “greatest ruffian” or “hardened violator of the laws of society” is altogether without it. Also Bernard Mandeville, who otherwise lectured about the virtues of self-interest, held that pity was the most gentle and least mischievous of all our natural passions. All men, he said, have a considerable share of it. While it is in some sense to be considered a weakness, Mandeville claimed it to bear several resemblances to virtue and it is also a condition without which society probably could not exist.

Rousseau was thus in good company when he argued that pity was deeply rooted in the human soul and indeed in the mind of every other sensitive being, including brutes. Hobbes had thus erred completely in believing that the natural man, simply because he was void of reason and moral insights, had to be considered wicked and without feelings for his fellows. On the contrary, it was exactly this lack of reason that hindered him from misusing his powers and accordingly, it is neither understanding nor civil laws that make men good but rather the innocence of their natural passions and the ignorance of vice, glory and property. (2D I 35-6/151) Moreover, Rousseau said, if the “terrible system of Hobbes” was proven to be right, all human virtues had to be considered as the outcome of a corruption of human nature. In other words, “a humane human being would be as depraved and unnatural as a pitying wolf” or, as he put it in a fragment: “the benevolence that makes

---

310 Smith Moral Sentiment Part I sec. i ch. 1. See also Force, especially chapter 1.
311 Mandeville “Essay on Charity” in the Fable of the Bees vol. I.
312 When Rousseau first discusses the innate and natural form of fellow-feelings he refers to “the natural repugnancy to see any sensitive being perish or suffer” and term this feeling pitié (i.e. pity). However, he commonly also speaks of the same natural feeling as commisération (commiseration) and links it to benevolence and mercy. These two terms are also the ones that are used when he later refers to the developed and genuine way of sympathising with the feelings of others. Accordingly, although there is a substantial difference between the basic natural pity and social compassion, this dissimilarity is not reflected in his use of terms. The French term compassion figures sparingly (e.g. in RJJJ).
us share in the happiness of our fellows, compassion that identifies us with the one who is suffering and afflicts us with his pain, would be feelings unknown and directly contrary to nature.” In other words, “a man who was sensitive and subject to pity would be a monster”.313

Interestingly, in a similar way as Rousseau regarded pity as one of man’s most basic principles, Kant assigned compassion the status of a completely natural instinct that can produce good actions even before reason has achieved the necessary strength. In that, Kant added that compassion can be considered as a “temporary substitute for reason”. (Anthr. 7:253) This also reflects his early claim in the Observations that the universal feelings of compassion and complaisance [Gefälligkeit] are grounds of beautiful deeds that would otherwise perhaps be altogether suppressed by a stronger impulse of self-interest.314

These comments raise some dilemmas related to the topic of this chapter. First, because of the naturalness and innateness of compassion, Kant’s declaration that it needs to be “cultivated” and strengthened in order to become an effective incentive, may seem paradoxical; especially if we recall his general idea that feelings, because of their impulsive nature, need to be constrained (disciplined) in order not to become determining grounds of our action (i.e. allowed to be the incentive that are given priority in the internal order of the maxim). Moreover, my suggestion that the possibility of an innate feeling like compassion necessarily rests on other aspects of our mind – including reason – might sound contradictory but I believe that it can be shown that this is the case. As was mentioned above, I count among these underlying factors the principle of sociability that, I will argue, is the most central determination of the human soul. This is shown by the fact that Kant included this principle in man’s most basic natural predisposition (animality) and, as we have seen, also Rousseau held sociability to be at least latent in the human soul. Still, the diligent reader might object on the grounds that Rousseau said that pity and amour de soi alone could produce all sense of natural right without any ideas of sociability at all. (2D preface 15/126) Yet, as we will see, this is, like his early account of amour propre, a simplification which is only applicable for the natural man in the state of nature.

313 See E IV 287/595-6 and the fragment The State of War (CW XI 63/OC III 611). Cf. also 2D where Rousseau says that even a thinker like Mandeville realized that men would be mere monsters if they did not feel sympathy for a mutilated fellow-being. The use of monster as a description of a being who is insensitive, or indifferent, to the well or woe of others is also found in Hume; see Enquiries Concerning the Principles of Morals 6.1.5.
314 Beob. 2:217f. Kant here adds that although these feelings cannot be properly called virtues they may nevertheless deserve the name adopted virtues. It might be mentioned that already 35 years before the Metaphysics of Morals and the Anthropology, Kant had claimed that providence has given us several supporting drives that are analogous to moral instincts and among them he counts honor, storge (i.e. familial love), pity [Mitleiden] and sympathy [Sympathie]. He admits that when these are present, morality is not pure but adds that a morality that excludes all such so called “motiva auxiliaria” is chimerical. (Refl. no. 6560, 19:77 dated 1762-3).
Accordingly, while Rousseau held that natural pity does not require sociability he admits in his later works that it is sociability that is the ground of all kindness and genuine compassion among civilized men. As we have seen, natural pity – pity in its most basic sense – is, after all, nothing but the instinct that hinders us from harming others. Only when we are able to connect to others emotionally and, in so doing, leave the savage state of isolation, does this feeling obtain its genuine force as incentive. Therefore, it is only in the enlightened and rational soul that this instinct can be used to actively help others and not only passively avoid harming them. (2D I 37-8/156) For this, as for all social affections, we also need enlightenment for pity, Rousseau says, “although natural to man’s heart, would remain forever inactive without imagination to set it in motion”. (EOL IX p. 306/395) Already in the first part of the Second Discourse Rousseau widens the idea of mere natural and passive pity, and defines genuine pity, or commiseration, as the feeling by which we “put ourselves in the position of him who suffers”, and adds that this feeling becomes more energetic the more intimately one identifies with the suffering being. (2D I 37/155) Thus it is not entirely true, as Michael Frazer claims, that Rousseau’s pitié is a matter of “sheer instinct” that, in contrast to Hume’s enlightened conception, lacked all signs of “cognitive elements”.315 In order to perform this transport of ourselves into the other, we need to be aware of what we and the other being, in spite of our separateness, share in common; namely we are all sentient beings and that the sentiments of the others are probably similar to our own. However, in order to reach these insights we need not only make use of our power of imagination but we must also compare our own state with the state of the other. While both imagination and comparison are to be seen as conditions of compassion they are, as we have seen, also the main sources of all social passions and deceptions, and thus also the cause of denaturalization and wickedness.

This is the reason why the fundamental conditions of genuine compassion, which must be understood and cultivated before we advance to other feelings, are sociability and humility. The latter is of special importance in that it can be said to be a condition for almost any virtue in the human soul. The former reveals the hazardousness of the process of becoming a true human being since this development necessarily involves comparison. As we have seen, Rousseau saw the root of human evil in the denaturalization of humanity that took place through the misuse of the powers that were acquired through the social process, especially those born out of inflamed amour propre. Similarly, Kant believed that vice was the result of competition caused by unsociable sociability and relative self-love. In both cases, the hazards are due to comparison. It is clear that we, by comparing ourselves with others, expose ourselves to the risks of being vain and self-conceived.

315 Frazer p. 42.
egoists by means of the competitive tendency that is more or less automatically set in motion through this process. Yet, at the same time, without comparison there can be no true compassion and hence no tender humanity. Again, we may also recall Kant’s claim that without antagonistic forces, all useful powers of the human mind would forever slumber undeveloped in our soul. Accordingly, comparison must not be dismissed; on the contrary, we must look around and learn how to compare in the right manner and to what extent this interpersonal comparison is to be practised. This, I will argue is what both Kant and Rousseau must hold to be an important part of moral education.

7.2.2 The Importance of Humility

In order to see the complexity in the cluster of sensitive features that operate in our mind we need to return to §35 and take a closer look at what Kant actually says. The plain reading tells us that we ought not to shun the places where the poor “who lack the most basic necessities” are to be found, but we must rather “seek them out” in order to learn that some of our less fortunate fellow-beings suffer from various miseries. If we do that, we will likely be overwhelmed with feelings of distress, and the compassion that follows will, according to Kant, make the rational decision to help them much easier. Obviously, Kant believes that the action that follows will still be “based on moral principles” so the feeling of compassion is not to be considered the main incentive here. However, whatever the status Kant grants this feeling, much more than plain awareness of inequality and the pain of others is needed in order to excite feelings of this kind and Rousseau makes this point clear. After having read the Second Discourse, Mr. Philopolis wrote to Rousseau:

I will only say one more word. It is about pity, that virtue so celebrated by our author and which was, according to him, the finest endowment of man in the infancy of the world. I beg Mr Rousseau to reflect well on the following question. – Would a man or any sensitive being, who had never known pain have pity, and would he be moved by the sight of an infant being strangled?316

Rousseau did not waste any time contemplating this question and his immediate and concise answer was simply “no”.317 He did not provide any arguments in support of this standpoint in that letter, but his comment in Emile leaves no room for doubting this idea:

316 Letter from Mr. Philopolis 1755 in CW III p. 131, OC III p. 235.
317 See Rousseau’s Letter to Philopolis 1755 in CW III 131/ OC III 236.
The man who did not know pain would know neither the tenderness of humanity nor the sweetness of commiseration. His heart would be moved by nothing. He would not be sociable; he would be a monster among his kinds. (E II p. 87/314)

I will soon develop Rousseau’s argument but I will first return to Kant for it is exactly in this argument that we find what we need in order to demonstrate the conditions of compassion and thus also the need for cultivating these conditions before we can proceed to the cultivation of compassion itself. From §35 it is clear that Kant believed that we, in order to feel pity, must recognize that there are miseries and evil in the world but, if Rousseau is right, we must also know what it is to suffer. Kant does not make the last point explicit, but it seems, as Rousseau says, to be an inescapable fact that we cannot respond emotionally to states we are not familiar with. If I am to really feel compassion for the suffering person, I must, after all, be aware that suffering is something terrible and hence something that ought to be prevented. Most importantly, however, I must also understand that I could have been the one in need. In other words, I must realize my weakness, vulnerability and limitations and thus what I, and all other human beings, in spite of different appearances, always have in common, namely that we are all human beings. It is clear that Rousseau had something like this in mind when he rhetorically asked why Kings are so short of compassion, and why the rich so often indifferent to the glooms of the tramps at the street. The answer is explicitly articulated in the “Second Maxim” namely:

- We pity only those ills from which we do not feel ourselves exempt. (E IV 224/507-8)

In relation to this moral axiom Rousseau refers to a phrase from Virgil’s Aeneid: “non ignora mali, miseris succurrere disco” and adds that he knew nothing so beautiful, so profound, so touching and so true as these words.318

The essence of Rousseau’s argument can be explained as follows: Kings and nobles do not expect to be a poor commoner and they cannot, therefore, imagine the position of their less fortunate fellow-men. This indifference, Rousseau says, is the effect of man’s alienation from nature, which began already at a very early stage. While reason could make us identify with other beings, the thinking man could likewise put himself aside of the show allowing himself to say: “perish if you will, I am safe.” (2D I 37/156) As Marshall points out, this act of being a mere external spectator reflects Lucretius’ example of the shipwreck where someone watches the disaster from a distance,

318 In Bloom’s translation the sentence is rendered: “Not ignorant of ills, I learn to assist the needy”. E IV 224/507 from Virgil, Aeneid I: 630. In J. Dryden’s translation of the Aeneid, this passage runs: “For I myself, like you, have been distress’d, Till Heav’n afforded me this place of rest; Like you, an alien in a land unknown, I learn to pity woes so like my own.” (Aeneid New York, P. F. Collier and Son 1909 p. 98).
contented about not being there himself. 319 Reason and philosophy provided us with this “admirable talent” so totally unattainable to savage man. The philosopher, the most estranged of all men, could only in his “tranquil sleep” distract himself with reflective thoughts while men could die before his eyes. (2D I 37/156)

