



**A University of Sussex PhD thesis**

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

# **AGENT PARTICULARISM: THE ETHICS OF HUMAN DIGNITY**

André Lúcio Santos de Almeida

Philosophy PhD

University of Sussex

September 2018

## **Declaration**

I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

**Signature** .....

**UNIVERSITY OF SUSSEX**  
**ANDRE LUCIO SANTOS DE ALMEIDA**  
**PHILOSOPHY PHD**  
**AGENT PARTICULARISM: THE ETHICS OF HUMAN DIGNITY**  
**SUMMARY**

The thesis proposes an ethics centred on the notion of human dignity. In Chapter One I introduce the position the thesis proposes, Agent Particularism, according to which who you are is relevant to determining what you ought to do. I reject the thesis of the universalizability of moral judgements that says that if you judge that X is the right thing for you to do, you are necessarily committed to the view that X is the right thing for everybody to do in relevantly similar circumstances.

In Chapter Two I present an Agent-Particularist conception of freedom. I offer an Agent-Particularist conception of the self. I make a distinction between negative freedom, which is being free from external interference, and positive freedom, which is developing into the ideal version of yourself (in accord with your particular nature).

In Chapter Three I present Agent Particularism as a kind of virtue ethics. I offer a solution to an epistemological problem that the thesis faces: once I have rejected the existence of exceptionless moral principles, how can there be moral knowledge and what kind of knowledge that would be? I argue that the problem can be solved by understanding moral knowledge as consisting on the deliverances of a perceptual

capacity. I position Agent Particularism in relation to traditional virtue ethics.

In Chapter Four I present the Agent-Particularist conception of human dignity. I show that the Agent-Particularist position developed in the first three chapters issues in a peculiar conception of human dignity. I present the basic elements of an Agent-Particularist conception of dignity. I present Kant's conception of dignity and contrast it with the Agent-Particularist conception.

## Table of Contents

<b>Acknowledgements .....</b>	<b>vii</b>
-------------------------------	------------

<b>Introduction.....</b>	<b>1</b>
--------------------------	----------

<b>Chapter 1: Agent Particularism and the Universalizability of Moral Judgements.....</b>	<b>9</b>
---	----------

1. Situation Particularism.....	9
2. Universalizability of Moral Judgments .....	12
3. Universalizability as a Requirement of Consistency.....	15
4. Moral Supervenience .....	19
5. Consistency Without Truth?.....	27
6. Legitimate Morally Relevant Differences Between People.....	29
7. Ways of Denying Universalizability .....	33
8. David Wiggins' "Winchian" Rejection of Universalizability .....	41
9. The Problem With Winch's and Wiggins' Rejections of Universalizability .....	48
10. Genuine Morally Relevant Differences Between People .....	54

<b>Chapter 2: An Agent-Particularist Conception of Freedom ....</b>	<b>66</b>
---	-----------

1.1. Rousseau's Richer Conception of a Person .....	66
1.2. Berlin's Two Concepts of Liberty .....	70
1.3. An Agent-Particularist Negative Freedom .....	84
1.3.1. The First Level of Negative Freedom.....	85
1.3.2. The Second Level of Negative Freedom.....	93
1.3.3. The Threat Coming From the Blank Conception's Morality ....	98
1.3.3.1. An Agent-Particularist Critique of Scanlon's Contractarianism .....	100
1.3.3.2. An Agent-Particularist Critique of Kant's Account of Freedom.....	107
1.3.4. A Natural Development of Agent Particularism: .....	131

2. An Agent-Particularist Account of Positive Freedom: .....	143
2.1. Negative and Positive Freedom: .....	144
2.2. The Process of Positive Freedom: .....	147
2.2.1. Strength of Character: .....	152
2.2.2. Developing as Fully as Possible in Accord with the Potential of One's Particular Genius: .....	155
2.2.2.1. Realizing One's Potential: .....	155
2.2.2.2. Developing One's Uniqueness: .....	158
 <b>Chapter 3: Agent Particularism and Virtue Ethics.....</b>	<b>164</b>
1. Introduction: .....	164
2. Agent Particularism as a Kind of Virtue Ethics: .....	166
2.1. Agent Particularism and Uncodifiability: .....	167
2.2. Virtue as Sensitivity to the Demands of Situations: .....	181
3. The Traditional Virtue Ethicist as a Universalizability Theorist:...	187
 <b>Chapter 4: An Agent-Particularist Conception of Human Dignity .....</b>	<b>192</b>
1. Introduction.....	192
2. Some Key Elements of an Agent-Particularist Conception of Human Dignity.....	193
3. Kant's Conception of Human Dignity .....	210
3.1. Problems with Kant's Conception of Dignity and the Contrast with the Agent-Particularist Conception.....	227
4. The Essence of the Agent-Particularist Conception of Human Dignity.....	239
 <b>Bibliography.....</b>	<b>242</b>

## **Acknowledgments**

This thesis is greatly indebted to Professor Michael Morris, my PhD supervisor. In general I believe it is safe to say that Michael is a very knowledgeable philosopher who is sincerely committed to his supervisees and to their projects. In my particular case I can affirm without a shadow of a doubt that this thesis would never have happened if it weren't for Michael's support. In particular, I would never have finished this PhD if it weren't for Michael's consistent and generous support when things became really rough. I am profoundly grateful for that.

I am also grateful to FDC and to CAPES, institutions that have financially supported my studies. Their support was essential to giving me favourable conditions to focus on the development of this work. In particular I am deeply grateful to Wagner Veloso for always believing in my potential and for his enthusiastic support to my project of becoming a philosopher. Without Wagner this thesis wouldn't be more than a hope and a dream. I am equally grateful to Subramanian Rangan, who over the course of the years became a true mentor for me. If I took the initiative of undertaking a PhD at all it was due to Subi's generous and humane support and encouragement. Subi instigated, guided and enlightened me since our first conversation in Brazil, out of nothing but his genuine love for humanity.

My most profound gratitude and unconditional love go to my father Emerson, my mother Nilda, my brother Ado, and my sister Ana Carolina. They gave me my life and nurtured it. From my first moves as a baby until this day they have always been there for me, loving,



encouraging, supporting and teaching. This thesis and all my future works are dedicated to them.

It is only fair to finish by thanking myself. Besides all the determination and hard work, I am particularly grateful for never betraying my deepest philosophical intuitions and my vision of a decent and dignifying life for each and every human being.

## **Introduction**

This thesis does in particular two things. Firstly it offers a particular conception of human dignity that comes out of an ethics based on a substantial view of what it is to be a particular person. This conception of dignity is in itself a contribution to the understanding of this notion, providing an alternative to the prevailing view of human dignity. Secondly it gives substance to a kind of theoretical position in ethics which, though not previously proposed (as far as I am aware), was already there as it were waiting to be formulated.

One of the things which this thesis does is develop a substantive view of what is essential to a particular person, which is traceable to Jean-Jacques Rousseau but articulated more fully here. I call this the richer-genius conception of the self. One of the novelties of the thesis is that I develop a new theoretical position which arises naturally out of that conception of the self. This is one of the original contributions of this thesis. It involves the presentation of a distinct kind of moral particularism, which is an alternative to the traditional rejection of moral universalism. As I will explain below the traditional rejection of moral universalism insists that variations in the surrounding circumstances and contingent facts about agents are morally relevant, and on that basis rejects universalism. The position I propose acknowledges not only that variations in circumstances and contingent facts about agents are morally relevant, but also that variations in the natures of the agents themselves are morally relevant. In fact, it is on this latter point that I focus on this thesis. Since the varied features of individual agents which I take to be morally relevant are objective rather than subjective the position turns out to be new and in some sense radical.

Another novelty presented by this thesis is the implicit understanding of authenticity that it offers. I don't explicitly contrast it much with the understanding of other authors in the actual thesis, but there are places (like parts of Chapters Two and Four) in which I reject for instance the Kantian conception of different notions and offer my own in its place. This is part of rejecting a more traditional conception of the self (adopted for instance by Kant) and allowing space for what I call the richer-genius conception of the self, from which my particular understanding of authenticity can be derived. Rousseau is the basis of my conception of authenticity, and he is also the source of other philosophers' conceptions. In fact, I arrived at Rousseau through philosophers like Charles Taylor and Alessandro Ferrara, who use Rousseau almost exclusively as the source of their conceptions of authenticity. But the conception that can be derived from this thesis is original, and substantively distinct from theirs. This novel conception of authenticity is presented implicitly rather than explicitly here because my focus was in other issues. If this thesis were longer I could spend some time addressing authenticity directly and explicitly. But since it is not I have saved that part of the project for a future occasion.

Another novelty of the thesis is its presentation of the contrast between negative and positive freedom. Building on the work of philosophers like Rousseau, John Stuart Mill and Wilhelm Von Humboldt I draw out a distinct way of thinking of this contrast which is quite different from the one that for example Isaiah Berlin encourages. This new conception also arises out of the richer-genius conception of the self.

Finally, as a development from the points mentioned above the thesis also offers a novel conception of human dignity. Much of the approach to the notion of dignity that we have has been Kantian and the thesis offers a plausible alternative to that approach that has real substance to it. This contribution becomes necessary as the Kantian approach in

a sense misses the whole point of a conception of dignity, as it precisely fails to dignify the individual human being.

I offer here an ethics centred on the notion of human dignity. The claim is that to respect the dignity of a person is to let them be who they really are rather than impose anything on them or subjugate them to anything other than themselves. What grounds the demand to respect the individual is the general moral demand for people to develop in the direction of the ideal version of themselves, as it is in that way that they are able to develop into a moral being.

The thesis starts in Chapter One from the idea that who the agent is is relevant to determining what they ought to do. This means that variations in what is essential to a person being the person they are are morally relevant. I call this position Agent Particularism. I introduce Agent Particularism as a rejection of the thesis of the Universalizability of Moral Judgements. The universalizability claim is that if a person judges that X is the right thing for them to do in a certain situation they are committed (rationally) to the view that X is also the right thing for everybody else to do in relevantly similar circumstances.

I distinguish between Agent Particularism and the more traditional kind of rejection of universalism, which I call Situation Particularism (commonly known as moral particularism), in order to make clear the kind of position I am not taking. While the focus of Situation Particularism is on the moral relevance of variations in the surrounding circumstances and contingent facts about agents, the focus of Agent Particularism is on the moral relevance of variations in the individual natures of agents.

I explain in the chapter that the universalizability theorist insists that it is universalizability that brings consistency to morality. Anything different from the idea that everybody ought to act in precisely the same way in relevantly similar circumstances is seen as arbitrarily

making an exception in one's own case. But there are two distinct aspects to universalizability: one is consistency, and the other is the insistence that who you are doesn't matter. I show that we can keep consistency while rejecting the view that who the agent is is irrelevant in moral terms. I explain that all that is actually required for consistency is that we have a coherent application of the thesis of the supervenience of the moral on the non-moral. The supervenience thesis means that there can't be a moral difference in a situation without a non-moral difference.

I show that the real concern behind insisting on the universalizability thesis comes from the belief that universalizability is an indispensable prerequisite if we are to avoid what the universalizability theorist refers to as the evils of arbitrariness. Their view is that either we accept the universalizability principle or we fall into a position in which anything goes. I show that though they want universalizability in order to have consistency and avoid arbitrariness, by holding a position such as Agent Particularism we come to see that there is no real problem. Supervenience works perfectly well within the Agent-Particularist position. There is also no problem of consistency if one holds such a view. All it takes is that whenever there are two relevantly similar people in relevantly similar circumstances we judge them similarly. We can have that while holding that who you are makes a difference.

I show in Chapter One that there is a reason why people think that unless there is universalizability there can't be consistency and we are left with arbitrariness. The universalizability theorist has a minimal conception of personal identity. Their view is that effectively a person is just something like a blank in a situation. They don't think of a person as something with a whole character which is essential to being them. I argue that it is not the case that a person is a kind of a blank in a situation and that the only relevant variation is variation between situations. Universalizability requires an insufficiently robust

conception of personal constitution. To be a certain person is a much richer thing, and Agent Particularism acknowledges that.

Agent Particularism rejects the blank conception of the self. So I start Chapter Two by presenting a richer understanding of what it is to be a particular person, which I call the richer-genius conception of the self. For that I rely on Jean-Jacques Rousseau's substantial conception of what is already natural to a person. In particular I build upon his notion of a person's particular genius. The view is that each person's particular genius, which makes them the person they are, is unique and that there is real substance to what is unique to them.

I show that once we are working with a richer-genius conception of the self a distinctive kind of account of freedom becomes necessary. The Agent Particularist conception of freedom is composed of negative freedom and positive freedom. In order to give a contrast to the Agent Particularist conception I look at Isaiah Berlin's account of freedom and point out problems with his view. I then move to developing the Agent Particularist conception of negative freedom. Negative freedom becomes relevant for Agent Particularism because life in society poses a threat to a person's particular genius through pressures towards uniformity of thought and behaviour.

I show how on the Agent-Particularist conception the need for negative freedom is justified by positive freedom. In order to be a plausible rejection of universalizability, and so avoid arbitrariness, Agent Particularism requires the notion of the ideal development of a person. In order to rule out the possibility of the agent avoiding their duties when for example what is required is difficult for them to do, there must be a demand for them to develop into the ideal version of themselves – which is the Agent-Particularist understanding of positive freedom.

In order to develop the Agent-Particularist conception of negative freedom I rely on Rousseau and John Stuart Mill. I then present and criticize two conceptions of ethics that work with the blank conception of the self: Thomas Scanlon's and Immanuel Kant's. Finally I move into the Agent Particularist conception of positive freedom. I explain that the process of positive freedom requires virtues of character that individual acts do not require. I present those virtues of character by relying on the views of Rousseau, Mill, Wilhelm Von Humboldt, and Ralph Waldo Emerson.

In Chapter Three I explicitly present Agent Particularism as a kind of virtue ethics. I show the differences and similarities between a traditional kind of virtue ethics and Agent-Particularist virtue ethics. The notion of the ideal development of the person is essential to both views, but while the traditional virtue ethicist assumes that everyone should develop towards the same ideal Agent Particularism claims that the ideal development of each person is peculiar to them and accords with their particular individual genius.

I come back to the distinction between Agent Particularism and Situation Particularism (which I have introduced in Chapter One). Situation Particularism is understood in this thesis as the rejection of what I call Situation Universalism, which is the view according to which the business of ethics is to come up with and apply universal principles of behaviour. The key distinction between Agent Particularism and Situation Particularism is that the Situation Particularist may accept the universalizability principle and in so doing reject Agent Particularism.

I explain that the traditional virtue ethicist is a Situation Particularist. They hold this view because they reject the thesis of codifiability – the view that ethics can be properly captured in a single set of universal principles that guide one's moral actions. I explain that my particular kind of Agent Particularism also is at the same time Situation

Particularist. I present David Wiggins' and John McDowell's rejections of codifiability and show that they become particularly compelling when developed in an Agent-Particularist way. I explain the particular character that an Agent-Particularist rejection of codifiability takes – it becomes the defence of negative freedom.

I rely on McDowell in order to solve an epistemological problem that I am faced with once I reject codifiability: if moral knowledge isn't a matter of conceiving and applying universal principles, how can there be moral knowledge and what kind of knowledge is it? McDowell's view is that moral knowledge consists in the deliverances of a perceptual capacity, which is what virtue is. I finish the chapter by explaining that the traditional virtue ethicist is in the end a universalizability theorist.

In Chapter Four I explicitly present the Agent-Particularist conception of human dignity. The ideas developed in the first three chapters of the thesis are what is necessary to explaining what is really striking about pursuing the idea of a richer-genius conception of the self properly. I show in Chapter Four that taking that idea seriously gives one a very distinctive conception of what human dignity is and why humanity has a call upon us, providing us a way of giving substance to the idea of an ethics of human dignity.

I explain that an account of human dignity is focused around what deserves respect in people, and that the view that I offer in this thesis is precisely that each person is worthy of respect. I offer a conception of respect for each person in which the key idea is in letting each person be who they are.

I explain that Agent Particularism acknowledges that there is a limit to how much someone can be criticized. In essence the structural feature of the position is that there is no further moral criticism that we can make of the person once they have developed themselves into the ideal version of themselves. I also explain that for Agent Particularism there



isn't a way in which someone could be compared to another person and found less good because they are a different person. Agent Particularism does not require people to be other than they are. It requires them to reform themselves to become the ideal version of themselves but it does not require them to be different people.

I then present Kant's conception of dignity in order to use it as a contrast with the Agent-Particularist conception. I explain that Kant's view is that what deserves respect in people is their humanity, which is understood as a characteristic or set of characteristics of human beings. In the end the view is that the dignity of the person is grounded on their rational capacity for autonomous action. I point out what is problematic about Kant's conception of dignity and make a contrast of those points with Agent Particularism. I then show how some of the central notions in Kant's conception of dignity (which are quite useful) actually work better if developed in an Agent-Particularist way. I finish the chapter and the thesis by presenting the essence of an Agent-Particularist conception of human dignity.

## **Agent Particularism and the Universalizability of Moral Judgements**

### 1. Situation Particularism:

I argue for a particular rejection of a particular kind of universalism. But it is different from the one that most people have in mind. The standard rejection comes from a position in ethics commonly known as moral particularism. I will call this position "Situation Particularism". I will call "Agent Particularism" the different position proposed by me.

John McDowell introduces universalism, the position that Situation Particularism aims at rejecting, in the following way:

"It may seem that the very idea of a moral outlook makes room for, and requires, the existence of moral theory, conceived as a discipline that seeks to formulate acceptable principles of conduct. (...) On this view, the primary topic of ethics is the concept of right conduct, and the nature and justification of principles of behaviour" (McDowell 202, p. 50).

Jonathan Dancy puts it in more direct terms:

"The idea is that, if we are doing our moral thinking properly, we approach a new case with a set of principles, and that we look to see which of those principles the case falls under" (Dancy 2006, p. 3).

Situation Particularism aims at rejecting the view according to which "... the very possibility of moral thought and judgement depends on the provision of a suitable supply of moral principles" (Dancy 2006, p. 73). Universalists may advocate a set of principles or one single universal principle that would be applied every time. In any case they will start from the basic assumption that doing the right thing is a matter of

conducting oneself in accordance with the moral principle that is appropriate for the occasion. And what Situation Particularism denies is precisely the existence of true exception-less moral principles.

For Situation Particularism what is the right thing for someone to do depends on the circumstances, because variations in the surrounding circumstances and in contingent facts about the agent's history might make a moral difference. Situation particularists believe that the difference in circumstances which it will depend upon can't be codified. That basically means that it is not possible to fix in one or more principles all that may be a morally relevant variation on circumstances.

There are in contemporary moral philosophy two main roots of the Situation Particularist position. They have two distinct lines of justification for the claim above, known as the uncodifiability claim. One line of justification comes from McDowell. Following Aristotle, McDowell holds that "... the best generalizations about how one should behave hold only for the most part" (McDowell 2002, p. 58). For that reason it is not possible to have one or more moral principles on which we could rely in every circumstance with no exceptions. But a system of morality that risks not being there for us when we need it the most is hard to defend. If one actually tries to guide one's life by such a system,

"... cases would inevitably turn up in which a mechanical application of the rules would strike one as wrong ... because ... one's mind on the matter was not susceptible of capture in any universal formula" (McDowell 2002, p. 58).

So for the Situation Particularist the application of principles in certain cases would lead to a mistake. And not because the agent is applying his principles wrongly, but rather because "... practical matters are mutable, or lacking in fixity" (Nussbaum quoted in Crisp 2000, p. 27). Which makes it the case that "... the complexity and unpredictability of human affairs are such that circumstances will arise in which the rules

are inappropriate for determining what we should do" (Crisp 2000, p. 27).

The other main root of the Situation Particularist position comes from Dancy, who offers a distinct line of justification for the claim about uncodifiability. He believes that "... the behaviour of a reason in a new case cannot be predicted from its behaviour elsewhere" (Dancy 1993, p. 56). The thought is that there are things that form general features, but these are features that can't be collected together to form general laws. Dancy calls his view holism of reasons. He explains it: "... there are many different ways in which the presence or absence of a consideration can make a difference to how one should respond..." (Dancy 2006, p. 73). That is because "... a feature that is a reason in favour of action in one case may be no reason at all in another, or even a reason against" (Dancy 2006, p. 190). Thus, in Dancy's view the same general feature might sometimes count in favour of doing something and sometimes count against doing it, and there is no way of adding qualifications that determine when it counts in favour and when it counts against.

So if the universalist suggests for instance that producing pleasure makes every act that does that a better act, Dancy's kind of Situation Particularism would answer that:

"... while the property of producing pleasure makes an act better in some circumstances, this property makes an act worse in other circumstances. That an act would give pleasure to the sadist is not merely an overridden positive feature of the act. Rather, sadistic pleasure actually makes the act morally worse than it would be if it didn't afford sadistic pleasure. As Dancy comments, the pleasure people get from watching hangings makes letting them watch morally worse" (Hooker 2000, p. 7).

I won't pursue any connections with Dancy's arguments in this thesis. I also won't get into the details of McDowell's arguments in this chapter. In Chapter Three I will come back to this issue, and look in detail at

McDowell's arguments for Situation Particularism. I will then show how every argument that McDowell uses for particularism of his sort applies even more so once you add in Agent Particularism to it. The importance of me presenting the Situation Particularist's rejection of a form of universalism here is to make clear what kind of approach I am not taking. My focus is not on how variations on the surrounding circumstances and in contingent facts about the agent's history might make a moral difference. I am concerned with a different issue, where it is not so much a question of the circumstances but rather of the person. I will explore the possibility that variations in persons' individual natures may be morally relevant. That is, who a person is may make a difference to what they ought to do. And that is an issue with which the Situation Particularist doesn't generally engage. The Situation Particularist generally assumes that who a person is isn't relevant – all that is relevant are variations in the surrounding circumstances.

## 2. Universalizability of Moral Judgments:

The simplest way to understand Agent Particularism is to see it as a way of rejecting a form of the thesis of the Universalizability of Moral Judgements. This thesis has been formulated in different manners. But R. M. Hare, perhaps the most obvious proponent of the thesis, is clear about what is for him its most important aspect: "... all the universalist is committed to in making a moral judgement is to saying that *if* there is another person in a similar situation, then the same judgement must be made about his case" (Hare 1963, pp. 48-49).

According to this formulation the universalizability thesis is thus the view that which person you are makes no difference to what is right

for you to do. I will from now on take that to be the key feature of what I will call the universalizability thesis. The view entails that if I judge that the right thing for A to do in a situation is X, I necessarily commit myself to the view that X is the right thing for everybody else to do in relevantly similar circumstances. In this view moral judgements are universalizable in the sense that they apply to everybody, because the "... particular identity [of the agent] is completely irrelevant in the determination of the correctness or appropriateness of the judgment" (Jollimore 2011).

Some more general formulations of the universalizability thesis have been offered, and they don't explicitly involve the key feature of what I am calling the universalizability thesis – the commitment that who the agent is makes no difference in moral terms. J. L. Mackie for instance, at a certain point states that:

"Anyone who says, meaning it, that a certain action (or person, or state of affairs, etc.) is morally right or wrong, good or bad, ought or ought not to be done (or imitated, or pursued, etc.) is thereby committed to taking the same view about any other relevantly similar action" (Mackie 1977, p. 83).

The differences in those two kinds of formulations of the universalizability thesis are by no means irrelevant. And that is because in holding certain philosophical positions one can accept some sense of the latter formulation at the same time that they reject the former. That is to say, one can without being inconsistent accept the view that reasons apply to morality while one denies the view that who the agent is is irrelevant in moral terms.

Hare also has offered a more general formulation: "... by calling a judgement universalizable I mean only that it logically commits the speaker to a similar judgement about anything which is either exactly like the subject of the original judgement or like it in the relevant respects" (Hare 1963, p. 139). The thing about a formulation of the

above kind is that there are different senses of phrases such as “like the subject of the original judgement”. In effect there is a sense of the above formulation that takes into account not only the situation but also the agent as the ‘relevant respects’ to be accounted for in the judgement.

The point becomes clearer when we look at yet another kind of formulation of the thesis offered by the universalizability theorist, which explicitly takes into account the agent as a relevant aspect of moral judgements. Henry Sidgwick has offered the following formulation of universalizability: “... whatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances” (Sidgwick 1962, p. 379). Putted like this, the thesis becomes a necessary requirement if one wants to see morality as involving reasons. But note that one can accept the view that reasons apply to morality, in fact one can hold the view that morality is such as being susceptible to plain truth (and as requiring consistency for that reason) while one agrees with a formulation like this. And all this can be done at the same time that one rejects our initial formulation of the thesis and so rejects the view that who the agent is is irrelevant in our making of moral judgements.

The issue becomes even more interesting when we see that the same author offers all these different formulations of the universalizability thesis. Hare has also formulated the thesis in this third version:

“It follows from universalizability that if I now say that I ought to do a certain thing to a certain person, I am committed to the view that the very same thing ought to be done to me, were I in exactly his situation, including having the same personal characteristics and in particular the same motivational states” (Hare 1981, p. 108).

What is most problematic here is that universalizability theorists such as Hare not only offer these different kinds of formulations, but they

also treat the differences between them as non-existent or irrelevant. That is particularly problematic because an important implication of the differences in the kinds of formulation, as we have seen, is that there are different commitments involved in each one of them. Philosophers who hold certain positions might be perfectly comfortable in accepting a formulation - which brings consistency to morality - of the form 'if A and B are alike in all relevant respects we must judge them alike' but would reject right away universalizability as I understand it.

### 3. Universalizability as a Requirement of Consistency:

Many of us believe it is important to ensure that when we judge an action to be morally right and another to be morally wrong there are reasons (and good ones) for our thinking that one is right and the other is wrong. To accept this view is to accept a requirement of consistency in morality, in such a way that when we have two cases in which all the morally relevant reasons apply in the same way we must judge them alike. In fact as described by John Renford Bambrough such demand for consistency, if accepted, is not a peculiarity of moral thought:

"... the requirement of consistency which all reasoning, in all fields, must satisfy. Whenever I make two different judgments, of whatever kind, I must always be prepared to discharge the onus of showing that there are differences between the cases that justify the difference in my judgment of them" (Bambrough quoted in Ward 1973, pp. 289-90).

Proponents of the universalizability thesis believe that universalizability plays precisely the role of safeguarding the consistency of moral judgments, as Andrew Ward claims: "... the thesis of universalisability in morality is nothing more than the demand that a person should be consistent in making his moral judgments" (Ward 1973, p. 289). The



universalizability theorist's point with claims of this kind is to suggest that universalizability is a kind of obvious and necessary principle if one is to keep consistency in morality: "... in the thesis of universalisability one is merely stressing the rationality of moral judgments, in the sense that one is demanding that they be consistent" (Ward 1973, pp. 289-90).

But the universalizability theorist is not content with claiming that universalizability is a requirement of consistency. They also insist that universalizability exclusively is what can provide such consistency: "It is ... because moral judgements are universalizable that we can speak of moral thought as rational" (Hare 1963, p. 6). Thus the universalizability theorist believes that universalizability is a feature that is intrinsic to morality. It is of the very logic of moral terms that when they are being properly used the judgements in which they are contained will be universalizable.

But note how in the latter quote Hare uses the more generic kind of formulation of the thesis. And although this kind of formulation doesn't in itself involve the commitment that who you are is irrelevant in moral terms, it has all it takes for ensuring consistency. Don Locke helps us to clarify the point: "The claim is not that if I ought to do X then there is some reason why anyone ought to do X, but that if I ought to do X then there is some reason why I ought to do X, (...) such that anyone else to whom that reason applies ought also to do X (logical consistency)" (Locke 1968, p. 32). In the above generic form of universalizability offered by Hare, 'different judgments on cases identical in their non-moral properties' would indeed be a sign of lack of consistency, but only if it takes into account the case as a whole – including as relevant not only features of the situation but also what kind of agent the agent is.