While Kant did not share Rousseau’s pessimism with regard to science and thirst for knowledge, it nevertheless seems as if he, at least in his early days, had thought in similar terms with regard to the idea of sympathy between people from different parts of the social hierarchy. First, in the Remarks (1764) he says that we are generally more susceptible to the needs of our own kind and thus a citizen is compassionate towards others who are oppressed by the prince but the nobleman is only compassionate towards other noblemen and hard on the peasants. (BB 20:134-5) Similarly, we find in the anthropology notes from 1772 that while the inferior [geringe] person has sympathy with the higher-ranking one, this distinguished person has no sympathy with him. Kant adds that this is the unfortunate thing with kings, since they, because of their narrow minds, cannot imagine the misfortune of their subjects. He adds that they do not even want to do so. (VA [Friedländer] 25:606-7) Obviously, power corrupts and makes men insensitive. Put in Rousseauian terms, by this denaturalization they forget that they also were born as naked and weak as the tramp and that they, like all men, one day will die. (E IV 221/503) In perfect contrast to Rousseau’s heroine Julie, who knew “too well the ills of life” and how easily “even honourable men can be reduced to their lot by the misfortunes of life”, the king, who has lived a protected life, has narrowed his mind in this respect. (Julie V:2 441-2/540)

Nevertheless, since no one likes being treated as inferior, the choice to follow the lead of our ambitious vein – passions raising from our relative self-love, or amour propre – must, after all, be regarded a very natural choice. Still, one must not forget that however natural this ambition may be in our present state, it is nothing but the first step from innocence and, as we saw in Chapter 6, the foremost threats to morality consisted in the perverted form of self-regard, called self-conceit that is the ultimate effect of inflamed amour propre. Besides being a delusive feeling that ignored the moral law completely, it also involved the idea of one’s superiority to others. Although self-conceit is not the sole cause of emotional coldness, it is suitable to use it as example since it leads us to the core concept of the proper treatment of all transgressions of law: humility. In Chapter III of the Critique of Practical Reason, Kant assigns respect [Achtung] the role of this humiliation [Demütigung] or, as he puts it, down-striking of self-conceit. So, it seems as if moral awareness is needed in order to treat this exaggeration of one’s self-esteem.

319 In De Rerum Natura I part ii. Marshall draws attention to this example at p. 139.
Now, while respect is the moral answer to the problem of actions contrary to the law, it is clear that this kind of humiliation by means of respect, or, as in the case of Rousseau, via the voice of conscience, is a cure rather than prevention. In the ideal world, self-conceit or inflamed amour propre would never be born, so the real question here must be how one may preclude, as far as possible, the birth of such vain illusions. Still, as we have seen, in order to develop our social skills we must, so to speak “take the rough with the smooth” and admit, as did both Kant and Rousseau, that without refinement man would remain a half-finished creature that would be as alienated to the social state as the modern man would be in relation to the natural condition. In view of this, Rousseau’s initially puzzling idea that man needs much art in order to avoid being artificial becomes comprehensible. (E IV 317/640)

7.2.3 Comparison and Communication

In the preceding chapters we have learned that human beings can only be fully human in a social context. Even Rousseau admitted that there is a communicative vein inside us for as soon as we realized that we are surrounded by other thinking and sentient beings, we felt a need to communicate these feelings. This was the reason why also the originally timid man came to join a group of equals and in this way founded what eventually became the social state. (EOL I 289-90/375)

To communicate means to interact; for Rousseau the most important means for this was language (speech) and compassion. Performed in the right way, communicating feelings are the most beautiful expressions of kindness. It is also the way by which compassion works, for, as Rousseau says:

How do we let ourselves be moved by pity? By transporting ourselves outside of ourselves; by identifying ourselves with the suffering being. We suffer only as much as we judge he suffers; it is not in ourselves, it is in him that we suffer. (E IV 223/505-6)

No one, Rousseau adds, becomes sensitive until his imagination is animated and begins to transport himself outside of himself in order to take on the other person’s being. (EOL IX p. 306/395; E IV 223/505-6) Also Kant recognised this noble practice of our sensibility and in perfect coherence with the idea of a humble soul and a non-narrowed heart, Kant sees the ability of interpersonal identification as expression of broad-mindedness. 320

320 See e.g. KU § 40 esp. 5:293-5, IV A [Busholt] 25:1481-2. Admittedly, Kant does here not explicitly discuss compassion but he talks about the ability to put oneself in the position of everyone else by abstracting from contingent limitations, such as our own judgements. In this way one expresses broad-mindedness [erweiterter Denkungsart]
Interestingly, this method of identification with the feelings of others was one of the reasons why Kant denied that compassionate feelings could be regarded proper duties. The reason is that when we, by means of the power of imagination, put ourselves in the place of the one who suffers, simply increase the amount of misery in the world. Therefore, compassionate feelings can be likened to infectious diseases, which spread from person to person and which affect also those who were not originally affected. This is why it cannot be a duty to feel compassion. (MS 6:457) In response, one might here remark that Kant seems to overlook a central feature of compassion for even if it might be true that compassion will make us “suffer”, we will still, as Rousseau points out, “feel the pleasure of not suffering as he does”. (E IV 221/504) This is why compassion is sweet while envy is bitter for when we envy, we do not put ourselves in the place of the other but we rather regret that we are not there. This is further explained in the Dialogues where Rousseau makes clear that “the truly egoistic man does, in fact, no longer love himself but he rather hates what is not himself.” (RJJJ II 157/863) This rule of emotional response is expressed in Rousseau’s “First Maxim” saying that:

- It is not in the human heart to put ourselves in the place of people who are happier than we, but only in that of those who are more pittable. (E IV 223/506)

Rousseau said that though there might indeed be rare exceptions to this rule, these are generally more apparent than real. Since we are by nature not disposed to identify ourselves with the condition of the happier, wealthier or more successful man by transporting ourselves into him, we may here recall what we saw already in Chapter 1, namely that in order to cultivate the right kind of sentiments and thereby “turn [the child’s] character toward beneficence and goodness” one must not expose his eyes to the “deceptive image of the happiness of men”.321 Kant obviously draws on this idea when he remarks that man “is poor (or considers himself so) only to the extent that he is anxious that other human beings will consider him poor and will despise him for it.” (Rel. 6:93-4) Again, this anxiety, which arises as soon as he is among human beings, almost exactly mirrors Rousseau’s description of the reflecting person who immediately starts to ask himself: How many men have I already surpassed? How many can I still reach? Why should my equal go farther than I? (E I 62/281)

As we have seen, what determines whether the passions of a man will be fundamentally humane or competitive are his personal goals. In the modern man, these goals are always planted in him from outside. Again we may recall Rousseau’s warnings of not putting the seeds of pride, vanity and envy

321 E IV 221-2/505. Cf. section 1.3 above.
in him by providing him with the deceptive image of luxury or glory. A parent must also refrain from stoking the child’s pride by flattering his “childish prattle”, by saying that he is a “clever little boy” and thereby make him believe that he is smarter than he actually is.\textsuperscript{322} It follows, therefore, Rousseau says, that, in order to incline a young man to humanity, one must avoid “making him admire the brilliant lot of others” and instead one must “show him the sad sides of that lot, one must make him fear it”. (\textit{E IV} 223/507) By doing this, we will avoid comparisons that may end up in envy and an agitated wish to advance in the social hierarchy.

This maxim thus demonstrates the importance of \textit{true sympathy}, that is, \textit{real identification} and not only a mere “passive awareness” of the other person’s state in comparison to our own. To be moved to help my fellow-being I must, of course, know (intellectually) that he is in need but I must also recognize the actual \textit{inequality}, that I am, indeed, better off. Obviously, this reflection requires that I realize my superiority and without the act of \textit{proper sympathizing} – my transporting myself into him and imagining myself in his place – I could equally well feel contempt, pride or even malicious pleasure [Schadenfreude]. The mere observation of a state of need from outside may also result in what Kant calls an “insulting kind of beneficence” where I, in the act of helping, also display my state of being in a better position and thereby making him feel unworthy and pitiful. (\textit{MS} 6:457) Still, another and more serious problem is, as Richard Boyd points out, that human suffering may, in this way, become a kind of “spectacle” that is “showcased, and manipulated for the education of the few”. In this sense, Boyd continues, “suffering humanity exists for the sake of Emile’s education and eventual happiness, and not vice versa.”\textsuperscript{323} Using Kantian terminology one might say that these unfortunate beings are in some sense regarded as more or less necessary \textit{means} to our moral refinement. Accordingly, without being exposed to other beings’ suffering, we would not develop the compassion that, as we have been told, will make us better human beings by making us more susceptible to the moral law. Clearly, treating other human beings as means is a violation of the humanity formulation of the categorical imperative.\textsuperscript{324}

Perhaps one may escape this possible dilemma by objecting that one does, in fact, \textit{not} treat these unfortunate beings \textit{merely} as means and that the very act of feeling sorry for them implies that one also, \textit{at the same time} treats them as ends in themselves. If we, as Julie suggests, refuse to describe them by the “contemptuous name of \textit{tramp} [gueux]”, which – by depriving them of their humanity – only fits into the mouth of a man who has adopted maxims that “flatter the harshness of the wealthy”, we will still, in spite of

\textsuperscript{322} \textit{E IV} 217/497-8, \textit{E IV} 221-2/505, \textit{Julie V} letter 3.

\textsuperscript{323} Boyd p. 525.

\textsuperscript{324} This formula commands that we shall never act in a way where we treat humanity (in ourselves as well as in others) merely [bloß] as a means but always at the same time [zugleich] as an end in itself.
our higher rank, see them as human beings and fellows. And, as Julie adds, it does not really matter whether our compassion actually helps them or not, since it at least testifies to the interest that we take in their troubles. It is, she says, a sort of greeting we give them that “mitigates the harshness of a refusal”. (Julie V:2 441-2/540)

In a way, the essence of this opinion is implicit in Rousseau’s “third maxim” which tells us that:

- *The pity one has for another’s misfortune is measured not by the quantity of that misfortune but by the sentiment which one attributes to those who suffer it.* (E IV 225/508)

Here Rousseau adds that one pities the unhappy man only to the extent one believes he is pitiable. In other words, our ability to feel distress at the sight of an unhappy being is to a great extent dependent on our own thoughts about his ability to suffer. Since the effects of unhappiness and suffering are often increased by memory and imagination we only pity those that we believe to possess these faculties. This, Rousseau says, is the reason why so many men are insensitive to the ills of animals for, as he puts it, we do hardly pity the horse relaxing in his stable since we do not believe that it remembers the whip; nor do we believe that it fears the hard work that is expected of it tomorrow. Neither does the grazing lamb move us to tears even though we know that it will eventually be slaughtered. Since we judge that animals do not foresee their fate, our hearts are generally hardened in this respect. Thus the pity we feel for the poor can, as Julie just said, be seen as a greeting, or a sign, that we regard them as our fellows or even equals. Once again we see the effects that a sense of superiority has on our sensibility for in the same way as men in general close their hearts towards the feelings of animals, the king and rich – by looking down from their thrones – are hardened against the fate of ordinary men simply because they assume the latter to be stupid enough to feel nothing of it. (E IV 225/508-9)