We need here to clearly distinguish between two things. One is what is required for consistency, and the other is the point also insisted by universalizability theorists according to which who you are doesn't matter. They present universalizability as a requirement of consistency. But all that is required for consistency is that relevantly similar cases should be judged in the same way. It is not a requirement of consistency that it should make no difference who you are. That is a separate point. In effect it seems that even universalizability theorists such as Hare, in the middle of focusing the discussion about universalizability on the issue of consistency, seem to be aware of that fact:

"If I say 'I ought, but there is someone else in exactly the same circumstances, doing it to someone who is just like the person I should be doing it to, but he ought not to do it,' then logical eyebrows will be raised; it is logically inconsistent to say, of two exactly similar people in exactly similar situations, that the first ought to do something and the second ought not" (Hare 1989, p. 179).

One point that the above quote acknowledges is that in order to accept a consistency requirement for morality one doesn't have to also accept that it applies independently of who you are. We can draw an analogy between the requirement of consistency in this more modest version of the universalizability principle and the way in which rules are supposed to function. The "rule" coming out of the universalizability principle must be applied, for the sake of consistency, whenever you have 'two exactly similar people in exactly similar situations'. Now that is very different from saying that the rule must be applied whenever you have any two people in exactly similar situations. Let's rely on Isaiah Berlin to further clarify the analogy: "In so far as rules are general instructions to act or refrain from acting in certain ways, in specified circumstances, enjoined upon persons of a specified kind, they enjoin uniform behaviour in identical cases" (Berlin quoted in Macintyre 1957, p. 327).

The view that universalizability is intrinsic to morality is debatable. Obviously those of us who reject the universalizability thesis deny it. We can clarify the issue by investigating whether the requirement of consistency is truly all that universalizability amounts to. Whether or not all that universalizability is is a kind of requirement of consistency is controversial and depends on the kind of formulation of the thesis that one endorses. But what does become clear by the way in which the universalizability theorist treats the different formulations of the thesis is that this is all it is supposed to be.

Consistency in morality is something that many (including me) want to hold on to. But what precisely gives substance to such consistency? What is truly required of moral judgements in order for them to respond to the demand for consistency properly?

The following quote from Hare can give us a hint in terms of which direction to go in trying to answer these questions:

"If a person says that a thing is red, he is committed to the view that anything which was like it in the relevant respects would likewise be red. (...) 'This is red' entails 'Everything like this in the relevant respects is red' simply because to say that something is red while denying that some other thing which resembles it in the relevant respects is red is to misuse the word 'red', and this is because 'red' is a descriptive term..." (Hare 1963, p.11).

Hare's thought is that if someone says that a thing is red, it is in virtue of its possession of a certain property, i.e. its redness. Therefore if a person says that a thing is red, he is committed to the view that anything that is like it in the relevant respects (in the present case, its redness) would likewise be red. Thus for Hare there is universalizability about redness, which entails that: 'everything like this red thing in the relevant respects would likewise be red'. And he thinks that such universalizability about redness is a straightforward requirement of the fact that red is a descriptive term, which is another way of saying that

there are truths about redness. This view seems reasonable. Let's now work out carefully what is going on here.

#### 4. Moral Supervenience:

As a principle of consistency, all that is required is that relevantly similar cases be judged similarly. As it is usually understood the universalizability thesis goes beyond that bland statement in two ways. One is in terms of the view that there couldn't be a moral difference between two situations without there also being a non-moral difference. In other words, "moral properties do not vary quite independently of non-moral properties, but are in some sense consequential or supervenient on them" (Hare 1963, p. 19). The non-moral difference of two cases is precisely what grounds the moral difference between them, and that is known as the principle of the supervenience of the moral on the non-moral. The other way is saying that who you are doesn't matter. Now, it is perfectly possible to accept the supervenience commitment without accepting the other one. For that reason these need to be separated.

It looks as if the only thing that is truly required for consistency in morality is acceptance of the principle of moral supervenience. Nonetheless the universalizability theorist claims that all that universalizability amounts to is consistency. Let's see if a closer look at supervenience helps to clarify the issue. As we have seen above there are truths about redness, which makes it the case that if something can correctly be called red it is in virtue of it possessing the property redness. It is reasonable to then accept the claim that in virtue of 'red' being a descriptive term some form of universalizability about 'red' holds, in the sense that anything that also possesses the same property

redness can be properly called red. That is where the consistency of judgments about red lies.

When it comes to moral judgments the consistency applies in a similar way. What grounds the consistency of moral judgments in Hare's view are certain non-moral properties on which the moral status of the thing in question supervenes. Hare gives a helpful example (Hare 1961, pp. 80-81). Let's imagine that we are discussing whether a certain picture X hanging on the wall is a good picture (in the sense of a good work of art). Let's imagine further that there is another picture Y hanging next to X. One is the replica of the other but we don't know which is which. They were painted by the same artist around the same time. Hare points out that in a case like this we can't say that the two pictures are precisely like one another in all respects but one, that one picture is good and the other is not. There must be some further difference between them that makes one being good and the other not. Bringing the issue to moral valuation, "... by calling [any actions] 'duties' we only mean that they have, *in addition*, certain non-ethical predicates" (Moore 1948, p. 181). Such a view has important implications. One of them is that if two actions share all their non-moral properties they will also share all their moral properties. At the same time an object that possesses a moral property can't cease to have that property (or have it for a lesser degree) without a change in its non-moral properties.

But there is a worrying aspect about Hare's account of supervenience, presented by Matthew Kramer. He synthesises Hare's formulation of moral supervenience in the following way:

"Necessarily, if  $r$ , then there is a valid inference of the ' $p, q$ , so  $r$ ' form, the two premises of which hold" (Kramer 2005, p. 194).

What Kramer thinks is that effectively we can isolate the features which determine which is the right or the wrong action. According to this view

anything that can be seen as right, or wrong, or good (and so on) can be so seen:

"... in virtue of being subsumable under a universal principle ... which provides that anything endowed with certain non-moral properties is also endowed with rightness or wrongness or some other specified moral property" (Kramer 2005, p. 194).

Kramer takes supervenience to require a kind of universalism, an anti-particularism of some form. In fact, it is not uncommon among universalizability theorists to phrase their view on moral supervenience along the following lines: 'If one's reason for saying that an action is good is that it has qualities ABC they are committed to the view that any action that has qualities ABC is also good'. But that is a problematic view. Jonathan Dancy points out why this can't be right:

"... the second action might have some further quality D which defeats any tendency that ABC had to make the action good. Thus if my reasons for calling this girl good are that she is chaste and pious, I am not committed to calling another girl good if, for example, besides being chaste and pious she is also cruel" (Dancy 1981, p. 377).

According to this refined view on supervenience, moral properties supervene not on all the non-moral properties, but only on those of them that are morally relevant to the judgment. Now, it is not the case that 'anything endowed with certain non-moral properties is also endowed with rightness or wrongness'. What Kramer says effectively is that what supervenience commits you to is that "if 'r' (which is, say, 'it is right to do this') then there must be a 'p' and a 'q' from which 'r' follows". But Dancy shows that we never have that. All we ever have is "if 'p' and 'q' in the circumstances, then 'r'. Since we can never fully specify the circumstances, it is perfectly plausible for one to accept supervenience even if one denies moral universalism.

After clarifying this point, let us now look at Hare's consistency formulation of universalizability:

"it is logically inconsistent to say, of two exactly similar people in exactly similar situations, that the first ought to do something and the second ought not" (Hare 1989, p. 179).

All Hare is saying here is that similar cases must be judged similarly. That much needs to be accepted if there is to be consistency involved in morality at all. For Hare judgments about morality aren't themselves true or false because they are imperatives. But he thinks they also have a descriptive component and that is what brings in supervenience. So he thinks that moral judgments are things which have a descriptive component and a command attached to them. And effectively what it supervenes on gives you the descriptive component. The thing about moral judgments in a view like this is that you have these two components but the structure of it isn't evident. So what you need is some form of access to the purely descriptive aspect, which is provided by supervenience.

My view on the other hand is that moral judgments are the kind of thing that can in effect be true or false. According to this view supervenience comes in from a requirement of consistency. But in that case a question remains to be answered. Why there need to be supervenience? In other words, if moral judgments are just true and false, why can't they be barely true (or false)? Why is it that one have to think that moral judgments are true (or false) in virtue of something else?

Here is a possible position which makes perfect sense. Let's think about the colour case, introduced by Hare through the example of red. Hare himself doesn't consider the possibility that redness supervenes on other things but surely in fact it does. Why does one think that red supervenes on something? Well, as we know sight is authoritative

about colour. But why is that so? Because sight is the sense which is affected by light and colour is a matter of reflects of light. So changes of colour will always have to do with changes in reflects of light. And that explains supervenience in the case of judgments of red. I believe we can explain supervenience in moral judgments through an analogy that holds between judgments of colour and moral judgments.

The colour case works like this: we have the argument of supervenience to explain why it is that sight is authoritative. We also have views about when someone is in a good position to make a moral judgment in the situation. Effectively that is a matter of what kinds of factors they are aware of. Now, our eyes are affected in a certain way by light. And in the moral case when the agent is faced with a situation and perceives what needs to be done he is affected in a certain way by the non-moral features of the situation. So someone is in a good position to make judgments about colour when they can see and the light is good. And that is because colour supervenes on light reflections. In the same way someone is in a good position to make a moral judgment when they are aware of the relevant facts of the situation. And that is because those are the things that moral judgments supervene on. So changes of moral judgments will always have to do with changes in the relevant non-moral features of the situation. It places a parallel argument here.

The role of supervenience here will vary according to one's conception of moral judgments. As it happens mine is a strong realist line. But there is a sense in which it makes no difference in fact. Because you can accept consistency, you can accept supervenience, but still disagree with universalizability. And as we have seen, according to the universalizability theorists themselves all that universalizability is supposed to do is give us a requirement of consistency. And where there is supervenience consistency requires an application supervenient on a certain level, which is fine. But the problem is that



there are two aspects to universalizability as I understand it. One is the bland consistency and then there is also the insistence that who you are doesn't matter, and that is a separate issue. As pointed out by Dancy:

"... it is desirable at this stage to make some effort to distinguish between the supervenience of moral properties on the non-moral and the universalisability of moral judgements as propounded by Professor Hare. Narrowing the subvenient class in the way suggested simply collapses the two notions into one. But there would be room on the alternative proposal to reject Hare's doctrine of universalisability without rejecting the doctrine of moral supervenience, and this is an option which ought to be left open as long as possible" (Dancy 1981, p. 374).

I believe that the option mentioned on the quote above – the alternative proposal of rejecting Hare's doctrine of universalizability without rejecting the doctrine of moral supervenience - works perfectly well if one accepts that there are truths in morality. Looking from that angle, we could say that what Agent Particularism does is precisely give substance to that option. There is no problem of consistency in this position. There is also no failure of supervenience. Yet the position still acknowledges that who you are makes a difference. As Peter Winch points out:

"... considerations of consistency, intelligibility, and rationality do apply in moral matters; (...) But to insist that, to speak intelligibly, he must also be prepared to say: 'And other people too, if they are to judge rightly, must make the same judgements as I have made concerning these situations', is to make a much more sweeping claim and one which seems to me highly questionable" (Winch 1965, p. 199).

I have said above that the Agent-Particularist position I am presenting is a strongly realist one. I would like to clarify what I mean by realism and also by objectivity (which is a notion that will become relevant for us in this chapter and others). According to Michael Dummett, realism is:

"... the belief that statements of the disputed class possess an objective truth-value, independently of our means of knowing it: they are true or false in virtue of a reality existing independently of us" (Dummett 1978, p. 146).

On my view realism is basically defined in terms of the independence of something from thought and representation. When I say that mine is a strong realist view what I mean is that it is very realist, i.e. there is no dependence at all on thought. Later in this chapter I will introduce the idea of a demand on people to develop according to their ideal process of development. A realist view of one's proper development will be one according to which how one should develop is not dependent on how one (or anybody else) thinks or wants to develop.

It is helpful for us to make a contrast between realism and objectivity, but for that we need to distinguish between two understandings of the notion of objectivity. I will follow Michael Morris in referring to them as an epistemic conception of objectivity and a metaphysical conception of objectivity. According to him:

"Something is *epistemically objective* if and only if knowledge of it does not depend on any particular mode of access" (Morris 2015, p. 916).

In this conception, for something to be objective is for it to be accessible from any number of points of view. This is what philosophers have in mind for example when they say that colours are not objective whereas physical properties are. The idea is that the physical properties of something can be found out in any number of ways. But colours are not objective in the same way, because authoritative judgement of something's colour depends upon sight.

Morris defines metaphysical objectivity in the following way:

"Something is *metaphysically objective* if and only if its nature is wholly independent of any way of knowing it or thinking about it" (Morris 2015, p. 916).

Metaphysical objectivity is basically the same as being real on the realist view I have presented above – essentially, it means independence from thought and representation. This is the notion of objectivity that I will use in this thesis. For instance, I will claim that there are objective features of a person that make them be who they are. What I mean by that is that there are certain features of a person that make them the person they are, and that is the case independently of them or anybody else for instance acknowledging which these features are.

The thing is that these two notions of objectivity don't go together. One is an epistemic notion and the other is not. And a similar thing happens with the notion of subjectivity. One understanding of 'subjective' is that it is not epistemically objective; that is the case for instance if you can only get at something through experience. On the other understanding 'subjective' means that there is no truth beyond what people think – which is an anti-realist view.

This anticipates a point which I will continue to develop in Chapter Four. As I mentioned above, I will claim that there are objective features of each one of us that makes us the person each one is. The view is that each person is particular, and at least part of that particularity is essential to the person. But I want to clarify that even though what is essential to a person is agent-relative it is not subjective in any serious sense because it doesn't depend upon what anybody thinks or feels or anything like that. This issue also anticipates a point which I will start to make in the next section. The universalizability theorist has a concern that if you don't have universalizability anything goes. With

this thought they want to insist that the moment in which we make obligations agent-relative anything goes. But the whole point of this chapter is that that is just not true – we can make obligations agent-relative and maintain realism without universalizability.

## 5. Consistency Without Truth?

Hare thinks that consistency about judgments of redness follows from the fact that 'red' is a descriptive term, since this kind of term is susceptible to truth. So, he wants to say, it is a requirement which follows from redness being subject to questions of truth that the universalizability thesis in cases of red in some form holds.

As we started to see earlier, if one holds the view that there is truth in morality Hare's example of judgments about redness function as a nice analogy for moral judgments. As it happens in the case of 'red', if one sees moral terms as descriptive the requirement of consistency in moral judgments follows from that fact. All that would be required is that one judges similar cases similarly. The consistency of moral judgments would follow from moral judgments being subject to questions of truth, and such consistency would then be explained in terms of supervenience.

But for Hare, "... the differences between them [descriptive judgments and moral judgments] in other respects are ... sufficient to make it misleading to say that moral judgements are descriptive" (Hare 1963, p. 10). So there must be for Hare something special about morality. For he is committed to there being a difference in the cases of judgments of redness and moral judgments - in the case of redness consistency just follows in fact from the descriptivism of the term. The question is how and why it is different in the case of moral judgments?

Hare believes that "... though prescriptive, [moral judgments] have a descriptive element in their meaning" (Hare 1963, p. 5). He doesn't think there can be truths about morality because he thinks moral judgments are imperatives, and imperatives can't be true or false (though he sees them as answerable to reason). Yet he thinks that there are nevertheless requirements of consistency in morality, which he then explains in terms of universalizability. So there is an important question for Hare to answer: Why is there in his view a requirement of consistency once he claims that there are no truths in morality?

It seems that the universalizability theorist's motivation for insisting in consistence in morality even if they hold that there is no objective moral reality is in their view that morality is nothing but a system that we invented to help guide society in a good direction. In Mackie's words: "... the function of morality is to counter the effects of limited sympathies..." (Mackie 1977, p. 152).

The view held by Mackie is common among many philosophers who think that although there is no truth in morality, it is important that there is a connexion between morality and what you can argue about. The idea, held by the Contractualist kind of philosopher, is that you must be able to present arguments for people who differ from you. Thomas Scanlon, for example, thinks that morality is determined as those principles that anybody rational could be brought to accept without being forced. He sees moral judgments as: "... judgments about what would be permitted by principles that could not reasonably be rejected, by people who were moved to find principles for the general regulation of behaviour that others, similarly motivated, could not reasonably reject" (Scanlon 1998, p. 4)<sup>1</sup>. Such principles would be like good rules for the running of society. But it has to be the kind of rules that everybody could agree we should appeal to in order to decide

---

<sup>1</sup> I will explain in detail Scanlon's position on Chapter Two.

any complicated issues. The whole point is to fix morality to a domain that is unproblematic, over which there will be no problem of agreement. And supervenience is a kind of structural way of relating the moral judgments to non-moral judgments that are relatively speaking unproblematic. Now because morality is just a system for guiding society in a good way, and its function is to be argumentatively persuasive there has to be the principle of universalizability – or so they claim. So the status of universalizability there is a kind of side constraint on what morality is.

In effect we have two sources of the universalizability principle. One is just a principle of consistency, which follows from a requirement of truth. You have truth as an antecedent commitment and you can't have truth without consistency. The alternative view is that there is no truth but the point of morality is to be something that people can address in argument. The important thing to keep in mind is that, as we can see, there are various reasons why one might want to believe in consistency. In the same way there are various reasons why one might believe in supervenience, and the different kinds of supervenience one can subscribe to. One can accept all such things as consistency, supervenience, and rationality, but still disagree with universalizability in the worrying form. The question then is, how is this possible?

#### 6. Legitimate Morally Relevant Differences Between People:

One thing that we are doing in this chapter is “clearing the terrain”, in terms of carefully distinguishing acceptance of consistency and acceptance of supervenience from any other commitments involved in the universalizability thesis as I understand it. That is important because such a distinction remains blurred throughout the tradition of

this debate. Let's look at the following formulation of the universalizability thesis by Sidgwick:

"... whatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances. Or, as we may otherwise put it, if a kind of conduct that is right (or wrong) for me is not right (or wrong) for someone else, it must be on the ground of some difference between the two cases, other than the fact that I and he are different persons" (Sidgwick 1962, p. 379).

In the first claim of the quote above Sidgwick basically introduces the principle of consistency by presenting things in terms of "all similar persons in similar circumstances". Whereas in the second claim he seems to introduce the worrying form of universalizability, by stating that the difference between two cases must be something "other than the fact that I and he are different persons". There are two importantly distinct possible senses for the second claim made by Sidgwick. One is highly problematic in that it contradicts the first claim with the claim – insisted by universalizability – that who you are doesn't matter. The second sense is unproblematic and helpful in that it clarifies the point that though one may acknowledge that different people may be demanded to act in different ways in relevantly similar circumstances, there needs to be substance to the differences on people that ground their different demands. In Mackie's words:

"It may be that what is wrong for you is right for me; but if it is, this can only be because there is some qualitative difference, some difference of kind, between you and me or between your situation and mine which can be held to be, in the actual context, morally relevant. What is wrong for you cannot be right for me merely because I am I and you are you, or because I am John Mackie and you are, say, Richard Roe" (Mackie 1977, pp. 83-84).

On the quote above we can see once more how what the universalizability theorist is really concerned with is consistency. We can also see that Mackie is ready to accept that "what is wrong for you is right for me", as long as what grounds the difference is "some

qualitative difference, some difference in kind, between you and me". Otherwise it mustn't count as a morally relevant difference. But there is something interesting going on here. For Mackie (as for most of the other universalizability theorists) one is making the same point when one says that 'similar people in similar circumstances must be judged similarly', and when one says that 'in similar circumstances everybody ought to do the same thing'.

While keeping that in mind let's ask ourselves, why in the end Mackie says that 'What is wrong for you cannot be right for me merely because I am I and you are you, or because I am John Mackie and you are, say, Richard Roe'? What the rhetoric of that is suggesting is that if it depends on who you are that is like depending on what your name is. So effectively the universalizability theorist has a kind of minimal conception of personal identity. Their view is that effectively a person is just a kind of blank in a situation. They don't think of a person as something with a whole character being what is essential to being them. Mackie imagines that none of the relevant differences between situations attach to the person. So he has a minimal conception of what it is to be that person. It is just to be the bearer of that name. Basically Mackie is saying that if we were to take into consideration the differences between people, it can't be merely because I am André Almeida and you are Richard Roe. But since there is nothing else that is essential to people, then differences between people don't matter in moral terms.

In the following quote Sidgwick makes a point quite similar to the one made by Mackie:

"... no one will deny that there may be differences in the circumstances — and even in the natures — of two individuals, A and B, which would make it wrong for A to treat B in the way in which it is right for B to treat A. In short the self-evident principle [the Golden Rule<sup>2</sup>] strictly stated must take some such

---

<sup>2</sup> Sidgwick's formulation of the Golden Rule is: 'Do to others as you would have them do to you'.



negative form as this; 'it cannot be right for A to treat B in a manner in which it would be wrong for B to treat A, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment...' (Sidgwick 1962, p. 380)

Here Sidgwick not only acknowledges that differences on the natures of individuals may be morally relevant, but goes even further in saying that "no one will deny" that. But for him things like "they are two different individuals" don't count as a morally relevant difference. Sidgwick basically treats the requirement of consistency and the universalizability thesis as if they were the same. And it appears that this happens because like Mackie he has a minimal conception of personal identity.

As we have seen, for authors like Hare, Mackie and Sidgwick consistency in morality doesn't follow from a requirement of truth. And they want to use universalizability as a way of enforcing it:

"... an agent whose judgments are universalizable will be morally consistent, in the sense that she will judge her own actions by the same standards she applies to others. Such an agent will not make an exception of herself by allowing herself to break a rule she regards as binding for others..." (Jollimore 2011).

What the universalizability theorist seems not to see is that though they acknowledge that all that consistency requires is that similar cases be judged similarly (what includes the thought that certain substantive features of the agent are in fact morally relevant) they also end up insisting that taking the agent into account is a matter of 'making an exception in one's own case' without noticing that this involves a substantial extra commitment. And they arrive at this result out of their "... concern about the evils of arbitrariness" (Kramer 2005, pp. 173-4).

What is puzzling is that though the universalizability theorist holds universalizability in order to have consistency and 'avoid the evils of

arbitrariness', if one holds a position such as Agent Particularism one comes to see that there is no real problem here. There is not a problem of lack of consistency, there is no failure of supervenience, there is no lack of objectivity, or anything like that. All those things apply perfectly well while holding that it does matter who you are.

Some philosophers have already hinted at the point that there is no real problem at stake, but in a different sense. In Alasdair Macintyre's words:

Where there is real moral perplexity it is often in a highly complex situation, and sometimes a situation so complex that the question "What ought I to do?" can only be translated trivially into "What ought someone like me to do in this kind of situation?" (...) where a situation is too complex, phrases like "someone like me" or "this kind of situation" become vacuous. For I am the only person sufficiently "like me" to be morally relevant and no situation could be sufficiently like "this kind of situation" without being precisely this situation" (Macintyre 1957, p. 335).

Ironically universalizability theorists such as Mackie went on that same road, "... in practice no two cases will ever be exactly alike" (Mackie 1977, p. 83). And the same can be said for Hare:

"Since we cannot know everything about another actual person's concrete situation (including how it strikes him, which may make all the difference), it is nearly always presumptuous to suppose that another person's situation is exactly like one we have ourselves been in, or even like it in the relevant particulars" (Hare 1963, p. 49).

## 7. Ways of Denying Universalizability:

Different kinds of rejection of the universalizability principle have been offered. Some of them want to hold onto some form of consistency

while others give up consistency in morality<sup>3</sup>. In this thesis I will only look at rejections of universalizability that attempt at holding consistency as these are the ones that can offer a helpful contrast to the kind of rejection of universalizability that I am proposing.

Not only the universalizability theorist is concerned about what they call the 'evils of arbitrariness'. Some philosophers that explicitly reject universalizability also have that concern. Therefore they agree with Mackie and Sidgwick in thinking that there needs to be substance to what grounds differences in judgments of what two people ought to do in relevantly similar situations. Peter Winch is one such philosopher. Winch holds that who you are matters, at the same time that he wants to distance his position from any form of arbitrariness in morality.

Winch points out that from the point of view of the universalizability theorist:

"It will be said that if I do not admit that the right thing for him to do would be the right thing for anyone to do in the same circumstances, I am ruling out any possible distinction between what a man thinks he ought to do and what he in fact ought to do. And if that is so, how can it matter to a man what he thinks is right; since whatever he thinks is right will be right" (Winch 1965, p. 209).

For the universalizability theorist either you hold that who you are doesn't matter or the result is arbitrariness in morality. There is no chance of consistency in between. But Winch wants to show that this view is a consequence of a misunderstanding of positions like the one he introduces. He explains that his position is misunderstood for "... a version of the Protagorean 'Man is the measure of all things', according to which if A asserts p and B asserts not-p, we cannot ask who is right, but only say that p is true-for-A and not-p is true-for-B" (Winch 1965,

---

<sup>3</sup> A well-known position that rejects universalizability at the same time that gives up consistency is the one offered by Alasdair Macintyre. Macintyre's rejection is focused around the idea that each person has their own private morality. In his view the agent's only commitment should be with what they believe is the right thing for them to do.

p.209). This quote presents an extreme form of subjectivism in morality which Winch wants to clearly distance himself from. In this kind of view whatever reasons you present for having acted in a certain way must be accepted as a good reason as long as you are being honest, in such a way that the notion of good reasons in a sense loses its meaning. According to him when two people have different moral judgments about relevantly similar circumstances:

"... it may be that neither what each says, nor anything entailed by what each says, contradicts anything said or implied by the other. But this certainly does not mean that, if A believes that X is the right thing for him to do, then X is made the right thing for A to do by the mere fact that he thinks it is" (Winch 1965, p. 209).

Winch wants to hold a form of Agent Particularism, in which who you are is morally relevant, at the same time as holding onto consistency. In Winch's words, his view is that "... if A says 'X is the right thing for me to do' and if B, in a situation not relevantly different, says 'X is the wrong thing for me to do', it can be that both are correct" (Winch 1965, p. 209). Winch therefore rejects the universalizability thesis. He leaves no doubt that the kind of view his position is meant to reject is the one that holds that:

"... ethics is a sort of calculus of action, in which actions are considered as events merely contingently attached to particular agents. What determines the rightness or wrongness of an action is the situation in which it is to be performed. (...) as in the case of mathematical propositions, the question who is making this judgement is of no logical interest. That is what I wish to call in question" (Winch 1965, p. 198).

In order to make his case, Winch relies on the story from the book "Billy Budd, Sailor" by Melville. Winch summarizes the case in the following way:

"I shall consider the moral dilemma of Captain "Starry" Vere, R.N., captain of

H.M.S. Indomitable, on active service against the French in the period immediately following the Nore mutiny, when further mutinous outbreaks aboard H.M. ships were feared at any time: a situation 'demanding,' Melville writes, 'two qualities not readily interfusible - prudence and rigour.' Billy Budd, a foretopman of angelic character, is impressed into service on the Indomitable from the merchantman Rights of Man on the high seas. He is persecuted by the satanic master-at-arms of the Indomitable, Claggart, in a campaign which culminates in Claggart's falsely accusing Billy, before Vere, of inciting the crew to mutiny. In the stress of this situation, Budd is afflicted with a speech-impediment which prevents him from answering the charge. Frustrated, he strikes Claggart, who falls, strikes his head and dies" (Winch 1965, p. 200).

Captain Vere presents the case to those involved in the improvised trial as a moral dilemma. On the one hand there are the demands of 'natural justice' and 'conscience in favour of acquitting Billy Budd. On the other hand there are the demands coming from the military code (from the eyes of which Billy Budd kills a superior officer on duty, which constitutes a capital crime whose penalty is death). Winch further stresses that:

"... Vere, while he sees the military code as opposed to certain of the demands of morality, does not see it as to be contrasted with morality sans phrase, but as something to which he himself is morally committed. For him the conflict with which he is faced is an internal moral one ..." (Winch 1965, p. 201).