7.2.4. Imagination and Theatric Life

It is clear that communication and imagination are essential for the development of social virtues such as sociability and pity for, as Rousseau said, he who imagines nothing, or does not participate in society, feels only himself and is, “alone in the midst of mankind”. This is why he could conclude that however natural pity is to the heart of man, it would nevertheless “remain forever inactive without imagination to set it in motion”. (EOL IX p. 306/396) In Emile we were told that in order to become sensitive and pitying, “the child must know that there are beings like him who suffer what he has suffered and who feel the pains he has felt, and that there are others whom he ought to conceive of as able to feel them to”. In order to reach this
state of identification one must nourish the child’s nascent sensibility and thereby excite in him “goodness, humanity, commiseration, beneficence, and all the attractive and sweet passions naturally pleasing to men”. (E IV 222/506)

Yet again, few of the means for the development of our benevolent passions are entirely safe, and although Rousseau in both Emile and in the Essay on the Origin of Language encourages the use of imagination and interpersonal identification for the birth of our sensibility, we have also seen the possible effects of an incorrect use of them. In Emile we are told that while the real world – just as our natural capacity – has its limits, the imaginary world is infinite. Since we cannot enlarge the former, we need to restrict the latter for, as Rousseau puts it, “it is from the difference between the two alone that are born all the pains which make us truly unhappy”. (E II 81/305)

Now, this condition of nascent sociability – the idea of identification and the awareness of the mutual relationships in the social state – lead us towards a well-known dilemma in Rousseau’s eternal balance between his criticism of artificial culture and his clear and sincere awareness of what is needed in order to become a morally good being in the social state. Rousseau’s problem here consists in the fact that when we develop sensibility by recognizing our relations, we at the same time leave our original authenticity. These thoughts also seem to expose an obvious inconsistency in his relation to art, and especially drama. As we saw in Rousseau’s answer to Philopolos, men go to theatre in order to experience the sweetness of pity and commiseration by gladly shedding their tears at the sight of the miseries presented at the stage. They do it in order to feel their innermost human feelings and in this letter, Rousseau clearly did not condemn this practice. Yet in other writings it is clear that he saw this kind of pity as nothing but a chimera that did more harm than good to the human character. His frustration with regard to this matter was articulated with force in his famous Letter to d’Alembert; this open letter was also his response to d’Alembert’s article on the subject “Geneva” in the Encyclopédie. Although this article was of great interest for Rousseau, who read it with pleasure, there was one thing that he could not leave without comment. Rousseau’s concerns were due to d’Alembert’s suggestion that the city would benefit from a theatre.325 In direct opposition to this idea Rousseau argued that a theatre or, more precisely, a “théâtre de comédie” would rather destroy the character of his hometown, a small city that was still relatively undamaged.

It is here natural to ask why Rousseau, who wrote several plays himself, felt a need to express such an antipathy towards this kind of art. The perhaps most obvious answer is that a drama is not real but a play or spectacle. Most importantly, a theatrical performance is built on the idea of pretending to be something that one is not. As such, it embodies the whole idea of the cor-

325 d’Alembert p. 244. (in CW vol. X).
ruptness of the present human condition, namely our abandoning of the genuine for the artificial.

In his classical book *Rousseau and His Era*, John Morley says that the *Letter to d’Alembert* was essentially a supplement to the *First Discourse* that aimed to apply its ideas to a factual case. At the same time, the *Letter* can equally well be seen as a deepening of the *Second Discourse* as the ultimate illustration of a state in which men became “actors” and “spectators”, a state where human beings performed their roles simply because they knew that they were observed. Again we may recall Rousseau’s insistence that in the genuine state of nature, *amour propre* did not exist since each individual regarded himself as the sole spectator to observe him and thus as “the sole being in the universe to take any interest in him and the sole judge of his own merit”. (2D II 91/219) On the contrary, the social state turned into a state where we all became enslaved to public opinion. Accordingly, the idea of theatre is the very opposite of the state of nature where there was no difference between what *was* and what merely *seemed to be*. In contrast, the state of society became, so to speak a *theatrical* state since the man of society, as Rousseau put it, is *always in his mask*. “What he is, is nothing; what he appears to be, is everything to him” (E IV 230/515) In that way, humankind lost its original authenticity and security since they could no longer see through each other. (1D I 6/8) As he put it in the Preface to Narcissus:

> Every man who occupies himself with agreeable talents wants to please, to be admired, and he wants to be admired more than others. Public applause belongs to him alone; I would say that he does everything to obtain it if he did not do still more to deprive his rivals of it. (*PN* 193/668)

In sharp contrast to the aforementioned natural state in which the only one who took notice of us and our merits and shortcomings was ourselves, we now had to handle the anxiety of being observed and judged by the mass. Accordingly, by our well-known comparative-competitive germ and the wish to be seen, applauded, and esteemed, men became not only a *spectacle* for others but also a *spectator* sine these two are more or less inseparable.

We now see why the method that we found in Rousseau’s early comment to Philopolis, where the stories performed at the stage make us feel the sweetness of pity, cannot properly be regarded as a moral practice. At this “temple of illusion and falsehood” we can never experience any authentic feelings but only “fleeting and vain emotions” which only last as long as the illusion on the stage. As soon as the curtain falls, all feelings subside and the pity that we feel at the tragedy is to be considered *sterile* since we cry at

---

326 Morley vol. I p. 320.
327 Recall again Kant’s claim that man “is poor (or considers himself so) only to the extent that he is anxious that other human beings will consider him poor and will despise him for it.” (*Rel.* 6:93-4).
imagined ills and such tears have never produced the slightest act of humanity. This is not only a waste of feelings but also to lie to ourselves since we believe that we have satisfied all rights to humanity and we thereby ignore the true needs that are present all around. Again, Rousseau lets Julie be the exception who, instead of paying the actors to make her shed some “sterile tears” prefers to help a real needy person who, in contrast to the actor, moves her heart to genuine pity. Accordingly, instead of feeling the pleasure in the love of “other’s good deeds”, which are feelings that never last, she will perform some good actions herself. Feelings resulting from such practice, she adds, are forever removed. (Julie V:2 441-2/540)

Still, it is important to see that it is only the “actor part” that is here to be condemned. In a state where communication is a necessary condition for sociability and moral feelings, we need to take on the role of spectators. It is, as Rousseau puts it in the Origin of Language, much easier to be moved by someone who tells you everything that he feels. Only in this way, Rousseau adds, “do the scenes of a tragedy have their effect”. (EOL I 291-2/378)

7.2.5. Kind or Weak – Compassionate or Hardened?

As was shown in the previous sections, the direction of the human heart towards goodness and benevolence requires that the human being opens his heart to the world and exposes his soul to external impressions of the fates of his fellows. In doing this, he will be soft and caring and the insight of his semblance to them will lessen the risk of breeding inflamed amour propre. Still, cultivation of specific feelings, however gentle, is not always unproblematic and too much emotional training may end up in less desirable results. As we saw already in chapter 1, the chief reason why men found themselves committed to enter a social order was the awareness of their own vulnerability. In Rousseau’s words, “it is man’s weakness that makes him sociable”. (E IV 221/503) All attachment, we were told, is a sign of insufficiency and without the need of others, man would hardly think of associating with them. Again, we may recall that this insight of our weakness is the criterion of sociability and humility that, in turn, are requirements for the possibility of compassion, benevolence, beneficence and kindness in general. The modesty that makes us realize our susceptibility to ills will most likely make us less hard-hearted and this is why Mandeville concluded that though we are all more, or less, affected with compassion, the weakest minds contain the most.329 Now, while weakness is the ground of sociability and thus of all social virtues, Rousseau also admits that “all wickedness comes from weak-

329 “Essay on Charity” p. 254. See also “Inquiry into the Origin of Moral Virtue” p. 56 (both included in the Fable).
ness. The child is wicked only because he is weak. Make him strong; he will be good.” (E I 67/288)

Although these two claims about weakness might, at least initially, appear to conflict, the tension dissipates in the light of the main cause of the social problem, namely the comparative-competitive tendency that arose because of our anxiety of being regarded as inferior to others. The passion that we all know too well, *amour propre* is – as was mentioned in chapter 5 – actually nothing but the feeling of our own weakness. (E I 67/288) In other words, weakness or, more accurately, *the feeling of one’s weakness* is thus the origin of both of our relative sentiments (i.e. *amour propre* and pity) and again we must recall that whether these sentiments will result in beneficence and commiseration or in covetousness and envy depends solely on which position we seek and that obstacles that we need to overcome in the striving for it. (E IV 235/523-5)

Human weakness is, in both these senses, to be understood as dependence, either in terms of a real need of other beings in general or as a felt need of being admired by others. This is why Rousseau insists that before we have become strong enough to ignore the pressure that our growing *amour propre* excites in us, we need to be masters of ourselves. Once again we may recall that one must not expose the child to the “exterior of high society” before we have put him in a condition to evaluate it in itself. As we have seen, “to show him the world before he knows men is not to form him, it is to corrupt him; it is not to instruct him, it is to deceive him”. (E IV 221-2/505) Again we see the effects of denaturalization for as we saw above, it is the gap between our capacities and our illusory needs that make us unhappy. Accordingly, Rousseau says, man is very strong when he is content to *be what he is* whereas he is very weak when he wants to elevate himself above humanity and to *become something that he is not*. As long as we believe ourselves to be sufficient unto ourselves we will never have to complain of our weakness simply because we shall never feel it. (E II 81/305)

While the process outlined in Rousseau’s *Oeuvres* may appear appealing and straightforward one must not neglect the potential pitfalls. If we return to Kant’s claim in §35, we see the idea that we should not shun sick-rooms or prisons clearly borrows much from Rousseau’s plan for Emile who is sent to meet the unfortunate. Again we may recall that in order to become a sensitive and pitying person, the child must learn that there are beings similar to him who suffer from various ills and misfortunes. Still, Rousseau warns of dangers of an excessive exposure and Emile must thus not be sent from one hospital to another, to every prison in town or to the executions at the square. Too much impression of suffering will rather harden him since habit accustoms us to everything. (E IV 231/517) This will eventually destroy a man. His soul becomes narrowed. Vile passions will stifle his otherwise delicious sentiments and he no longer transports himself into others. He feels neither
joy, nor sadness. This unfortunate man, Rousseau says, “no longer feels, no longer lives. He is already dead.” (E IV 287/596)

Kant does not directly address this possible problem, nor does he say much about the risk of becoming hardened by too much exposure to ills and evil. Rather, critics like to point out the fact that he, in a positive manner, mentioned the “sublimity” of the wise Stoic who could simply say “what is this to me?” when he realized that he was unable to help a friend in need. (MS 6:457) For Kant, to be unmoved, therefore, is not a bad thing but it rather signifies a “strength of soul” that may be highly useful since such a person is less inclined to be overwhelmed by feelings. Hardening of the heart seems, in his view, not to be the result of seeing too much, but rather of repeated cruelty. This, as we shall see, is a real violation of the duty to oneself.