Winch's approach is to restrict his discussion to only one class of moral judgments – that involving conflicts within morality. By looking at "... a conflict between two genuinely moral oughts ..." he avoids the possibility of "... the temptation to which we are all prone of making an exception in our own case, for no good moral reason, in order to evade a distasteful duty" (Winch 1965, p. 203). Here we see the first important difference between Winch's kind of Agent Particularism and the one proposed by the present thesis. While Winch's position takes into account only one class of moral judgments, my position encompasses moral judgments in general.

Winch points out that though the universalizability theorist claims that

"[i]f a person says 'I ought to act in a certain way, but nobody else ought to act in that way in relevantly similar circumstances', then ... he is implicitly contradicting himself" (Hare 1963, p. 32), such claim doesn't help them in establishing universalizability. And that is because "I am not claiming that there are no cases in which anyone would be morally justified in going on like this. I am denying only that, in all cases, a man who refuses to accept such a corollary is thereby misusing the word 'ought'" (Winch 1965, p. 206). Thus in order to deny universalizability it is enough for Winch to demonstrate in one class of moral judgments that who the agent is does matter.

In the Billy Budd example Vere ends up by concluding that the demand from military law is too strong for the demand from 'natural justice' to resist it. Vere then solves his moral dilemma by condemning Billy Budd to the death penalty. Winch puts himself in Vere's position, of a committed military officer responsible for the ship and facing the danger of a new mutiny. And from that place having to make a decision about Billy Budd's case. He then concludes that:

"Having done this, I believe that I could not have acted as did Vere; and by the 'could not', I do not mean 'should not have had the nerve to, but that I should have found it morally impossible to condemn a man 'innocent before God' under such circumstances. (...) It is just that I think I should find the considerations connected with Billy Budd's peculiar innocence too powerful to be overridden by the appeal to military duty" (Winch 1965, p. 208).

Contrary to Vere, for Winch in relevantly similar circumstances the demand from what Vere called 'natural justice' would have made for him impossible to condemn Billy Budd. Not because he is being weak-willed, but rather because it would be wrong for him to do it. Winch continues:

"According to Sidgwick, and those who think like him, this must mean that I think Vere acted wrongly, made the wrong decision. However, I do not think this. The story seems to me to show that Vere did what was, for him, the right thing to do" (Winch 1965, p. 208).

So in effect Winch thinks that in those relevantly similar circumstances the right thing for Vere to do differs from the right thing for Winch to do. Winch says as we have seen that it would be morally impossible for him to act like Vere did. But what exactly does he mean by that? In order to answer that question we need to take a few steps back in Winch's argument. The following claim is central to his moral views:

"As Melville suggests, it may well happen that when I am confronted with an actual situation demanding a delicate moral decision from me, I find that things strike me rather differently from the way they struck me when I was thinking only generally, or as a spectator" (Winch 1965, p. 198).

Winch claims that there is a primacy of the agent's perspective in moral judgments, i.e. in some sense the agent's judgment has more legitimacy than that of any spectator in terms of determining what is right for him to do. And the reason is that – according to this view – I may imagine myself as best as I can as facing a certain situation and having to decide what to do, and conclude that 'X' is the right thing for me to do. But when I am in fact faced with the situation I find that things strike me rather differently. It is only in being actually confronted with the situation that you are truly in a good position to make the moral judgment. Winch then makes the following claim:

"A man in a situation like Vere's has to decide between two courses of action; but he is not merely concerned to decide to do something, but also to find out what is the right thing for him to do. (...) the deciding what to do is, in a situation like this, itself a sort of finding out what is the right thing to do" (Winch 1965, pp. 209-210).

According to Winch one thing is for the agent to merely decide which course of action to take. A different thing is to perceive what the situation is requiring of him. Winch's point is that the deciding what to do is key to finding out what is right to do. In his words: "It is because

I think that deciding is an integral part of what we call finding out what I ought to do that I have emphasized the position of the agent in all this" (Winch 1965, pp. 209-210).

So for Winch it is when you are actually in a position of having to decide what to do that you are able to find out what is the right thing for you to do. And that is because in having to decide, the situation strikes you in a certain special way. Winch continues:

"It seems to me that what one finds out is something about oneself, rather than anything one can speak of as holding universally. (...) What a man finds out about himself is something that can be expressed only in terms of the moral ideas by consideration of which he arrives at his decision" (Winch 1965, p. 212).

For Winch in finding out what is the right thing for him to do the agent is finding out something about himself. We can see at this point the full-bodied subjectivism of his account. First he distances himself from an objective sense of finding out something about himself by making it clear that he "... certainly do not wish to endorse any 'self-realization' theory of morality" (Winch 1965, p. 212). Then he introduces his subjective understanding of finding out something about himself, having as a key element 'the moral ideas by consideration of which he arrives at his decision'.

In order to properly understand Winch's rejection of universalizability, it is essential to clarify what he means by 'the moral ideas' that the agent considers when he needs to decide what to do. Since Winch is clear in pointing out that what the agent finds out about himself isn't anything objective, and since what the agent finds out are 'moral ideas', we are left with exploring the possibility of a strong kind of subjectivism as what grounds the different moral demands that distinct people may have in relevantly similar situations. Coming back to the Billy Budd case, Winch believes that the right thing for him to do would be to



judge Billy Budd as innocent, "... [b]ut somebody else in such a situation, considering those very same arguments, might conclude that the moral possibilities were different for him..." (Winch 1965, p. 213).

What the agent finds out in making the decision is which moral ideas he already holds, and that will inform his decision. That is what Winch means by 'what is morally possible for someone', i.e. roughly what is in line with the moral beliefs and values he already holds. We can conclude that because for the author, as we have seen, how a situation strikes the agent is essential to determining what is the right thing for them to do. And for Winch "... if we want to express, in a given situation, how it strikes the agent, we cannot dispense with his inclination to come to a particular moral decision..." (Winch 1965, p. 213).

According to Winch's view it is those subjective moral beliefs that ground the difference between what is right for Vere to do and what is right for Winch to do in Billy Budd's case. For him those differences in moral beliefs are enough to justify their being required to act in distinct ways in the same relevant circumstances.

In this view how the situation strikes the agent is central in finding out what is the right thing to do. And the same situation may strike different agents differently. That will depend on what kind of moral belief the agent already holds in the background, i.e. his inclination to come to a particular moral decision. Such inclination will determine, in Winch's words, what is morally possible for the particular agent to do. So effectively what Winch's idea here is doing is saying that if when you look at yourself you find that you are already committed to a moral point of view which is inconsistent with acting in a certain way, then you should not act in a way that someone else should. But the problem is that that is just radically subjectivist.

## 8. David Wiggins' "Winchean" Rejection of Universalizability:

David Wiggins is another philosopher that could be properly seen as offering a version of Agent Particularism. Like Winch at the same time that he rejects universalizability he understands the importance of keeping consistency in morality. Wiggins' position can be seen as quite similar to the one offered by Winch, but contains some refinements. In fact, in discussing Winch's position Wiggins goes as far as claiming that "... much that is said here seems right..." (Wiggins 1987, p. 169).

Wiggins pretty much follows Winch's line of thought in accepting the view that if A judges that the right thing for him to do is X and B judges that the right thing for him to do is Y in a relevantly similar situation it may be that both are correct – though he makes a minor amendment to it which we will look at later. Wiggins finds it particularly useful to look at (the problematic, on his view) cases in which I judge that the right thing for me to do is X and you judge that the right thing *for me* to do is Y. He thinks that those cases can be solved:

"... if we can confer a special moral authority on one perspective. The only one that seems privileged is the perspective of the agent or participant. If those outside a situation fail to concur in a moral belief that those within it tend to converge upon, and converge upon seemingly non-accidentally relative to the content of their judgment, have the ones outside fully grasped the situation?" (Wiggins 1987, p. 165).

Wiggins follows Winch in emphasizing the privileged position of the agent. Like Winch he justifies that privileged position with the view that "... it matters in a special way ... how things strike him" (Wiggins 1987, p. 170). And he presents Winch's reasoning that culminates in the claim that what is right for him to do differs from what is right for Vere to do in the Billy Budd case. Wiggins then suggests that if Winch is right about that the explanation would be that:

"... while many ways of being struck by the situation are no doubt excluded as morally misguided or maniacal or simply purblind, what we always may lack reason to assert ... is that there is just one way it must strike him if he is sincere" (Wiggins 1987, pp. 170-71).

Wiggins thus makes the point that for us to judge properly what is the right thing for Vere to do in that situation we need to take into account how the considerations made about the case strike him. And according to him someone else might be struck by those same considerations in a different way.

We see also in Wiggins' position how the 'moral ideas' already held by the agent are key in determining what is right for that agent. For the author "... here we are concerned... with what is morally possible for this or that agent" (Wiggins 1987, p. 168). Depending on the moral ideas already held by the different agents involved in relevantly similar circumstances the situation might strike those different agents differently.

Nonetheless Wiggins believes there is a difference between his position and that of Winch's, as he endorses what he calls a refined view on the primacy of the agent's perspective. According to him: "Much more needs to be said about what confers priority upon the participant perspective, and what sort of priority this is" (Wiggins 1987, p. 165).

Basically Wiggins thinks that Winch's position accepts a conception of the privileged position of the agent that is too permissive. In Wiggins' words: "Even when the agent meets Winch's three requirements ... we surely ought not to treat the agent's finding as above criticism" (Wiggins 1987, p. 172).

For Winch as long as the agent meets three requirements in his judgment – 1) he is not mistaken in seeing himself in a genuinely moral context when he is really not in one; 2) it is not the case that "... his ideas of right and wrong differ so profoundly from our own, that we are

unwilling to accept his claim that he acted rightly" (Winch 1965, p. 210); and 3) he is being sincere – his position should be considered a privileged one in such a way that his judgment ought to be seen as right for him. Wiggins' view is that even if the agent meets those three requirements we are still supposed to look for possible problems with his judgment. And he illustrates the point by coming back to the Billy Budd example. According to Wiggins, Winch grasps well one part of Melville's thinking with the story, in which the agent - the person who is closer to the predicament - is in a position to see better the case. But there is:

"... another equally important trend in Melville's thinking to which Winch pays insufficient attention. (...) It was not quite right for Winch simply to declare 'The story seems to me to show that Vere did what was for him the right thing to do'. Even if it is out of place to lay the whole blame for any of this madness and its doings at the feet of individual men, criticism still has to reach down to their particular practical judgments, and the route by which they were arrived at" (Wiggins 1987, p. 172).

Wiggins thinks that Winch refrained from criticizing Vere's judgment while having no real grounds for doing that. And he wants to prove his point by claiming that:

"... there is one important form of criticism we still need to engage in that prevents us, even on the strength of such thoughts being out of Vere's reach, from simply acquiescing in Vere's decision not to wait till HMS Indomitable rejoined the rest of the fleet, to elide the normal procedures, and to dispense with a hearing of Billy's case at a properly constituted court martial. Of course, Vere's own finding is important, and integral to a morally informed narrative. But so is the disquietude that is evoked by his precipitancy" (Wiggins 1987, p. 173).

The point is that restricting ourselves to Vere's reasoning in the trial may be misleading. We should step back from the trial's circumstances and criticize Vere's very decision to have an improvised trial instead of simply waiting to join the rest of the fleet, when a standard procedure would normally be put in place to deal with the case. But though

Wiggins sees this difference between his position and that of Winch, he thinks that except for this point Winch is right. After objecting to that part of Winch's argument Wiggins claims that:

"In the presence of any case where neither finding is open to the sort of decisive criticism we have found we can mount of Vere's, and where the agent's actual finding is one we can make sense of (if only retrospectively) in the light of what the agent becomes, I am prepared to treat this finding as morally decisive of what was right for him" (Wiggins 1987, p. 181).

Wiggins basically follows Winch's objection to the universalizability principle. He makes an important amendment to it through the claim that though the agent is indeed in a privileged position and even if he is sincere we still need to be able to criticize his judgment. But then the question that emerges is, what are the grounds for criticism?

In order to understand what are Wiggins' grounds for that criticism we need to start by looking at his view of moral judgments, according to which:

"Any judgment of these kinds appeals by virtue of its content to a point of view that is not only subjective but also inter-subjective, not only mine (it is at least mine) but also common to me with others, and to this extent impersonal" (Wiggins 1987, p. 60).

Like Mackie's, Wiggins' view on the source of morality comes from his idea that "[h]uman beings need norms of reciprocity and cooperation that can counteract the settled tendency of things to turn out badly rather than well" (Wiggins 1987, p. 60). Wiggins thinks that morality is something that we came up with in order avoid that things turn out badly between us. Morality is for him a sort of convergence of perspectives that leads to what he calls 'the moral point of view':

"... we expect a point of view that can be shared between the members of an actual society to give expression to a potentially enduring and transmissible shared sensibility. To adopt the moral point of view is to see one's thoughts, feelings and actions as answerable to the findings of such a shared sensibility" (Wiggins 1987, p. 60).

In the quote above we seem to have Wiggins' grounds for the sort of criticism on the agent's judgment that he claims is lacking in Winch's position. So with the Billy Budd case presumably the objection is that if Vere's decision conflicts with the shared sensibility after we lay on it the necessary criticism then it is wrong even if it is sincere. But that does not mean that Wiggins accepts the universalizability principle, i.e. Vere's judgment being wrong doesn't make it the case that there is only one judgment people can legitimately make in those same relevant circumstances. He points out that:

"We have no assurance, not even ordinary assurance that, always, wherever there is a practical judgment that survives criticism at the level of the reflective, self-critical agent who has attained the distance that seems to him (then and later) to be the right distance, all sufficiently informed intelligences that understand his predicament ... will converge on that judgment ... and will converge in a way that leaves everyone who properly explains that convergence with no alternative but to concur in the judgment" (Wiggins 1987, p. 174).