7.3. The Moral Status of Humility

In order to return to one of the most central of the conditions of the compassionate soul we now need to take a closer look at Kant’s own treatment of humility and its role in his ethics. In line with what has been said above, humility can be understood as the opposite of pride or self-conceit; or, in Rousseauian terms, it is the absence of inflamed *amour propre*. As long as we understand humility as a correct (i.e. sound) estimation of one’s own worth which does not relapse into self-abasement, there is nothing original or puzzling with the idea that such an attitude can serve as a means to good behaviour. In his early days, Kant explicitly rejected Hume’s idea that humility [Demut] was to be considered *a monkish virtue* and claimed, on the contrary, that it was needed also in natural morality. It is thus important to implant both humility and self-esteem in human beings in order to avoid unwanted effects of a spoiled upper-class, upbringing, luxury and pomp; in the light of this, we are told: “Rousseau’s ideas become attractive”. (VM [Herder] 27:39)

However, in later works Kant does not seem to count a humble attitude in relation to other human beings as a moral quality at all. Not only does he explicitly deny that humility in this sense counts as a duty, he also says that such an attitude and the comparison that is required in order to be humble in relation to other finite beings often leads to an ambitious idea of oneself as being a better human being. This is, of course, directly contrary to the duty to others. Moreover, such behaviour will also easily fall into hypocrisy and flattery and is thus nothing but false humility. This humility, Kant says, can also be seen as a diminishing of one’s own worth with the aim of gaining worth in the opinion of others and this is contrary to the duty to oneself since one reduces the worth of one’s own personality. (MS 6:435) As Jeanine Grenberg puts it:
To feel humiliation while lacking a respectful awareness of what humiliates (that is, authoritative moral principles), or while lacking the self-approbation which these principles also ground would not be a moral feeling at all; it would be simple self-abasement. Similarly, to feel an unrestrained self-approbation without an awareness of its proper ground, and with no awareness of the need to constrain the corruption of that agency which is simultaneously the ground of one’s worth, would not be a moral feeling; it would be moral fanaticism, or simple arrogance.\(^{330}\)

While I do not think that these lines provide the only possible way to understand humility within a Kantian framework, Grenberg’s idea basically resembles Kant’s own depiction of moral humility. Admittedly, it seems as if Kant, when he rejects humility as a quality of moral significance, is referring to humility in a more general sense as a kind of “submissiveness” and not as the “correct estimation of one’s own worth” that was the meaning of the term in Herder’s notes from the early 1760s. As is clear from Kant’s writings, moral humility as the “correct estimation of one’s own worth” can only be felt in relation to the moral law. Accordingly, the awareness of the law and the respect that this awareness brings about also causes the right kind of humility that in Kantian terms consists in the insight of the “worthlessness” of our own wants when they are compared with the endless value of this holy command. This characterization – which I will here call “Kantian humility” – seems to be the only kind of humility that he explicitly acknowledges as being of moral significance. Still, the idea that every form of humility except “Kantian humility” must, as Grenberg’s comments indicate, necessarily end up either in “simple self-abasement” or in “moral fanaticism or simple arrogance” is only correct if one accepts that the alternatives that she attributes in her presentation cover all possible forms of humble behavior save the practice of true “Kantian humility”. Yet, by doing that, one ignores the possibility of accepting a conception of “ordinary humility” that, though not qualifying as a moral feeling, still must be regarded as a feeling that is necessary for the possibility of the moral awareness which, in turn, is necessary for the possibility of Kant’s moral notion of humiliation. As we will see, ordinary humility is not “simple self-abasement” nor does it, because of its non-moral status, run the risk of lapsing into any form of false humility. In order to get a better picture of the need for “ordinary humility” in Kantian morality and to link this feeling to the overall topic of this chapter, I will consider Kant’s arguments for the duty of beneficence.

7.3.1. Humility and Beneficence as Duties to Oneself
Kant divides the set of duties to oneself into duties to oneself as “merely animal being” and “as a moral being”. Our duties to our animal nature in-

\(^{330}\) Grenberg p. 154.
volve the preservation of our life and the correct use of our basic natural drives. These demands forbid suicide, excessive consumption of food and drink and unnatural use of the sexual inclination. (MS 6:420) However, for our discussion here, we are interested only in the duties that have to do with our exclusive human nature or our nature as rational beings having an absolute moral worth. The foremost violations of these duties are lying, avarice and servility (i.e. false humility – or humilitas spuria) and by violating the duties that one has to oneself one throws oneself away and makes oneself an object of contempt. These duties are thus to be considered as perfect. However, besides these, we have a general duty to develop and increase both natural and moral perfection. While these last duties are only imperfect, we will see that they are of the greatest importance for the conception of the moral human being.

Accordingly, as Kant himself puts it, duties to oneself do not consist in material things, like dressing oneself in fancy clothing or eating delicious food. Nor can these duties entail the satisfaction of desires or inclinations in general for that would contradict a central virtue of human beings namely that they be moderate and temperate [mäßig und enthaltsam]. Rather, Kant says, what we owe to ourselves in this respect is simply to not deny the dignity of our own person. (Päd. 9:488) Consequently, in order to fulfill the duties to ourselves, we ought to realize our proper worth as objects of respect in virtue of our humanity. This insight is basically what Grenberg referred to as the form of self-approbation that the law grounds in us; a proper self-esteem or, in Rousseau’s terms, amour de soi. At the same time, we must, of course, not believe that this incomparably high value allows us to behave self-importantly or that we have a right to do whatever we want – not even to things or beings that are not objects of respect. The reason for this is simply that by being harsh, or mean, towards other beings, including those one has no direct duties to (like animals) one makes oneself a worse human being.\footnote{331 This is spelled out in §17 of the “Doctrine of Virtue” (MS 6:443) and in Collin’s notes on ethics VM [Collins] 27:458-9 as well as in Friedländer’s manuscript from the anthropology course. (VA [Friedländer] 25:727) Admittedly the argument is clearly anthropocentric since the reason why cruelty to animals is considered bad is that such human beings are likely to be cruel also to other human beings. By being mean to animals one may also upset people who care about them. Still, the fact that Kant saw the dog’s faithfulness or the horse’s long service as something that is analogous to merit, and that we thus have a duty to take care of them into their dotage, at least indicates that such animals are not mere “things”. However, the main reason why I ought to treat animals in a good way is not that they have any rights to such a treatment but that I thereby cultivate my humanity by not damaging the kind and humane qualities of myself.}

Accordingly, the accurate way of understanding the essence of the duties we have to ourselves is, as we have seen, in terms of a duty of self-perfection [Vollkommenmachung]. (VM [Powalski] 27:188) In other words, we ought to make the best of ourselves and develop our full potentiality. We must be trained for all sorts of ends and not let our talents rust inside us. (MS
This agrees well with the arguments of the *Anthropology*, as well as with the *Pedagogy* where we are repeatedly told that the human being ought to strive towards perfection, no matter how strong his animal impulses may be. Man’s destiny is to fight crudity and to *cultivate*, to *civilize* and to *moralize* himself in order to bring forth morality in himself. (*Anthr.* 7:324-5; *Päd.* 9:446) As we will see, this claim of perfection also includes the demands for benevolence and beneficence.

The understanding of duties to oneself in terms of moral self-perfection explains why the duty to always tell the truth, be grateful, generous and so on are not, as one might believe, primarily duties to others but to ourselves.332 Moreover, this conception also shows that some of the duties that we have to other people may contribute to our perfection and that they, thereby, can be seen as *indirect* duties to ourselves. Among these duties we find the duties of love, benevolence, beneficence and compassion. This idea is spelled out in the draft to the *Metaphysics of Morals* in which we are told that duties to others are, in fact, indirect duties to ourselves consisting in the demand not to destroy one’s soul by hating a friend. (*Vorarb.* [MS] 23:403) This is in perfect coherence with Powalski’s notes from the early 1780s where we are told that beneficence, though only an imperfect duty, is to be regarded as a duty to oneself with regard to the fulfillment of one’s duty of humanity. (*VM* [Powalski] 27:188) One needs not love others, or even find them loveable, in order to make their happiness an end that is also a duty for us. Practicing active benevolence by being beneficent will thus produce actual love of them, which in Kant’s view is truly a general love of humanity. In the long run, this love can be seen as an aptitude for further benevolent deeds. (*MS* 6:402)

### 7.3.2. Humility and the Law

Because of Kant’s focusing on the divine nature of the moral law and our immediate duty to subject our will to it, there is no doubt that humility is a key feature in Kantian ethics. It is only through the awareness of the insignificance of our own interests as compared to the law that we can feel respect for it, but Kant strongly rejects that there can be such a duty when it

---

332 According to Kant, a man who lies throws away all his dignity and does not deserve more worth as if he were a mere thing. (*MS* 6:429) The reason why truthfulness is a duty is spelled out in Kant’s short text *On the Supposed Right to Lie from Philanthropy* where we are told that in virtue of being an object of respect, man has a right to the truth, including his own. This is why it is a duty to tell the truth even to the murderer since, by lying, the man would rob himself of his dignity. This text was written as a reply to Benjamin Constant who suggested that there is a duty to tell the truth only to someone who has right to the truth. For Kant, this very phrase “right to the truth” expressed an “original falsity” since it ignores the fact that everyone has the right to his own truthfulness, that is, to the subjective truth in his own person. (*RL* 8:425-6) See also VM [Vigilantius] 27:702.
comes to humility in relation to other beings. From what is said in the *Meta-
physics of Morals*, this form of humility entails comparison with finite be-
ings that too often produces competitiveness that, in turn, may lead to self-
conceit. At the other extreme, comparison with others might cause shame
and servility (i.e. false humility), which, as we saw, is contrary to the respect
for humanity in our own person.

In spite of Kant’s clear rejection of “ordinary” humility as a *moral* quali-
ity, the understanding of humility as opposed to self-conceit is essential for
the possibility of *moral receptivity* and not only of the *moral subjection* that
is caused by the feeling of respect for the law. To see the importance of this,
we must once again recognize the full sense of self-conceit. First of all one
must accept that all human beings are, to varying degrees, affected by self-
love, or *philautia*, which means that we tend to act from subjective incen-
tives rather than from the objective principle of respect for the law. While
self-love may be (and are often) in accordance with the law, it need, because
of its unreliable nature, to be restricted by moral awareness. Self-conceit, on
the other hand, is not merely deviation from this law. To be conceited is, as
we have seen, to make one’s subjective wants into law-giving principles. In
this case it is not enough to simply “restrict” and “modify” these principles
into conformity to the law; self-conceit must, as we have seen, be entirely
eliminated, struck down and in this way be *humiliated* by the law.333

Now, the definition above expresses a classical Kantian idea, namely that
the law’s absolute authority makes us aware of our limited nature and this
causes respect and humiliation. However, as we saw in chapter 6, there is
also another dimension of self-conceit which deals with conceitedness, not in
regard to the law, but in relation to other human beings. In this sense, self-
conceit is expressed through one’s immoderate claims on other peoples’
respect and through one’s representation of oneself as superior to others with
respect to personal worth. (Cf. *MS* 6:462, 6:465) The person who is conceit-
ed in this sense is indifferent to the needs of others and he thus thinks that
they have reasons to despise themselves. As Kant’s Latin definitions of
Eigendünkel (*arrogantia* and *superbia*) indicates, self-conceit is a perfect
description of *arrogance* or *haughtiness*, and this is as far from moral behav-
ior as one can possibly come. The person who is conceited in this sense is –
because of his flawed idea of the worth of his own person as compared to
others – short of every sign of compassion, humility as well as respect for
other human beings. In view of this, Kant’s claim that humility in relation to
others is not to be considered a duty becomes quite surprising. But, what
seems even more puzzling is that a trait that he *does* treat as a duty, namely
*moderation* [Mässigung] of one’s demands regarding love and respect from
others, is derived from the duty of respect to them. (*MS* 6:467)

In the light of this, I believe that true Kantian respect can arise only if we already possess a humble and modest disposition in the ordinary sense. Kant’s “higher” form of humility thus implicitly seems to entail the ordinary conception in the sense of awareness of one’s own imperfectness and limitations that make us subject to misfortunes and various needs. It is, after all, hard to see how someone who believes himself to be superior to others could apply the categorical imperative in the right way, and the need of ordinary humility for this practice is, in fact, visible also in Kant’s own argument. When he tests the first formulation, the fourth example aims to expose the impossibility to will – even for a man for whom things are going well – that it is up to everyone to take care of themselves and that there are no reasons for anyone to contribute to the welfare of others, nor to assist them in case of need. To this Kant adds that although this way of thinking is not contrary to any laws of nature, a rational being would realize that he could not possibly will this maxim as a universal law since he would thereby “rob himself of all hope of assistance” if he himself happened to be the one in need. (GMS 4:423) While Kant does not make this point explicit here, it is clear that this argument requires an insight of one’s subjection of the misfortunes of life. Without this humble understanding, the very claim that he, by accepting the maxim of universal egoism, would “rob himself of all hope of assistance” would, after all, be of no consequence to him.

The same idea reappears in the discussion of the duty of beneficence in the Metaphysics of Morals. Here, we are told that the awareness of our own need of love and help from other people leads to the idea of treating everyone’s ends as one’s own. The reason why I ought to be benevolent towards others is simply that I want everyone else to be benevolent toward me and, in order to qualify as a universal law, a demand cannot be allowed to be selfish. (MS 6:394 6:450) Even if such an insight might be rationally grounded, this ground presupposes the existence of humility and at this time, not in the Kantian sense, but in the ordinary sensible awareness of oneself as not being immune to misfortunes and that one might need other’s benevolence as well as their help.

Here again, we see the essence of Rousseau’s second maxim for if we – like kings – would look upon ourselves as if we were self-sufficient and entirely distinct from ordinary men, we would neither feel genuine compassion, nor would we be moved to help a fellow, unless we did so out of a disrespectful sense of superiority. Accordingly, even if Kantian humiliation is the effect of the moral law, this subjection will only work if we already have a sound picture of our worth as well as our status as limited beings. A self-conceited person would frequently ignore the law, no matter how much it tries to “strike down” or “humiliate.”

Perhaps, it is situations like this that made Kant conclude that beneficence is better performed from duty and not because of any “proud conceit” [stolzer Einbildung] derived from the impression of oneself as a good-
hearted “volunteer” independent of moral commands. \( KpV \ 5:82 \) Beneficence from duty would obviously involve no such degrading feelings, nor would it entail vain pride since the law itself has made us aware of what we should do and this is, as we have seen, the same law that makes us humble. At the same time, one might object that the conceitedness that the one who is beneficent from inclination might express rather indicates that there is no true benevolence behind the act but rather a self-interested drive. As we have seen, benevolence and beneficence have their root in ordinary humility, which is also a requirement for the love of humanity in general.

7.4. Becoming Humans by Forming Nature

As was mentioned already in chapter 1, Kant held that the human being can only become human through education. \( Päd \ 9:443 \) This indicates that for Kant, the mere belonging to the human species \( (\text{homo sapiens}) \) is not enough to be regarded as truly human. From what we have learned so far, it is clear that to be human in the proper sense involves a development of higher faculties like reason and especially moral reasoning and social skills.

In chapter 4 we were told that ethics refers exclusively to Gesinnung and that everything moral must refer to the goodness (or evil) of this fundamental ground. Moreover, in the discussions of a moral character we saw that this feature of the soul was wholly distinct from our natural aptitudes [Naturell] as well as from temperament that includes our sensibility and other non-rational aspects of the soul. In the Pedagogy Kant says:

\[ \text{We come now to the culture of the soul, which in a way can also be called physical. One must, however, distinguish between nature and freedom. Giving laws to freedom is something entirely different from forming nature.} \ \text{(Päd. 9:469, emphasis added)} \]

In spite of this emphasis on the difference between natural and moral education one must not fall into the trap and conclude that education towards perfection of one’s nature as a truly human being consists solely in moralization, by forming of the character and subjection to moral laws. Surely, Kant says, even a human being who is “highly cultivated in physical terms” (and thus has a well-formed mind) may still, because of a lack of moral culture, be an evil creature. While this is wholly consistent with Kant’s arguments from the conception of evil outlined in the Religion, one must not believe that a well-formed nature is entirely without importance for the moral human being. As we saw in the previous sections, without a sound idea of ourselves as limited beings we would neither be humble, nor would we be likely to respond in an appropriate way at the sight of a suffering fellow. Accordingly, while proper moral education is only concerned with the development of
character, the formation of the lower parts must not be belittled for men, we are told, fall either into *philautia* or *arrogantia* depending on differences with regard to their temperament. (*VM* [Collins] 27:358) In its basic sense, *philautia* (or self-love) is fairly harmless and can easily be moderated into conformity with the law while *arrogantia* (or self-conceit) is a real threat to moral behaviour. Hence, in the progress towards the moral character and a general moral awareness of our insignificance with respect to the holy law, we need to begin with a well-formed mind. In other words, before we can begin with moral culture we need to cultivate nature and sensibility.

Now, it could be objected that Kant does not claim that cultivation and civilization of our natural aptitudes are without significance in the overall educational plan but only that it does not contribute to our moral thinking. The good nature, we are told, is *passively* good while the good character is *actively* good. (*Refl.* no. 1126, 15:502) A good temperament is not stable but rather, as Kant puts it, as a watercolour painting. (*Refl.* no. 1518, 15:869) Goodness from temperament is not enough to make a good man since goodness must, as we know, have its foundation in character. However, if we recall what was said in the dramatic passage in the *Religion* where Kant tries to explain how man, in spite of his evil propensities corrupting the very root of his maxims can advance towards goodness through a revolution, we will see that sensibility has a role even here. For, as we have seen in earlier chapters, the only way by which we can make this change is through a “revolution of the Denkungsart” but we have also seen that this revolution must be accompanied with a “gradual reformation of the Sinnesart” since the latter “places obstacles to the former”. (*Rel.* 6:48) This is the only way by which our mind’s moral attitude can establish character, which also is the sole way of proceeding beyond the state in which we do not only fight specific vices but also removes the universal root of all human evil. So by refining sensibility we help the way of thinking to develop into a character that, as the law-making part in us, is the faculty that sets our ends, but which constantly has to struggle with inclinations. By “forming nature” we may cultivate the right forms of sentiment that provides beneficent inclinations conducive to moral thinking. A character created this way, i.e., not exclusively by way of intellectual improvement, will perhaps be better disposed to incorporate the incentive of duty. This understanding of character seems also to make sense of Kant’s demands of cultivation expressed in the *Metaphysics of Morals*.

### 7.4.1. Humane Human Beings

Since the chief aim of moral education is to form human beings with a strong soul and a good character we need to find out where to begin. It seems reasonable to assume that a “good human being” is someone who makes correct use of his human characteristics and especially of those that are exclusive to his species. The most obvious human trait is, of course, rea-
son but, as we have seen, rationality alone cannot capture the entire essence of human beings and much of what we appreciate in others can, after all, be seen as emotional qualities or social virtues like humility, compassion and kindness.

In Emile Rousseau exclaims “Men, be humane [humains], this is your first duty. Be humane with every station, every age, everything which is not alien to man.” (E II 79/302, FM 25/82) To be humane, Rousseau says, is not merely to give alms but rather to give charity. It is, he adds “in doing good that we become good.” (E II 95/326; E IV 250/544) Rousseau obviously seems to have adopted the idea that the practice of social virtues is the condition for moral goodness and he does clearly no longer treat sociability as alien to human nature. On the contrary, as he says in the Letter to Beaumont, “the most social man is also the most humane”. (LB 54/969) We may also recall the dictum from the Moral Letters and the Vicar’s speech that man is social by nature or at least “was made to become so” and without sociability, he would remain a half-finished creature. The importance of social virtues for morality as well as for humaneness irretrievably highlights the importance of proper upbringing aimed at refining what is already there, since this method will not involve the risk of planting any illusionary ideals in the child’s soul. Here we may once again recall what was mentioned above, what truly constitutes humanity and humaneness is the ability to, by means of identification and imagination, share the feelings of others.

Kant was, of course, not ignorant of the need for social relations for our progress towards the human vocation. Without the feelings that arise from sociability, the human being would be considered inhuman or, better, inhuman. To be humane, we are told, is to interest oneself in the fate of other men; inhumanity on the other hand, is to take no interest in what happens to them. (VM Collins 27:443-4; 27:419) The humane heart is that which, through feeling, puts itself in the place of the other while the inhuman man is heartless [Lieblos]. These two correspond to the good and the evil mind. (Refl. no. 1497, 15:767) Humanity is also linked to “the capacity for being able to communicate one’s inmost self universally” and this capacity signifies the “sociability that is appropriate to humankind”. (KU 5:355)

This basically reflects the idea expressed in §34 of the Metaphysics of Morals for we are told that humanity [Menschlichkeit] can be placed either in the capacity and the will to share in others’ feelings (humanitas practica) or in the natural receptivity to the feelings of joy and sadness that we experience in other human beings (humanitas aestetica). While only the former is said to be free and rationally grounded, it is the latter, because of its communicable nature, that actually make us feel what others feel. Accordingly, it seems clear that not only Rousseau, but also Kant, connected the term humaneness [Humanität] with a general feeling of sympathy or of “taking part”

---

334 E I 37/245, E IV 290/600, ML V 196/1109.
[Teilnehmungsgefühl]. It is a reciprocal well-wishing tied to a likewise mutual respect. Pity and sympathy are, therefore, to be called humane since they distinguish man from the beasts. Moreover, the very opposite of pity, Schadenfreude, or happiness at others’ misfortune, when someone “laughs when others weep, and feels pleasure when others feel pain” expresses genuine cruelty and can be seen as an example of inhumanity. 335

Accordingly, lack of pity is contrary to the concept of humanity, or humaneness. Schadenfreude, the very opposite of pity, is by Kant also considered a diabolic vice. (Rel. 6:27) Moreover, in the “Doctrine of Virtue” Kant explicitly put the duties of compassion and “love of human beings” under the title of “duties of humanity” (humanitas, Menschlichkeit) and in the draft to this work, Kant distinguishes between “Menschheit” and “Menschlichkeit” by explaining the former in terms of humanitas substantialis and the latter in terms of humanitas accidentalis. (MS 6:456; Vorarb. [MS] 23:398) Again we see that, for possessing “humanity” in its full sense, it is not enough to be the rational living beings that we are by nature, but we need to complete our existence with specific and occasional features that we must acquire and cultivate. The reason why some fields of study are called humaniora is that they refine people in this respect. (VM [Collins] 27:419)

7.5. Conclusions: Sensibility and Ethics

It is no exaggeration to say that §35 is an interesting passage that seems to add several perspectives to Kant’s idea of the moral person especially since it seems to conflict with most of what he previously had to say on this matter. At the same time, it is a highly difficult paragraph that requires much of the interpreter. In order to see the richness of it one must look beyond the feelings that are explicitly discussed in this passage, namely the “active participation” or compassion. Though the themes that are brought about by such an analysis reveal some of the most Rousseauian aspects that can be found in Kant’s corpus, Kant himself does not mention this origin. The discussion above, therefore, aimed to disclose the close interweave of Kant’s and Rousseau’s thinking with respect to this matter and, as we have seen, they are more often in agreement than in genuine conflict.

While §35 is confronting us with an ethical problem (which also might be treated from a purely moral point of view) I have argued that we, in order to reach a more complete understanding, ought to analyse it against the background of Kant’s theory of human nature. According to a plain reading, the section makes the simple demand that we ought to make ourselves aware of natural and social evils, and thereby open our souls and let the irresistible

feelings come over us. I do not doubt that this method of producing feelings of uneasiness will work for most human beings but, as I have argued, the possibility of cultivation of pity in this way does, after all, require an already well-formed soul. This pre-formation requires a sound idea of what we really are combined with a lack of excessive ideas of what we might have been. As I have argued, this insight can justifiably be explained in terms of humility or modesty and these are needed for the correct use of our social skills. Again we need to acknowledge the need for all those so called “subjective conditions” in our nature since these factors make up the very foundation of all applied morality.