Wiggins thinks that there are occasions in which such convergence will not happen. And the explanation for such lack of convergence is that moral judgments depend on people's subjectivities, people's particular moral sensibility. In his words:

"The reason why there is a special difficulty in conceiving of such assurance is this. Human interests and concerns are as indefinitely various and heterogeneous as are human predicaments. Even moral interests and concerns are indefinitely various and heterogeneous" (Wiggins 1987, p. 174).

Because of such diversity, in cases in which the sort of criticism that he makes about Vere's conduct can't be found, if the agent's judgment is reasonable then his view of the situation should be seen as morally decisive. But "I have explained what draws me in this direction, namely my sense of the scarcity of unique solutions to problems of individual or social choice" (Wiggins 1987, p. 174). Since we can't have perfect

rational consensus (at least in part of the cases), the best option that we have got is striving towards that consensus at the same time that we allow that the peculiar way in which the features of the situation strike the agent counts as morally relevant.

Wiggins's rejection of universalizability is one that is explicitly subjective, but still tries to hold to consistency in some form up to its conclusion. He is aware that with his position we can't have truth in morality, but according to him we can still gladly endorse "...the proposition that p as the best approximation to truth we shall find in this kind of matter" (Wiggins 1987, p. 183). And he concludes by presenting his moral account as a "... sort of subjectivism, oriented towards truth but tilted to allow for essential contestability..." (Wiggins 1987, p. 183).

Wiggins doesn't think we should rely on any moral system expecting to find perfect rational consensus. And that is because people have different values, desires and interests. For him agents A and B can be both right in concluding that the right thing for A to do is X and the right thing for B to do is Y in the same relevant circumstances because they are different people composed of different subjectivities.

In order to illustrate his point Wiggins introduces the following example:

"Consider a small extension of the regrettably short narrative given in Cyril Connolly's fantasy Ackermann's England, in which a dictator instructs his commander Lord Cavalcade to level all buildings built after 1840 and to prepare to restore England to the appearance that it is represented as wearing in Ackermann's Divers Views. Warming to his task and falling deeply in love with these representations and the ravishing beauty of what they depict, Lord Cavalcade reaches a point where he is perfectly prepared to will that, if he were in the position of one living in a house built in the 1870s or 1920s or whatever, then his habitation should be razed to the ground. There is nothing he is overlooking when he becomes prepared to will this—nothing moral and nothing non-moral" (Wiggins 1987, pp. 71-72).

According to Wiggins there is nothing in Cavalcade's 'frame of mind' that could be criticized, especially from the universalizability theorist's point of view, because he is in fact putting himself in the position of the inhabitants that he is removing from the houses and still willing the project to continue. His judgment is thus passing the test of the golden rule, 'do to others as you would have them do to you'.

But for Wiggins when the subjectivities of the different people involved in the case are taken into account things change. Now even Cavalcade would be able to see a problem with his judgment. Wiggins goes on:

"Changing Connolly's story, let us now suppose that in the course of the operation, Cavalcade becomes prey to doubt—but not to doubt about his own hypothetical willingness to be dispossessed of his own habitation wherever it dates from after 1840. Finding how ruthless he has had to be, and being shocked by the physical presence of those whom he has exposed to the weather, he starts to question whether he ought to complete his assignment. The philosophically remarkable thing is that, under these conditions, he could still will that if *he* were the inhabitant of an 1870 house then, whatever he then thought, his house should be razed. (...) But, even though he wills this for himself in the position of an inhabitant of an 1870 house and wills this regardless of what he might then think or feel, it still doesn't seem morally right to him that others (e.g., those without this attachment) should be treated as the dictator had dictated" (Wiggins 1987, p. 72).

In this new version of the example Cavalcade takes into account not only the circumstances with which he would be faced if he were one of the inhabitants removed. He also takes into account his subjectivity and that of the other people involved in the case. Though for someone that like him has developed an aesthetical attachment to the houses as they would become after the project might think that it is worth the trouble to have such a beauty come to reality, he realizes that a different kind of person who doesn't have such attachment might judge the case differently. Wiggins comes to the point of the example:

"What has happened is that, so far from the refurbished preparedness-to-will test restoring a fanatic to his senses, someone who is still in one good ordinary sense a fanatic is here restrained by recognizably moral considerations from acting out his fanaticism and doing to others what he is perfectly prepared for others to do to him in any situation he might be in. His moral hesitation



springs from a source quite other than any that the philosophical universalizer has tapped" (Wiggins 1987, p. 72).

The interesting thing about this new version of the example is that it is precisely by refraining from following the golden rule, and so by treating others in a way that is different from the way in which he would like them to treat him that Cavalcade is able to do the right thing. Thus for Wiggins the universalizability principle doesn't hold in this case. The right thing for Cavalcade to do to others is one and the right thing for them to do to Cavalcade is a different one in relevantly similar circumstances.

Wiggins points out that the difference on the demands to Cavalcade and to the current inhabitants of those houses comes from a source that is different from the source taken into account by the universalizability theorist. But that doesn't stop it being the case that his actual rejection of universalizability is also subjectivist. The reason why it is subjectivist in Winch's case is that if someone happens to think it is alright for them to do something, whether for a good reason or not, if they are really sincere about it that means it is ok for them to do it. Wiggins' view is basically the same except when what they happen to think is too much out of line with what the community happens to think. The difference in Wiggins is that it is not an individual subjectivity but it is a kind of collective subjectivity. This gives very little advance in relation to Winch, and is still subjectivist.

#### 9. The Problem With Winch's and Wiggins' Rejections of Universalizability:

We have looked at Winch's and Wiggins' rejection of the universalizability principle. They are both problematic in that they reject universalizability at the cost of abandoning reason. Almost all of us believe that there needs to be some use of reason in morality. Our moral judgments must accommodate the drawing of "... a distinction between sense and nonsense in moral discourse, with allowing for an intelligible relation between moral judgements and the reasons offered for them" (Winch 1965, p. 208). We think that when we judge an action to be morally right there must be reasons for our thinking that it is right. And we do so in such a way that when we have two cases in which all the morally relevant reasons apply in the same way we must judge them alike. This is exactly what Winch's and Wiggins' accounts lack.

Winch's and Wiggins' rejections of universalizability bring with them an important sense of trying to keep reason in morality and of having consistency in some form. But both Winch's and Wiggins' rejections of universalizability abandon reason at some point along the way, not by giving up consistency completely but rather in virtue of the fact that their positions end up being subjectivist.

In Wiggins' case the attempt to preserve consistency revolves around his notion of the moral point of view. In principle Wiggins' consistency would come out of the checking of one's thoughts, feelings and actions against society's shared sensibility. But he doesn't believe in actual societies reaching perfect consensus. In order to solve the problems that that creates he allows for a certain privilege of the participant's perspective, in Winch's style. It is true that at the same time he proposes that when there is no consensus we should keep looking for problems in the judgment. But Wiggins holds that where there is no decisive criticism such as the one he claims to have found in Vere's thinking, he is "... prepared to treat [the agent's] finding as morally decisive of what was right for him" (Wiggins 1987, p. 181). A problem

with Wiggins' view here is that there is no clear definition as to what counts as 'decisive criticism' and when 'how the agent is struck by the situation' will by itself determine the right thing for him to do.

As we have seen Winch is actually explicit in his intent to insist on consistency in morality. But his view on how much he has to deny in order to reject universalizability is problematic. For instance he presents and denies the following quote from Singer:

"Hence to give a reason in support of the judgement that a given individual ought or has the right to do some act presupposes that anyone with the characteristics specified in the statement of the reason ought, or has the right, to do the same kind of act in a situation of the kind specified" (quoted on: Winch 1965, p. 204).

Winch thinks that in order to deny universalizability he must deny all the different versions that have been offered of the universalizability principle. And thus by denying even the version that basically states that similar cases must be judged alike he denies what it is just a demand for minimal consistency. But as we have seen one can confidently reject universalizability while accepting supervenience and what is just a requirement of minimum consistency.

A question to ask ourselves is, why does Winch think he needs to go so far as saying that we don't need to judge similar cases alike? And the answer is that there is working in Winch the same minimal conception of personal constitution that we find in universalizability theorists such as Hare, Mackie and Sidgwick. Like them Winch can't see that who you are actually involves a lot of relevant difference, and that is what guarantees consistency. Winch can't see that there is an objective sense in which what is right for me to do depends on my character. Of course I can and should transform my character in certain ways but I can't become another person. I can't become relevantly indistinguishable from somebody else. And what is underlying this

whole issue is just an insufficiently robust conception of what it is to be a person.

There is a kind of sense in which although I have quoted Winch above as adopting a conception that rejects even the minimal consistency version of universalizability, it looks as if there is some sort of consistency that Winch could hold on to. The idea would be that what is required for having consistency is that if this is right for A to do then it must be right for anybody else to do if they are similar people and in similar situations. And it looks like there is a kind of sense in which whatever he says Winch could accept that because effectively he allows a difference between the kinds of person to emerge in the different moral ideas they have. So although that is not the way he actually proceeds, he could hold on to that by describing the differences between people in this way.

The reason why this is important for us is because it is helpful in clarifying the notion of consistency in my account. On my view it has got to be the case that the differences between people, which aren't just differences between people in the sense that they don't require everybody to act in the same way, can't be just thought of as being just differences of moral outlook. In a way they could issue in differences in moral outlook but those differences in moral outlook must be based in other differences of character. What Winch seems to want to say is that a mere difference in moral outlook on its own will be enough.

My view is that the differences in moral outlook that arise from a genuine difference of kind of person will be alright, but not any difference of moral outlook. And that is basically because you can't expect people to become other than they are. They can be better versions of themselves but they can't be different people. Underlying this is an application of the thesis that *ought* implies *can*. The thought

here is that you can only require someone to do something which they can do. And they can become better versions of themselves. But they can't become different people. So any change of moral outlook which would require them to be a different person is not ok.

The problem with Wiggins' view in this regard is twofold. On the one hand he also puts no constraint on differences between moral outlook, in the sense that genuine differences don't need to be based on genuine differences of character. On the other hand he thinks we need to strive towards consensus and in that striving there is nothing to prevent the agent being required to act in such a way that he could no longer be the same person.

But though clarifying this issue will be useful for making my view more precise, it is somewhere else that we can find my main contention with Winch's and Wiggins' rejections of the universalizability principle. The problem comes up in the way in which they set out to reject universalizability at the same time that they want to avoid that what the agent thinks is right or what the agent merely wants to do becomes the right thing for him to do.

Winch carefully distances himself from an extreme form of subjectivism and seems to be successful in avoiding that in his position the agent has room to evade a 'distasteful duty' and still passes for doing the right thing. In other words his position is successful in avoiding that the agent does whatever he wants to do (Wiggins likewise). But he fails to see another way in which the conclusion that 'what the agent believes is the right thing for him to do becomes the right thing for him to do' follows from a less extreme yet still radical subjectivism, namely his kind of position.

Winch thinks that in order to determine what is right for the agent to do one must take into account how the situation strikes them. He also thinks that this can be understood only in terms of "... the moral ideas

by consideration of which he arrives at his decision" (Winch 1965, p. 212). What justifies, in his account, the different moral demands that different agents face in relevantly similar situations is their prior moral outlook.

So the question is, does Winch succeed in avoiding that what the agent thinks is right becomes right? In other words does Winch manage to distance his position from a kind of subjectivism that he himself sees as a threat to the consistent use of reason in morality? And the answer is no, he fails his own challenge. Since what justifies the difference in demands on different people is nothing but the prior moral outlook they have, it comes down precisely to 'what the agent thinks is right becomes the right thing for them to do'.

Again, Winch does succeed in denying that 'whatever he thinks is right' in the sense of 'whatever he wants to do' becomes right. But it is still the case in his account that 'whatever he thinks is right' in the sense of 'whichever sincere beliefs in terms of what is right for him to do' becomes the right thing for him to do. That is still for Winch what justifies the different moral demands on different people. And the problem is that if you press hard you can't actually insist on reason while holding that kind of subjectivism. Now, that is exactly the motivation for the universalizability principle in the first place.

Winch's and Wiggins' positions don't really solve the universalizability theorist's 'concern about the evils of arbitrariness', as they don't allow room for reason in morality. At the same time the universalizability theorist is unable to solve Winch's and Wiggins' concerns about a view of morality that is intrinsically disrespectful towards human beings, as it leaves no room for acknowledging differences between people that are indeed morally relevant. Now the interesting thing is that there is no real problem here. We can accept the right kind of universalizability which is just minimal consistency. And we can also keep reason,

consistency and supervenience in morality at the same time that we allow that different people may be demanded to act differently in relevantly similar circumstances.

The interesting thing is that Winch rejects even the consistency version of universalizability even though it looks as if he could hold on to it and that is because in effect it doesn't occur to him to think that a moral outlook might depend upon what kind of person you are. And that goes back to his thin conception of personal constitution. Winch positions his view in this way because he can't see that there are morally relevant differences which actually do depend on who you are.

#### 10. Genuine Morally Relevant Differences Between People:

In order to illustrate how their rejections of universalizability could work Winch and Wiggins rely on cases which they consider to be moral dilemmas. But I think everyday cases can actually provide very convincing examples. I will now use a series of everyday cases to show how it can be quite plausible that who you are might make a difference to what you should do.

Suppose there is a child misbehaving in the supermarket. His mother is near him doing her shopping and you are also around. Imagine that the child is behaving really badly, screaming, shouting and drawing attention. In those circumstances it is ok for his mother to intervene in a way that is not ok for you to intervene. And it is ok for her to intervene and not for you in virtue of the circumstances. She is the child's mother and you are not. So there is a role here that is generalizable (being the parent), and separate from their being the person they are.

Whether or not you are the parent of the child that is misbehaving in the supermarket is a matter of circumstances. This is a case where you detach the person from the situation and the role. But not all cases are like that. Imagine that a small child is drowning in a little swimming pool of 5 foot deep and they are only 3 foot tall. So they are drowning in there but it is only tiny. If there is nobody there they drown. But if there is an adult with ordinary size and strength there the adult has an obligation to get in and save the child. There is no risk to him, he is the only person that can do something about it, and so it is a clear obligation. Now, if the only other person that is there is another child of the same size there is nothing they can do about it and so the obligation is not on them. We just end up with a bad situation. I think that features of personality can give us a situation that is in a sense like the latter one.

Let's come back to the supermarket case, but with a modification. Suppose we have the same child behaving really badly in the supermarket. His mother is not around this time because, say, she had to go to the toilet and asked the child to stay put with the shopping until she is back. Suppose the child takes the opportunity and starts to agitatedly scream and shout at people. We can from here think of ways in which what is morally permissible differ for two people, and this time not due to the nature of circumstances. So suppose there are two people near the child: person A, a small women, and person B, a big man. We can imagine a scene in which given the state of the child it would look very dangerous for anyone but a woman to intervene, otherwise it would look as if the intervention is a kind of use of force. So it is ok for person A to intervene, even though she is not the parent. But is not ok for person B to intervene.

And we can think of other examples that work in a similar way. Suppose for instance that you are a man and have a female friend who is in distress. Imagine that your female friend has very recently been



raped<sup>4</sup>. It is plausible to imagine that some people would be particularly in shock soon after the experience, so let's suppose that this is your friend's case. She has been recovering, but very slowly. Trying to talk her out of her state only makes things worse at this point. So the only way to comfort her is to give her a hug. Now, given the circumstances that prompted her shock, at this early stage the last thing that she wants is to be touched by a man (any man). This case involves circumstances in which it would be inappropriate for a male friend to comfort the person. So comforting her is something you can't do, yet there are other (female) friends who would be able to do it. Therefore there couldn't be a demand on you to comfort the friend because that would require you to be another person, and you can't do that.

At this point one can see how some philosophers may want to insist in explaining all morally relevant differences in the above cases as differences of circumstance. But the problem is that means treating your gender as something merely contingent to who you are. And it is just not plausible that the same person could have had a different gender. Yet the theory of universalizability in its standard form requires an insufficiently robust conception of personal constitution<sup>5</sup>. They operate with a schematic conception of what it is to be a person, a kind of blank. But to be a certain person is a much richer thing. So is not the case that a person is a kind of a blank that just is in a situation and that all the variation is variation between situations. And we can see how the gender case is an obvious instance in which the variation is a variation related to the people involved as opposed to being merely a variation of circumstances.

---

<sup>4</sup> I ask the reader to forgive the darkness of the example. Though "heavy" the example is a clear one.

<sup>5</sup> In this thesis I will refer to this insufficiently robust conception of personal constitution as the blank conception of the self. What I mean by that will become clearer in Chapter Two, when I will contrast the blank conception of the self with the alternative conception proposed by me: the richer-genius conception of the self.

Some people might want to say that one's gender can't be seen as essential to who one is, otherwise how could we explain the cases of transgender people? Surely, they could say, if someone starts off life as a man and decides to undergo a sex-change process to become a woman, being a man wasn't essential to them. But I think that transgender people actually are best understood as not starting off as being clearly one gender and then being the same person, and afterwards becoming clearly another gender. It is rather, roughly, that they start off being uncertainly either way but with a kind of what feels like an external determination in one way and then they remove that external determination. Because the general experience of transgender people as far as I can tell is that they don't think they are changing genders, but rather they think they are becoming what they always were.

Yet I am not only saying that a person is not a blank. I am also saying that a person is not merely a member of a slightly narrower class of near-blanks either, like man or women. It is just not intelligible that the same person might have been a woman (or man). Now once we have accepted that that means that the general principle is correct. So the question is how far it extends.

A line of generalization of the above principle comes down to questions about a person's style. And some features of a person's style are ineliminable. We know that people react differently in their role. We just don't expect people to behave in exactly the same ways in the same situations. And it is not just that we don't expect them to, they can't be required to.

Sports people give a good example of that. Let's look at the example of football for instance. Players have totally different temperaments and interesting cases arise in criticism of great players that have distinctive styles. There is a Brazilian football player known as Paulo

Henrique Ganso, which people refer to as a classic midfield organizer of the game. He does everything with just a kind of lazy style, and is all about perfect timing. His style of playing involves waiting for the perfect moment to make a move. In his case there is no force involved at any point, it is all just down to the style. Off course, playing with that style means that he is inevitably going to make a certain kind of mistake. For example he is sometimes criticized for not having, in particular instances, used enough strength in disputing the ball with an adversary. But there is a sense in which you can't criticize him for a certain kind of mistake, for that would be to ask him to be a different kind of player and he can't be that. The only player he can be is this one. We can contrast Ganso's case with Luis Fabiano, a kind of fierce Brazilian striker. He sometimes makes bad judgments and does stupid things out of being excessively avid in specific situations of the game, but it is absurd to criticize him for part of those stupid things because he can't play in any other way. He has to be constantly inventing aggressive things otherwise he can't do it. That is him.

The point would also apply to the distinctive style of great artists. Different great singers for instance have different styles. What gives the force of Metallica's singer James Hetfield's style is a kind of raw force that gives him a distinctive presence. And what gives the force of Morcheeba's singer Skye Edwards' style is a kind of gentle and sweet yet deeply penetrating way of singing. One might, following his taste, wish that James started singing more like Skye and criticize him for not doing so. But though singers mature over time - which involves refining, incorporating, and eliminating certain elements of their singing - and can learn certain specific techniques used by other singers, it can't be required of James that he sings in the way that Skye does.

We can also see questions of style coming in for instance in the case of teachers in school. To be effective teachers they have to maintain

discipline at some general level. But it is not required that every teacher responds in the same way to a particular incident, as teachers have totally different styles. And their whole personality is involved in a way, so you can't expect that they respond in the same way.

We don't in general expect people to behave in exactly the same ways in the same situations. And it would be surprising if that didn't carry across to the area of morality. There are ways in which you can't criticize this person for doing that because that would be a different person. We can see that if we think of these cases which depend on style, and style is so close to a person's character. In these cases the person's style restricts the kinds of things they can do while being that person. That means that what is required of them has to fall within that range. And so it is quite possible for something to fall within this person's range and not fall within the other person's range.

The thought here is that someone's style in these non-moral cases restricts the range of things that a person can do while still being, for instance, 'the player they are'. Now let's also here plug in the principle that ought implies can: you can only be obliged to do something if you can do it. Well then what someone can be required to do has to fall within the range of things they can do while being that person. And it is obviously possible for something to fall within the range of things that one person can do and not within the range of things that another person can do. So it looks as if it might be the case that this is the response that person A ought to have but it is not the response that person B ought to have. So this is the play person A should make but it is not the play that person B should make.

The issue is this: the only way of maintain universalizability now seems to be to hold that who you are is just a blank thing and everything else is circumstance, history, role or something like that. But once we have got questions of style coming in we can sense how those questions of

style can lead to other sorts of things coming in. There are whole ways of being that make persons be the person they are.

We can imagine for instance someone whose whole rhythm of life is measured in a certain way. There is a certain pace with which he does things, and that is the way in which he does everything he does. Trying to force this person to act in ways that obviously contradict that rhythm is trying to make them a different person, and that is not all right. The same kind of point can be made if we think of someone who is very extroverted and has a touch of sense of humour engrained to the way in which he expresses himself. Though it is ok to require that they be, say, more reserved in certain circumstances, it would still not be ok to require that they do some given things in certain ways. It just wouldn't be them. Because people can and should improve themselves, but improving themselves in terms of becoming better versions of themselves is ok and is one thing, while requiring that they become another person is something else and it is not ok. And again, there is no reason to suppose that that doesn't carry across to the area of morality.

Imagine for instance that we are friends and another friend of ours has a tragedy happen in their life. So we both feel that we need to give him some support. But while you are extroverted, light-hearted and have a highly humorous demeanour I am more introverted, sensitive and reflexive. We can imagine a scene in which the way that is right for you to support the friend (given that you are extroverted and so on) would be by, say, lightening the air, cheering him up or something of the like. But I couldn't be demanded to support the friend in the way in which it would be right for you to support the friend. And if I did try to do it we can see how it would end up in a kind of disaster.

It could be objected that though we can meet the obligation with different styles the demand is still the same for both of us, i.e.

supporting the friend. I would see that as a highly suspect claim because there are cases in which a difference in styles would issue in two different actions. So for instance in the same situation the right thing for you to do would be to say amusing things and cheer someone up, and so provoke a kind of momentary forgetting and letting go of the friend's worries and fears of the moment, whereas the right thing for me to do would be to engage in the friend's worries and help them see the situation from another angle which would be liberating for the friend in some important way.

But for the sake of argument let's suppose for a moment that we could accept the view according to which independently of those being distinct actions we should still consider that the demand on both cases is exactly the same, i.e. supporting the friend. We would only need to change the example slightly for it to work. Because it might be that in the circumstances the only way of supporting our friend would be to do something which I could do and you couldn't do. So in that case there will be a requirement for me to support the friend but not on you. There might be nothing you could do which would be a support. Whereas there is something I could do. In the new version of the example there is not a level of generalization which says, ok you are both required to support the friend – this is what you have to do and that is what you have to do. It might be the case that in the circumstances the only way in which this friend could be supported is something within this range, and it is plausible that would be outside of your range.

Now, once we have got to this point what we need is a principled way of avoiding the result that anything goes. Because it can't be the case that when we face any moral failing of ours we say, well I am just not the sort of person who can do that. But nevertheless it is not the case that everybody has exactly the same requirements on them. And we can see how the principle of universalizability looks as if it is in danger

of having the following consequence: in an ideal world we would all be indistinguishable from each other. Now the problem is that doesn't look like an ideal world at all.

We all ought to improve ourselves as much as possible. But there are limitations on what is possible and those limitations are not indications of failure. They are rather indications of a kind of structural limitation in life. If we think about it, a moral demand is that something be done, and the reason why it is a demand is roughly because things will be bad if it is not done. Now, from the examples I have been discussing one can perhaps sense a situation which is a bad one, which would exert a demand on some people but unfortunately there isn't anyone like that there. And we just end up with a bad situation.

Let's think of an example inspired in the song "She's Leaving Home" from The Beatles. Imagine a father and a child whose relationship has deteriorated over time in a way that appeared to be kind of inoffensive at first. The father genuinely wanted to offer the child something that accorded with his own view of what it is a good life. Suppose he happened to have had poor parents and indeed grew up as a child in poverty. Suppose that things got a bit better for his parents around his adolescence which helped him to secure for himself a living away from poverty latter on. And because of some rough times he has gone through in the past the father thought he needed to make his child manage to take good care of himself latter on, with a special emphasis on having a stable and comfortable material life. Suppose that he becomes overly concerned with the issue, and starts forcing the child in that direction without attending to the fact that he was neglecting other important aspects in the child's life. He ends up by treating the child in a disrespectful way for he fails to see that, as the Beatles' song says, there was "something inside [of the child] that was always denied for so many years". And being of an introverted, taciturn and dour

character his lack of proper communication of his intentions and of his love never really helped the situation.

Imagine now that though the child loves the father he has been feeling increasingly oppressed over time. Though he appreciates good instruction, he has his own view on things and his own interests he wants to pursue. It is not implausible to suppose that a child will have some of the same traces of the father, so we will suppose that this is this child's case. Being still immature and having never been good at expressing his feelings honestly he has in a way repressed a great part of his revolt against the treatment he received from his father until a point came at which he couldn't hold it anymore. At this point he became aggressive and intractable which only contributed to the escalation of tension between the two. That in its turn caused a serious of misunderstandings, the situation became really bad between the two and there emerged a serious risk of they drifting apart up to a point of never seeing each other ever again. So imagine that they are coming to what might be their last meeting. Suppose that both of them want to avoid the worse, but things were so bad between them that they wouldn't be able to sort things out unless one of them makes for instance a light joke about the whole thing. One of them needs to be able to say something with a certain character in order to bring the other in. But neither can. And that is not because of any failing in them, because to require them to do that would be to require them to be other than they are. They end up with a really bad situation and that is just unfortunate. There is nothing they could do in some circumstances. It is as if there is a kind of demand that circumstances place which they couldn't meet though other people could.

The deep thing this example shows is that there is a common kind of optimism in morality that is wrong. According to this view everything can be made good. But what we can see is that is not true. What you end up with if you take Agent Particularism seriously is a kind of



approach to moral theory which is both more forgiving and less optimistic. And this is what that example shows. You realize that this was a tragedy that could not have been stopped. There are situations that in a sense demand something that those around can't do. And not because of any failing in them as a person but because of them being the person they are. But on the other hand you realize that neither is to blame, because what could they do? After all we are all the time prepared to acknowledge that in relationships there are irreconcilable differences. And it is not the case that those are always blameworthy. Someone else could have done something but not these two.

We could accept the view that for instance the father from our last example needs to develop his character in some way, like developing the capacity to express his feelings more honestly and directly. But improving them as a person doesn't turn them into a different person. The improvement of a person as a person does not mean removing everything that is personal to them so they become a kind of universal template. Therefore what is required of them needs to respect the person they are.

There is a tricky thing: imagine someone who has a tendency to say, whenever they fail to do something they ought to do, "I am just not that kind of person, I am just always like that". Imagine that the way they say it makes it seem as if they couldn't really be obliged to do it because they are not the sort of person that can do it. And you feel this isn't good enough. There is something else that is interesting about that. Though they oughtn't to be like that, it is part of their character that they are going to find it difficult not to be like that. And they can't help that it is going to be difficult for them to not be like that. Now, while some people accept their own failings too readily, other people are too anxious about their own failings. And that leads us into the notion of self-improvement. Self-improvement involves making sense of our constraints, where we really are talking about what is intrinsic

to the person. Yet, we are not allowing anything to go. The whole idea of self-improvement is that for each person there is a way in which they could be better than they are - but it is not the same as anybody else<sup>6</sup>.