Kant clearly accepted that compassion, while in itself not a duty, can – if cultivated – assist us in the execution of imperfect duties – like the duty of beneficence, but in doing so, he must accept further feelings and social qualities that serve as conditions of the possibility of this cultivation. This was, however, not all; it should have become clear that some of these supporting aspects of sensibility also serve as conditions for moral awareness in general, even for the possibility of awareness of perfect duties. Humility, for instance, is crucial and not only in the Kantian sense of humiliation in relation to the moral law. Self-conceit – the very negation of moral attitude – can only be avoided through proper self-knowledge and the awareness that one does not have a greater worth than anyone else. This is something that must be grounded in the education of our empirical nature for the development of self-conceit (and hence also our avoidance of developing it) has its root in our predisposition to humanity. This predisposition is, admittedly, partly rational, but still physical and it is from this predisposition that all social passions originate. Social passions are, for example, envy, pride, ambition, jealousy and, too often, self-conceit. It is, therefore, not that far-fetched to say that this predisposition principally grounds a basic sensibility or temperament that, by means of reason’s guidance, serves as the foundation for our highest predisposition and the possibility of choosing “nothing but good ends”, which is also the end of moral education.

Still, one must not forget that the cultivation that is prescribed in §35 cannot be regarded as an act of moral culture. By cultivating natural (aesthetic) feelings we merely “form nature” and while this is, indeed, an important undertaking in the educational process, it is entirely different from giving law to freedom. Accordingly, it is clear that the worth of this feeling, as well as the cultivation of it, is only conditional. Therefore, a sympathetically cultivated temperament is not good in itself and compassion cannot serve as a moral incentive. This seems to mirror the familiar view from the Groundwork. In this work, we also see that the notion of hypothetical imperatives is connected to talents in the forms of skillfulness [Geschicklichkeit] and prudence [Klugheit] that often appears in Kant’s discussion of the use of talents and other gifts of nature, but not in relation to morality. If this is a plausible reading, the duty in §35 is not a moral command but merely some kind of
advice about what we may do in order to perform an act that is in accordance with duty. Since Kant here clearly states another goal for this cultivation (active benevolence), a categorical reading is excluded and the imperative is merely hypothetical. Active sympathy, that is beneficence, is brought about by means of a sensible cause, which is something other than respect for the law and cannot, therefore, be moral in the strict sense.

Still, before closing this chapter, I would only like to say a few words about the passage so frequently cited in order to highlight the role of Rousseau for Kant’s moral thinking. In an early remark Kant wrote:

I am myself a researcher by inclination. I feel the entire thirst for cognition and the eager restlessness to proceed further in it, as well as the satisfaction at every acquisition. There was a time when I believed that this alone could constitute the honor of humankind, and I despised the rabble who knows nothing. Rousseau has set me right. This blinding prejudice vanishes, I learn to honor human beings, and I would feel by far less useful than the common laborer if I did not believe that this consideration could impart a value to all others in order to establish the rights of humanity. (BB 20:44)

Clearly, it was the insight of the unjustified conceitedness that had been a part of his thinking – the idea that he, as a passionate academic thinker, was a better human being than the common man – that made him realize the proper ground for the unconditional value of humanity. Thus it was a sense of “ordinary humility”, as a humble feeling in relation to other human beings that revealed the common ground that they all shared. This common ground, in turn, brought about the new notion of “Kantian humility” based on a higher conception of the notion of humanity. Rousseau was the one who brought about that awareness. So, perhaps, one might conclude that Rousseau was the condition for the realization of the conditions that are, in turn, conditions for compassion as the feeling that we need in order to be fully human – or humane – in the true Kantian sense.
About half a century after Kant’s and Rousseau’s great days, Arthur Schopenhauer wrote in his magnum opus that it is clear that one can never reach the essence of things from the outside; no matter how much we look, we find “nothing but images and names”. We are, he says “like someone who walks around a castle, looking in vain for an entrance and occasionally sketching the façade.”

This illustrates plainly what is needed for a genuine study of human nature. Both Kant’s and Rousseau’s descriptions of the development of a social state involve a change of the perspective of the individuals. For Rousseau, the leaving of the state of innocence was essentially slipping out of ourselves and, in contrast to the savage, who lives completely within himself, the modern man lives outside and can no longer perceive anything without making comparisons. In other words, he lives in the opinion of others. Also Kant saw the problems that had to be dealt with for the human being to be able to hold on to his principles in a state in which men constantly had to relate to each other. Although a return to the woods is no longer an alternative, a return to ourselves might be possible. This, however, requires that we know which self we are to embrace as our true selves.

So, what can be said to have been the essence of this thesis? I believe that it may be summarized by a sentence written down by one of Kant’s and Rousseau’s favourite authors. In his highly popular work An Essay on Man, Alexander Pope wrote:

\[ \text{Know then thyself, presume not God to scan,} \]
\[ \text{The proper study of mankind is man.}\]  

We are thus thrown back to the ancient maxim that Rousseau once adopted as his leading principle. In knowing ourselves we learn about our subjective conditions. We learn how to make the best of ourselves. As Rousseau so clearly put it: “the jargon of metaphysics has never led us to discover a single truth, and it has filled philosophy with absurdities of which one is ashamed as soon as one has stripped them of their big words.” (E IV 274/577) By this, he wanted to say that we must speak of man as he is in-

---

336 Schopenhauer, Die Welt als Wille und Vorstellung I. Book ii §17.
337 Pope, Essay Epistle 2 (first sentence).
stead of hiding him beyond obscure terms. In contrast, Kant used the knowledge of man – anthropology – in order to facilitate our progress towards full moralization, that is, to acquire a character that always makes us transport ourselves into the moral realm as if we were independent of our pathological selves. Still, as we have seen, this man, as ideal moral being, is not a human being. Nevertheless, it is clear that Kant believed that the higher perspective, that is, metaphysics, is a necessary aspect of the theory of the human being by means of stating our ultimate purpose. From this follows that the purely natural man, whoever he is, or was, cannot be considered to be fully human.

Now, in order not to overstate the apparent difference between these two outlooks we first must bear in mind that Rousseau explicitly agreed to this last claim. However, rather than putting the ultimate source of morality in some transcendental source, Rousseau found it in man’s most basic principle – amour de soi – by claiming that it must be regarded as involving two aspects. Besides the natural instinct, it encompasses an intellectual perspective. This latter aspect needs to be awakened and fortified by reason. When developed, it bears the name conscience and is, in line with Kant’s moral compass, wholly independent of all social constructions. Second, we should heed Kant’s admission that also the “second part of morals” is needed in order to understand the full meaning of a human morality and that this indicates his belief that we need to take both parts into account if we will ever be able to understand also the full meaning of the human being. This need is obvious as soon as we realize that human beings are essentially moral beings or, at least, beings that must be regarded as susceptible to a higher form of representations in order to be something more than mere animals. In a way, both these philosophers presented an idea of the human being as a double-natured creature whose spirituality was disclosed to him in the consciousness of his ability to be free in the sense that he subjected his will to a freely chosen command that he found in the depth of his moral self.

Human Beings – Human Nature

In the preceding seven chapters I have brought up various themes that are actualized by the subjective conditions that constantly operate within the human mind. As we have seen, it is not always easy to navigate within this area, and commentators have here a never-ending source of discussions. While I do not believe that the topics that I have discussed have any indisputable solutions, my main aim has here been to expose how anthropology infiltrates numerous specific themes by intervening in some recent debates.

In an attempt to summarize this analysis I will return to Kant’s three fundamental questions about what we could know, hope and ought to do; ques-
tions that were unified through the fourth and anthropological one – what is the human being?

As we have seen, the last question is not easy to answer and history has made clear that it cannot be answered without being cut up into some sub-questions, since there is not one “man” or “human being” that can be the standard model for the answer of the question of what “man” is. Even if it is fairly clear that Kant, when using the phrase “human being” in this context, referred to the present man – and not to the pre-historical idea of a wild man in progress – one must not believe that this conception of man can be understood without its historical context. The present man is the product of a progress enforced upon him through his destiny. It is a necessary evolution from savagery to civilization via a lawless social state. At the same time, the present man is in constant progress towards full moralization. Man stands under moral laws which he, however, does not always subject himself to. He is therefore evil, or, at least, morally depraved, because of his reversing of the order of his various incentives that he, at the time of choice, makes the determining ground of his maxim. In this he is, in all relevant respects, the same man that Rousseau pictured as the wicked man, whose wickedness we could see with our naked eyes.

However, perhaps one can argue that the question is put in the wrong way. Kant underlines that anthropology in the strict sense is not a science that searches for a description of human beings but rather a science of human nature. While human beings constantly change, humanity does not. The conception of humanity that we establish must thus be applicable to all human beings, across time and because of this, we must find a reasonable explanation for the apparent changes that human nature has gone through during the path of history. While this might seem to stand in sharp contradiction to Rousseau’s suggestion that the human being of today is not the same as he was before the irreversible turn from the wood to society, one may here argue that the difference is most likely an illusion. As Rousseau himself put it in the Dialogues, true human nature will be revealed if someone simply removes the rust that hides it. (JJR III 214/936) We may here also recall his metaphor of the eroded statue that symbolised the present man, and this indicates that the man we see, however different, is nevertheless based upon the same ground as the original man. We are all born natural and free but end up behind the mask of society. Again we may recall Kant’s appraisal of his colleague’s wisdom: “Rousseau discovered for the first time beneath the multiplicity of forms human beings have taken on their deeply buried nature and the hidden law...”. (BB 20:58-9)

---

338 VA [Friedländer] 25:471. Cf. also EF 8:374 where Kant explains that to know human beings is not to know the human being. For knowing the latter we need the higher standpoint of anthropological observations.
Accordingly, there is said to be an essence that survives all external changes and, as we have seen, a total change of nature takes more than a change of our particular outlooks. We grow, develop, learn and are forced to cope with various external influences that trigger our subjective conditions that make us what we are. We may here once again recall the ambiguity of the term nature that, if taken in its most basic sense as the opposite or negation of reason and culture, would be a term of no moral import for, as we were told, by nature we are not moral beings. Yet, the most central conception of nature that we have been concerned with in this exposition is the one that Kant put forward in the Religion. Nature in this sense signifies the “subjective ground – wherever it may lie – of the exercise of the human being’s freedom in general.” (Rel. 6:21) Practically all of the themes we have encountered above are related to this conception of nature as the ultimate source of the ideas of good and evil. In spite of its epithet nature, this is a source that only a rational being can reach, and one that connects us to a higher order of things. The feelings caused by this super-sensible awareness filled Rousseau with love of that moral order, while for Kant it was a source of shuddering respect and awe. It is here, in this kind of nature, that we find what we need in order to become human beings.

What Have We Learned from Human History?

In order to find the entrance to ourselves that, in turn, is needed to carry on the search for the exit of the artificial theatrical life as objects of the opinion of others, we need to step back and once again search, not only for the answers, but also for the right questions. Accordingly, because of the vagueness of the plain concept of a “human being”, we must rather ask “what is man?” “who is the civilized man?” and, when we know the answers to both these questions, we may ask how these two men are related and which of them, if any, that is to be preferred to the other. Kant apparently had something like this in mind when he wrote the final pages of the Anthropology:

Rousseau wrote three works on the damage done to our species by 1) leaving nature for culture, which weakened our strength, 2) civilization which caused inequality and mutual oppression, 3) presumed moralization, which brought about unnatural education and the deformation of our way of thinking. These three works, I maintain, which present the state of nature as a state of innocence (a paradise guarded against our return by the gatekeeper with a fiery sword), should serve his Social Contract, Emile, and Savoyard Vicar only as a guiding thread for finding our way out of the labyrinth of evil with which our species has surrounded itself by its own fault. Rousseau did not really
want the human being to go back to the state of nature, but rather to look back at it from the stage where he now stands.\textsuperscript{339}

It is pointless, Kant says, to discuss whether any of these changes were for the better or the worse since there is no return. However, he adds, the change needs to be reviewed from two distinct positions, for it is obvious that it must, at least on the moral side, be considered a fall. Also on the natural, physical side this alteration was a loss since we were thereby exposed to so many new afflictions that may very well be seen as punishments for this fall. In a truly Rousseauian manner, Kant thus states: “The history of nature begins with goodness, for it is the work of God; but the history of freedom begins with evil, for it is the work of the human being”.\textsuperscript{340} Yet, it was solely through this change that we became human beings, and so, for nature – which sets its ends for the human being in the species – this was a gain.

Once again it is clear that Kant was deeply familiar with Rousseau’s thought about the matter, and he also realized that the appeal to history was most of all a method for answering the question of man and his nature. When Kant referred to Rousseau’s state of nature in his two historical works, he deplored the fact that his forerunner had so often been misinterpreted, and claimed that Rousseau was undoubtedly right when he preferred this state to the social state as presented in the first and second Discourse. Rousseau had also correctly demonstrated the unavoidable conflict between nature and culture. Kant’s faith for Rousseau’s conclusions concerning this matter is even more explicit in Mrongovius’ lecture notes where we are told that the intermediate state between the crude and the cultivated state is the worst, because it involved the appetite for luxury, refinement, taste and sociability.\textsuperscript{341} Although these notes indicate that Kant (or at least Mrongovius) accepted Rousseau’s favoring of the uncultivated state of nature to this one, he clearly states that none of them express the truly human state, as that in which man has, if not fully reached his vocation, then at least made considerable progress towards it. Indeed, Kant says, even Rousseau had to admit that the true problem of the understanding of humanity did not concern the division between the savage and the citizen, but rather the progress towards a state where each individual became fully skilled and educated while still remaining true human beings.

In another lengthy comment on his colleague Kant says:

\textit{in his Emile, his Social Contract and other writings, he seeks again to solve the harder problem of how culture must proceed in order properly to develop}

\textsuperscript{339} Anthr. 7:326-7. The three works mentioned are generally considered to be the two first Discourses and Julie. Although the latter is not primarily a philosophical work, it reflects the author’s views on several moral as well as social matters.

\textsuperscript{340} MAM 8:115 paraphrasing Rousseau’s “Everything is good when it leaves the hands of the Author of things; everything degenerates in the hands of man”.

\textsuperscript{341} VA [Mrongovius]. Cf. also Reft. 1498, 15: 778.
the predispositions to humanity as a *moral* species to their vocation, so that the latter no longer conflict with humanity as a natural species. From this conflict (since culture, according to true principles of education of human being and citizen, has perhaps not yet rightly begun, much less having been completed arise all true ills that oppress human life, and all vices that dishonor it; nevertheless, the incitements to the latter, which one blames for them, are in themselves good and purposive as natural predispositions, but these predispositions, since they were aimed at the merely natural condition, suffer injury from progressing culture and injure culture in turn, until perfect art again becomes nature, which is the ultimate goal of the moral vocation of the human species. (*MAM* 8:117-8)

One of the most central lessons of human history is that human beings are essentially *political* beings that need other laws than those of nature. This progress is also, as being the chief aim of the innate vocation, to be considered *natural* even if it appears to conflict with the human being’s original way of living. However, the last sentence in the quote is important and shows that the so called artificial culture – that once replaced the existent nature – may evolve and, thereby, become a *new form of nature* that, at least eventually, might become as natural as original nature was. Kant repeats this argument several times, often in relation to Rousseau’s elevation of “nature”. He is right, Kant says: we ought not do violence to nature but rather develop it to completion – let it flourish and bear new fruits. 342 This final state where “perfect art becomes nature again” represents the closing of the circle and is the best state that human beings can reach. (*Refl.* no. 1489 15: 778)

In the light of this, it seems that not only Rousseau, but also Kant, saw “nature”, or “naturalness”, as an ideal, at least in the sense that we should live in accordance with our *nature*. However, according to Kant, this required another conception of nature than the original Rousseauian one. Or perhaps, this state expresses a kind of “civilized naturalness” that, as we have seen, is the highest level of present humanity. We may once again recall Kant’s most philosophical idea of “nature” as the subjective ground within us that regulated the use of our freedom. We are essentially free but we do not necessarily use this faculty in the way we should. Let us also recall Kant’s claim that the human being necessarily needs a *master*. This idea is neither evil, nor misanthropic, but rather the sole and rational conclusion that can be drawn. He needs a master in order to control his stubbornness. He needs to be forced to subject himself to a universally common will – a moral will. In spite of some central differences, the Rousseauian idea of a general will is not that far away, and I believe that it is safe to say that without Rousseau, Kant would probably not have written his two historical works, and his anthropology would certainly been very different. However,

conjectural speculations about “what if…” or “what if not…” are, of course, always nothing more than just that; conjectures.

Human Morality: Aims and Abilities

In discussing Kant’s practical or moral anthropology, there are phrases that one cannot avoid returning to over and over again. Besides the general definition of this discipline as the one concerned with our “subjective conditions”, we may also repeatedly have to consider Kant’s dictum that the natural man was neither good, nor evil, for by nature man was not a moral being. (Päd. 9:45)

In light of this, we may conclude that in order to be regarded as a moral being, man needs reason since it is only through this faculty that the concept of a moral law is accessible to us. As should be clear by now, Rousseau agreed in this that without reason and the sense of being a part of a common whole, man would never be anything else than a stupid animal limited by his physical instincts.

Accordingly, no matter how modest or noble the natural man may be, he will still be short of all human characteristics. These characteristics develop only by social interaction, imagination and comparison – three usages of reason that require that we leave the inside-perspectives and take the risk of becoming transformed in a corruptive way. Moreover, although Rousseau regretted the outside-perspective, social man must at least be able to step outside temporarily in order to communicate with his fellows. In this way we elevate ourselves and, as Rousseau said, “To live is not to breathe; it is to act; it is to make use of our organs, our senses, our faculties, of all the parts of ourselves which give us the sentiment of our existence.” (E I 42/253)

Kant’s vocabulary may be less flowery; indeed Kant is known for his methodological way of explaining his carefully mapped out system, a style that also has contributed to the picture of him as somewhat ponderous and dry. We may here recall the common portrayal of him as the “cold Verstandesmensch” who did not seem to regard emotional kindness or friendship as morally better than self-seeking motivations or who did not appear to see any worth of human happiness unless this happiness was deserved by means of virtuous behavior. The purely rational motive of duty and the total abstraction from everything empirical or subjective were the indispensable conditions for a morally worthy action. This ultra-rigid portrait may, as Schilpp argued, be regarded an “unhistorical caricature”. Moreover, as we have seen, it is clear that Kant’s full idea of applied morality requires that we take man’s subjective conditions into account. Yet one cannot escape the fact that Kantian ethics is not a question of being happy but of deserving this happiness. It is a theory that idealizes a rational ideal rather than the existing man.
For Rousseau, the soft nature was at least as important as a rational soul, and for him it did not matter if the desired end was reached by virtue or by loving and peaceful nature. (RIJJ II 154/858) Kant would, of course, not accept something like this. In the light of such apparent disagreements it seems as if Kant, if he ever had a Rousseauian turn, at some point chose to turn away from Rousseau. Still, the continuous presence of Rousseauian themes hints that there might be areas that are not affected by such differences.

Rousseau and Kant

One remaining question is thus what Kant’s so called Rousseauian turn, or revolution, can be said to consist in. From the many discussions concerning the relationship between the thinkers one can at least draw the conclusion that there is no settled consensus here. According to Cassirer, one cannot point out any decisive theme or doctrine in Rousseau that Kant continued. Kant’s reading of Rousseau is rather to be seen as a total turning point that led Kant into a new course that he would never abandon.\textsuperscript{343} Schmucker agrees; Emile and the Social Contract did not merely inspire Kant but provided him with the matter that determined his entire forthcoming course of thinking.\textsuperscript{344} For Jeffery Edwards, Rousseau’s writings triggered a radical change in Kant’s political thinking by making Kant interested in the notion of popular sovereignty, which brought about Kant’s own idea of ethical law-giving through the idea of the general will. Yet, Edwards also stresses that it was this "counter-revolutionary" foundation that required Kant to adopt an “irredeemably egoistic moral psychology.”\textsuperscript{345} According to Beck, Rousseau’s most important impact on Kant is moral and consists in the adoption of the idea of autonomy.\textsuperscript{346} Also Reich sees an ethical impact since Rousseau made Kant see the problems of blindly believing in culture and this led him to a new idea of the value of man.\textsuperscript{347} Wood, as we have seen, holds that Kant’s theory of the human being is identical to Rousseau’s which thus suggests an anthropological impact. Finally we can recall the claims of Velkley that, in line with those of Cassirer, suggest that Rousseau was the single most important cause of a complete revolution that, in turn, was the condition of the critical revolution that led to Kant’s final account of reason and morality. The list could, of course, be made much longer.

Yet, while no one doubts that Kant read Rousseau, there are commentators who want to downplay the actual philosophical impact of this reading.

\textsuperscript{343} Cassirer (1945) p. 1.
\textsuperscript{344} Schmucker p. 143.
\textsuperscript{345} Edwards p. 425.
\textsuperscript{347} Reich p. 95.
Alfred Hegler sees no decisive imprints from Rousseau on Kant’s idea of psychology and thus, one may conclude, not much effect on Kant’s idea of the thinking being. More recently, in his review of Velkley’s book, Kuehn agrees that Rousseau was important for Kant but he was not the indispensable condition for Kant’s critical turn that Velkley wants to see. Kuehn also draws attention to the fact that almost no references to Rousseau are to be found in the moral works. Also Schwaiger seems to say that Kant’s so called Rousseaubegeisterung at that time is often overstated. A similar hesitation is visible in Peter Gay’s preface to the translation of Cassirer’s pioneering essay where he hints that Cassirer perhaps makes Rousseau more Kantian than the facts warrant. Schilpp, though devoting a considerable part of a chapter to Kant’s familiarity with Rousseau states that there are reasons to conclude that not even at his critical state, was Kant far from an unquestioning disciple of Rousseau but rather a highly critical one who was aware of the inconsistencies and extravagances that appeared as soon as one read the texts rationally.

Nevertheless, even if it may be true, as Schwaiger points out, that it has been “proven with all desired clarity” that Kant’s anthropology has its root in the Wolffian school of empirical psychology, this does not exclude the possibility that Rousseau was the one that excited the his interest of human nature. All the evidence of Kant’s intellectual love-affair with Rousseau combined with the lack of references to him in the moral texts, take us back to the query that has not been directly discussed but that nevertheless has figured in the background: namely, whether Kant owed more to Rousseau than the general story tells, and if perhaps there are Rousseauian roots all the way up to his most metaphysical critical system? Secondary literature is, as we have seen, full of interpretations of this relationship, and it is fairly clear that both Kant and Rousseau are benefited by this connection since it discloses the philosophical complexity that exists in both of them. However, the question of whether Rousseau really was a proto-Kantian or whether Kant’s discovery of the Citizen of Geneva resulted in a Rousseauian “turn” or “revolution” or just planted some seeds can be debated endlessly, so it is fortunate that this is not the decisive question here.

To verify the suggestion that Kant’s entire course of thinking is reliant on, and necessarily conditioned by Rousseau is and ought not to be an end in itself. The preceding text rather aimed at demonstrating the complexity of

---

348 Hegler p. 306.
350 Schwaiger p. 25, p. 74 n. As for the term “Rousseaubegeisterung”, see editorial note to AA 29:677.
352 Schilpp p. 51.
353 See Schwaiger p. 102 “Zwar ist schon seit langem mit allem wünschenswerten Deutlichkeit erwiesen, dass Kants Anthropologie in der empirischen Psychologie der Wolffschule wurzelt.”

246
Kant’s practical philosophy and particularly his moral anthropology. I have tried to show that although this division is titled the “second part” of morals, one must not conclude that it is thereby “secondary” in the sense of being inferior. Not even the most formal ethics can exist without a clear view of the empirical constitution of the human being who will always be a member of the world of sense. Given Kant’s arguments, it is only such imperfect beings that can be moral in the strict sense of the word since respect for the moral law only occurs if we realize the insignificance of subjective wants in relation to the command of duty. If we did not have inclinations affecting us, we would be holy and our wills would, like the divine will, always be in accordance with duty.

We have reached a new conception of human nature, distinct from the uncultivated and uncivilized stage from which man once begun. In this, the two philosophers converge; they both saw the need for a better understanding of the human being in its various forms. Rousseau saw himself as a “portrayer of nature and the historian of the human heart” or “an historian of human nature”. Again we may recall that Kant, who, according to Cassirer, was the only one who understood Rousseau, saw that Rousseau was the one who saw “beneath the multiplicity of forms human beings have taken on their deeply buried nature.” (BB 20:58-9) Kant’s intellectual career may be seen as involving turns but still, in the depths of it, one can see an ongoing thread; the interest in human nature accompanied by a Rousseauian heritage.

**Throwing the Mask…**

In order to end this expose I would like to say a few words about optimism. At the title page of Emile we can read:

> We are sick with evils that can be cured; and nature, having brought us forth sound, itself helps us if we want to be improved.  

Seneca’s words reappear in Kant’s Religion just before he felt forced to admit our fallenness. (Rel. 6:20) Because of all the evil that nature parades before us, Pope’s maxim that whatever is, is right must, as many satirists have declared, be refuted as nonsensical. The cause of the fall should be known by now but no matter what we have become, or what we have done

---

354 See RJJJ I 52/726 and III 214/935. Admittedly, the latter statement is the “Frenchman’s” words but Rousseau does not deny it. (The RJJJ, or Dialogues is one of Rousseau’s self-biographical works that is presented as a fictive dialogue between Rousseau and a unspecified Frenchman in which they discuss (judge and defend) the well-known spirit Jean-Jacques, his life and writings).

355 “Sanabilibus aegrotamus malis; ipsaque nos in rectum genitos natura, si emendari velimus, iuvat”. From Seneca’s On Anger (de ira) II:13.
to the world and ourselves, it might still be correct to claim, without thereby lapsing into naïve optimism, that “whatever was, was right.” It also seems as if everything may become good again if we really want to recover.

This recovery through nature, however, requires that we recall what true nature is. In some sense, we have abandoned nature and do no longer express it. What we see are not expressions of human nature but of the mendacious nature of an actor who pretends to be someone or something else than he actually is. Again, we may recall Rousseau’s saying that man is very strong when he is content to be what he is whereas he is very weak when he wants to elevate himself above humanity. (E II 81/305) This move exposes us to the great danger of being disclosed as the limited weaklings that we actually are. In order to become esteemed, we put on the mask without realizing that we, thereby, only become even more alienated from our true nature. The eventual revelation will be brutal. As Rousseau’s namesake Jean-Baptiste Rousseau famously proclaimed:

“the mask falls; the man remains, and the hero disappears.”

Kant seems to be in full agreement with this, or at least he admits that the human species is not particularly lovable if one really comes to know it. (MS 6:402) The present man is no hero but an evil and imperfect creature. Yet it is just such knowledge of our imperfection and subjective conditions that make possible our entering into our deeply buried nature. This inside perspective alone makes knowledge of true human nature possible.

This said, I guess that we may conclude that both Kant and Rousseau held that human beings have to make themselves into what they ought to be. Man does not have his character from birth even though several slumbering characteristics may be present already in the rudest state. Although Rousseau begins Émile by complaining about man’s obsession to modify and abolish nature, he admits that given the present state, everything would be worse without refinement, and therefore, “like the trees in the garden” we need to be moderated into a better shape. Still we must never forget where the truth is to be found.

Accordingly: “Let us return to ourselves, my young friend! - Let us examine, all personal interest aside, where our inclinations lead us” (E IV 287/595-6) or, as Kant says in the Remarks: Quod petis in te est nec te qua-siveris extra – what you desire is in you; do not look for yourself outside. (BB 20:5)

---

Bibliography

Kant’s Works

Unless otherwise stated, quotations from Kant are from the nearly complete edition of Kant’s texts *The Cambridge Edition of the Works of Immanuel Kant* (general editors P. Guyer & A. Wood. Cambridge: Cambridge University Press 1995- ). The list below features the works, translators and the location in the volumes of this edition as well as in the Akademieausgabe (AA). In cases where no translation has been available, translations are my own.


(1765) *Nachricht von der Einrichtung seiner Vorlesungen in dem Winterhalbenjahre* – Immanuel Kant’s Announcement of the Programme of his Lectures for the Winter Semester 1765-1766, trans. D. Walford in *Theoretical Philosophy 1755-1770* AA 2


(1783) Prolegomena zu einer jeden künftigen Metaphysik, die als Wissenschaft wird auftreten können – Prolegomena to any Future Metaphysic that will be able to come Forward as Science, trans. G. Hatfield in Theoretical Philosophy after 1781 (2010) AA 4


(1786) Mutmaßlicher Anfang der Menschengeschichte - Conjectural Beginning of Human History, trans. A. Wood in Anthropology, History and Education AA 8


(1790) Über eine Entdeckung, nach der alle neue Kritik der reinen Vernunft durch eine ältere entbehrlich gemacht werden soll – On a Discovery whereby any New Critique of Pure Reason is Made Superfluous by an Older One, trans. H. Allison in Theoretical Philosophy after 1781 (2010) AA 8


(1793) Welches sind die wirklichen Fortschritte, die die Metaphysik seit Leibnitzens und Wolff's Zeiten in Deutschland gemacht hat? – What real Progress has Metaphysics made in Germany since the time of Leibniz and Wolff?, trans. P. Heath in Theoretical Philosophy after 1781 (2010) AA 20


(1796) Von einem neuerdings erhobenen vornehmen Ton in der Philosophie – On a Recently Prominent Tone of Superiority in Philosophy, trans. P. Heath in Theoretical Philosophy after 1781 (2010) AA 8

(1798) Anthropologie in pragmatischer Hinsicht – Anthropology from a Pragmatic Point of View, trans. R.B. Louden in Anthropology History and Education (2007) AA 7
— Opus Postumum (selections), trans. E. Förster & M. Rosen AA 21-22
— Briefwechsel – Correspondence, trans. A. Zweig (2007) AA 10-13

Rousseau’s Works
Quotes from Rousseau’s work are taken from the edition The Collected Writings of Rousseau ed. G. D. Masters and C. Kelly. Hanover, N.H.: University of New England Press 13 vol. (1990-2010) Abbreviated CW below. References are also given to their location in the Oeuvres Complètes (OC) Emile is cited according to Allan Bloom’s translation from 1978. A revised version of this translation is also available in CW vol. 13.

(1752) Préface de Narcisse ou l'Amant de lui-même – Preface to Narcissus or the Lover of Himself, trans. C. Kelly in CW vol. 10 / OC vol. 2
(1755) “Letter to Voltaire Sept. 10 1755” in CW 3 / OC 3
(1756) “Letter to Voltaire Aug. 18 1756” in CW 3 / OC 4
(1758) Lettre à M. D'Alembert sur les spectacles – Letter to d'Alembert on Theatre, trans. A. Bloom in CW vol. 10 / OC vol. 5
(1757-8) Lettres morales – Moral Letters, trans. C. Kelly in CW vol. 12 / OC vol. 4

251
(1761) *Julie ou la Nouvelle Héloïse* – *Julie or the New Heloise*, trans. P. Stewart & J. Vaché in *CW* vol. 6/*OC* vol. 2
(1762) *Manuscrit Favre* – *Favre Manuscript* (*Emile*, first Draft), trans. C. Kelly in *CW* vol. 13/*OC* vol. 4
(1769) *Les Confessions* – *Confessions* (Published posthumously), trans. C. Kelly in *CW* vol. 5/*OC* vol. 1
(1776-8) *Les Rêveries du promeneur solitaire* – *Reveries of the Solitary Walker* (Published posthumously), trans. C. E. Butterworth in *CW* vol. 8/*OC* vol. 1

**Other Primary Sources**

d’Alembert, J. le Rond (1757) “Geneva” (Article from the *Encyclopedie*) trans. A. Bloom. in *Rousseau, Collected Writings* vol. 10 pp. 239-49
Diderot, D. (1755) “Natural Right” (“Droit naturel” Article from the *Encyclopedie*) in *Rousseau, Collected Writings* vol. III pp. 135-9
Herder (1784) *Ideen zur Philosophie der Geschichte der Menschheit [mit Kants Rezensionen]*. Berlin: Deutsche Bibliothek (1914)
Mandeville, B (1732) *The Fable of the Bees* 2 volumes (ed. by F. B Kaye). Indianapolis: Liberty Fund


Voltaire (1755) “Letter to Rousseau Aug. 30 1755” in *Rousseau, Collected Writings* vol. 3 pp. 102-4


**Secondary Sources**

While some of the titles included in this list are not directly cited in the text, they have all been of great value for this study by providing both general and detailed knowledge about Kant, Rousseau, human nature, history and morality. While far from complete, the list at least aims to show the multitude of scholarly texts on these subjects and may be seen as suggestions for further reading.


— (1920) “Rousseau and Conscience” *The Journal of Philosophy, Psychology and Scientific Methods* 17:7 pp. 186-191 (reply to Schinz’s review, se below)


254


Dieterich, K. (1878) *Kant und Rousseau*. Tübingen: Verlag der H. Laupp’schen Buchhandlung


255


— (2010b) “Recent Works on Kantian Maxims II” *Philosophy Compass* 5/3 pp. 228–39


Gurwitsch, G. (1922) “Kant und Fichte als Rousseau-Interpreten” *Kant-Studien* 27 pp. 138-78


Hegler, A (1891) *Die Psychologie in Kant* Freiburg: Akademische Verlagshandlung von J. C. B. Mohr


— (1899) “Rousseaus Einfluss auf die definitive Form der Kantischen Ethik” Kant-Studien 2 pp. 11-21
— (1921) Den Nyere Filosofins Historie. Kjøbenhavn: Gyldendalske Boghandel
Konhardt, K. “Was man sein muss um ein Mensch zu sein: Kant und die neuerlich wiederentdeckte Frage nach dem guten leben” in H. Busche & A. Schmitt (eds.) Kant als Bezugs punkt Philosophischen Denkens. Würzburg: Königshausen und Neumann
Menn, M. (1894) Kant’s Stellung zu Jean-Jacques Rousseau Freiburg: Wane’rsche Bucherei (Dissertation)

258
Schlick, M. (1930) Fragen der Ethik Wien: Julius Springer
Xie, S. (2009) “What Is Kant: a Compatibilist or an Incompatibilist? A new Interpretation of Kant’s Solution to the Free Will Problem” Kant-Studien 100 pp. 53-76
Dictionaries and Encyclopaedias

Eisler, R. (1904) *Wörterbuch der philosophischen Begriffe*. Berlin: E.S. Mittler und Sohn
Kirchner, F. & Michaelis, C. (1890) *Wörterbuch der Philosophischen Grundbegriffe*. Heidelberg: Weiss
Mayer, J. B. (1837) *Synonymisches Handwörterbuch der deutschen Sprache*. Kempten: Dannheimer