The aim of the thesis from here is going to be to find a way of accommodating that. It will be on the one hand more forgiving because it acknowledges that there are things that people can't do but on the other hand it will be less optimistic because it doesn't think that everything can be made good. We all like the idea that we can make the world perfect but actually we can't. And that is kind of a sad thing. But the counterpart of it is that we now recognize that is what people are. So in both respects it is kind of more realistic. And the challenge will be to provide a principled way of allowing that the substantial differences between different people can be morally relevant without just making it the case that anything goes. It will be to keep consistency while allowing the development of what is proper to a person.

---

<sup>6</sup> In Chapter Two I will enlarge on this issue as I develop my account of a person's proper development.

## **An Agent-Particularist Conception of Freedom**

### 1.1. Rousseau's Richer Conception of a Person:

We saw in Chapter One that what inclines people not to accept Agent-Particularism is a kind of blank conception of what it is to be a person. In this chapter I will start to build up a kind of ethical approach which is Agent-Particularist. The initial job will be to look at a way of thinking of the person more richly than just a blank.

In order to do that I am going to see what we can develop in terms of a richer conception of a person from Jean-Jacques Rousseau, a philosopher who clearly doesn't have the blank conception. Rousseau is in particular a promising place to look because he starts with a substantial conception of what is already natural, already integral to a person. That is particularly apparent in his work 'Emile, or On Education'.

Rousseau's view of what is already natural to a person comes from how much Rousseau attributes to nature in the first place: "It is an incontestable maxim that the first movements of nature are always right" (Rousseau 2010, p. 161). His view is that what comes from nature is good, and that grounds his claim that "[t]here is no original perversity in the human heart" (Rousseau 2010, p. 161). Therefore as long as man keeps its natural form he will stay good: "Let us not corrupt man! He will always be good without difficulty and always be happy without remorse" (Rousseau 2010, p. 456).

For Rousseau to corrupt a man is to corrupt his nature. But this doesn't mean to corrupt human nature in some general sense, because he believes each person has a particular nature: "We do not know what

our nature permits us to be. None of us has measured the distance which can exist between one man and another" (Rousseau 2010, p. 190).

The fact that we are by nature different from one another has implications for Rousseau's treatise on education. In general terms it implies that a person is only able to develop properly when their individuality is taken into account. Rousseau's conception of a proper education is centred around the view that:

"Each child has his unique genius and character. One must not try to change or restrain it but to allow it to perfect itself, for all human beings are originally good and all vices observable in human character come from the false forms society tries to impose on their natural geniuses. Education, then, is a matter of allowing natural penchants to develop fully" (Rousseau 1968, p. 417).

In Rousseau's view there are two aspects to education. There is first a negative education in the sense of preserving what is integral to the individual. But there is the possibility of something else given that there is an education to be had - there is a positive education in the form of fostering the particular genius in the child. The key thing about both of those aspects is that we must respect something which is unique to each individual. From this we can see that for Rousseau to be a person is a much richer thing than a mere blank. And his notion of a genius carries within it exactly this sense of robust individuality.

The Oxford English Dictionary reveals relevant applications of the term genius in the 18<sup>th</sup> century (Rousseau's life period). Under the entry on "Character, ability, and related senses" we find W. Derham's application in 1713: "There is the same Reason for the Variety of Genii, or Inclinations of Men also". We also find B. Franklin's application in 1729: "Different Men have Genius's adapted to Variety of different Arts and Manufactures". Rousseau is writing in French, not English, and we can see that the same kind of application of the term is also found in

its French counterpart, *génie*. For instance in the French Lexicon dictionary from the *Centre Nationale de Ressources Textuelles e Lexicales* we find the sense “...ensemble des aptitudes innées, des facultés intellectuelles, des dispositions morales”, which translates into ‘group of innate aptitudes, intellectual faculties, moral dispositions’.

The Oxford English Dictionary also offers an application of the term genius which is strikingly similar to Rousseau’s views on education, though it is actually an approach to gardening. It is centred around the notion of the ‘genius of a place’ introduced by T. Whately. The *Observ. Mod. Gardening* offers the following quote from 1770: “In this application, the genius of the place must be particularly considered; to force it is hazardous, and an attempt to contradict it is always unsuccessful”. Effectively what the gardener had to do is recognize that the place had certain distinctive unique capacities, and then develop a particular better version of those. Spence offers a practical comment on gardening that further illustrates the point:

“... but as to any large tract of ground, there is no saying anything in particular without being upon the spot; and having considered it well and often. (...) consult the genius of the place. What is, is the great guide as to what ought to be” (Spence 1966, p. 646).

In the case of gardening Spence treats the particular genius of the place as the main guide of its proper development. In education Rousseau treats the pupil’s unique genius as such reference. The following quote from Rousseau attest to the similarity of the approaches:

“One must know well the particular genius of the child in order to know what moral diet suits him. Each mind has its own form, according to which it needs to be governed; the success of one’s care depends on governing it by this form and not by another. Prudent man, spy out nature for a long time; observe your pupil well before saying the first word to him. To start with, let the germ of his character reveal itself freely” (Rousseau 2010, p. 227).

Clearly the 18<sup>th</sup> century understanding of the term genius has a lot to say in terms of what Rousseau had in mind when employing the term. Rousseau's view is not just that people are different. The most important point is that there is actually a lot of substance to that difference. Rousseau believes that the particular genius of a person needs to be respected and then developed. But the point is not merely that you need to respect a particular person's genius because it is theirs. It is rather that the actual content of such uniqueness, which is integral to the person, requires that respect. In order to understand how a person's natural character is good we need to look at the actual individual character and find out what is it that is good about it.

Since Rousseau attributes so much importance to people's natural genius he wants his pupil to:

"... esteem each individual but despise the multitude; let him see that all men wear pretty much the same mask, but let him also know that there are faces more beautiful than the mask covering them" (Rousseau 2010, p. 391).

There are different points being made in the quote above. One is clearly the remark that we need to respect what is particular to each individual. There is something beautiful about the particularity of each person, and that is what his pupil should consider. But Rousseau also points out to a danger coming from his view that people behave very similarly or in some sense adopt very similar attitudes. Since what needs to be respected is what is particular to each person we are immediately made suspicious about whatever has been done with the adoption of the same attitude. This point introduces Rousseau's concern with a threat to individual freedom coming from life within society. This is something to which we will come back to later, as it will play an important part when we come to our discussion on negative freedom.

## 1.2. Berlin's Two Concepts of Liberty:

I will offer in this chapter an Agent-Particularist conception of freedom. I will do that by building upon the ideas of Rousseau and John Stuart Mill mainly, but also Wilhelm Von Humboldt and Ralph Waldo Emerson – all of whom clearly work with a rich conception of the self. I will rely on Rousseau for constructing the central thread of my conception, using his conception of individual nature or genius as the basis for an account of freedom which will have both a negative and a positive aspect. And I will use ideas coming from Mill (together with Humboldt and Emerson) to enrich the conception.

But in order to clarify the sense in which this will be a conception of freedom it will be helpful to first set it in the context of recent philosophical discussion. In this sense Isaiah's Berlin work 'Two concepts of liberty' is particularly helpful, as Berlin introduces with it an influential distinction between negative and positive conceptions of freedom. So a starting point for the development of an account of freedom suitable for Agent Particularism is by thinking about problems with Berlin's account.

Berlin claims that when we look at the main conceptions of freedom offered throughout the history of philosophy, both in ethical and political traditions, we see that they can be divided into two main groups: Philosophers holding a negative conception form one group and philosophers holding a positive conception form the other. His view is that not only negative and positive accounts of freedom are distinct, but they represent "... two divergent and irreconcilable attitudes" (Berlin 1958, p. 212).

In the political sense the contrast is between negative freedom, which means not being interfered with, and positive freedom, which in some

sense means being my own master. For Berlin while conceptions of negative freedom are concerned with the question 'What is the area within which the subject should be left to do or be what he is able to do or be, without interference by other persons?', conceptions of positive freedom are concerned with the question 'What, or who, is the source of control?' (Berlin 1958, p. 169). Berlin believes that there is a clear distinction between his negative and positive conceptions of freedom. In order to see why he is not right about that we will look at his distinction in a bit more detail, paying special attention to the different uses to which the notion of external influences is put.

For Berlin freedom, in the negative sense, is understood in terms of "the area within which a man can act unobstructed by others" (Berlin 1958, p. 169). The notion of external influences from which one is free in the negative sense is therefore clearly understood in terms of other people. You are free to the extent that you are left to do what you want in the way that you want to do it.

Berlin's positive conception of freedom on the other hand, "derives from the wish on the part of the individual to be his own master" (Berlin 1958, p. 178). Here the notion of external influences is used in a different way: "I wish my life and decisions to depend on myself, not on external forces of whatever kind, to be moved by reasons, by conscious purposes, which are my own, not by causes which affect me, as it were, from outside" (Berlin 1958, p. 178). Berlin's view is that when I am my own master something like my true self is the source of control, rather than my passing desires and passions. According to Berlin the question asked by the positive conception philosopher to the proponent of the negative conception is: "I am my own master"; 'I am slave to no man'; but may I not be a slave to nature? Or to my own 'unbridled' passions?" (Berlin 1958, p. 179).



Berlin's positive conception of freedom comes from the view that man becomes free from slavery to nature, and in the course of doing that he becomes conscious of a true self that dominates the lower self. In this view the lower self that is brought to discipline by the true self is identified with nature. It is from there that influences that are external to one's true self come. According to Berlin's understanding of the positive conception tradition,

"This dominant self is then variously identified with reason ... with my 'real', or 'ideal', or 'autonomous' self, ... which is then contrasted with irrational impulse, uncontrolled desires, my 'lower' nature, the pursuit of immediate pleasures..." (Berlin 1958, p. 179).

That is the basic understanding of Berlin's conceptions of negative and positive freedom. We can find different problems with Berlin's understanding of freedom in the literature. I will present here the most important points that can be found in the literature, and develop them in light of my own account of freedom.

There are two main problems, on my view, in Berlin's theory. They are related to: a) What the distinction that Berlin makes amounts to; and b) The fact that there is a misconception of some of the central philosophers that Berlin relies on. Within the question what the distinction amounts to two sub-questions have already been pointed out by other philosophers. The first is the view that there is not really much of a distinction. The other is that Berlin's positive conception of freedom is actually not truly positive. Let's take a closer look at them.

It has been observed by Gerald C. MacCallum that Berlin "... cannot distinguish two genuinely different kinds of freedom..." (MacCallum 1967, p. 318). The point is that although there is a kind of formal difference between these two conceptions there is not very much difference, which renders the distinction unclear. This claim brings us back to the different utilizations to which the notion of external

influences (or interferences or obstacles) is put among philosophers taken by Berlin as being proponents of the negative or positive conceptions. The problem, according to Raymond Geuss, is that Berlin fails to acknowledge that "[t]he notion of an "obstacle" is a flexible one that needs interpretation" (Geuss 2001, p. 91).

On my view the issue is that although it seems as if there ought to be a contrast between positive and negative freedom there is not very much difference between the views as Berlin presents them. Basically both of Berlin's positive and negative conceptions come down to being free from external interference. And the difference is just that the conception of what counts as external changes between in some sense pre-philosophical (external understood as other people) and philosophical (understood as external to my "true self", including internal threats). For this reason it is very odd to describe the distinction that Berlin finds as a distinction between positive and negative freedom.

What this means is that although Berlin tries to make a contrast between positive and negative freedom the striking thing is that the positive isn't very positive – which is the second sub-question from the broader issue about what Berlin's distinction amounts to. Quentin Skinner presents the point in the following way:

"... the positive sense of the word refers to the idea of being my own master as opposed to being acted upon by external forces. But this too fails to isolate a separate concept of positive liberty. For the situation in which I am free to act in virtue of not being hindered by external forces is, according to Berlin's own analysis, that of someone in possession of their liberty in the ordinary negative sense" (Skinner 2001, p. 239).

As we have already seen Berlin's understanding of being my own master, which is what distinguishes the positive conception, involves not being enslaved by internal hindrances such as my irrational passions. But even though they are internal in the sense of coming

from me as opposed to coming from other people, they still count as external to something like 'my true self' – a common notion, in some form or another, among the philosophers to whom Berlin attributes the positive conception. Once the notion of something like one's true self is brought into the picture the notion of external interference or constraint comes to "... include psychic, internal forces in the universe of possible constraints" (Nelson 2005, p. 60), and thus "we are still speaking about the need to get rid of an element of constraint if we are to act freely, and we are still speaking in consequence about the idea of negative liberty" (Skinner 2001, p. 239).

We can immediately see how Berlin's positive conception of freedom sounds like just a different formulation of his negative conception. Eric Nelson concludes the point in pointing out that Berlin's conception of positive freedom:

"... does not employ a separate 'concept' of liberty: although the constraints it envisages are internal, rather than external, forces, the freedom in question remains negative (freedom is still seen as the absence of such impediments)" (Nelson 2005, p. 58).

If one were to ask oneself what the difference between Berlin's two conceptions is, the actual formal characterization just looks as if it is the same for both, so that all that the substantial account that Berlin can give of what positive freedom is just looks like negative freedom. And the only difference is whether some part of the person is counted as alien.

In other words, if I wanted to try and explain it just in the abstract what was the difference between negative and positive freedom on Berlin's own characterization it is very hard to make sense of 'being in control of one's actions' as being anything other than a contrast with not being subject to external interference. If one just reads what the words mean, it looks as if the only sense that Berlin is able to give to

being in control of one's action is this one which just means that not the other people are in control, i.e. non-interference. He gives no more sense to it. The only way in which he tries to find more sense is by distinguishing between those philosophers who count parts of the self as foreign to one side, and those who believe that the only external interference is coming from other people to the other. That is indeed a distinction, but it is understood in both cases as being free from external interference. So the fundamental conception is the same. Therefore in the end both conceptions just look like different forms of negative freedom.

The second broader problem with Berlin's account concerns a kind of curious misrepresentation of the authors he considers. MacCallum makes this point in the following way:

"The trouble is not merely that some writers do not fit too well where they have been placed; it is rather that writers who are purportedly the very models of membership in one camp or the other ... do not fit very well where they have been placed, [which is] conducive to distortion of important views on freedom" (MacCallum 1967, p. 322).

MacCallum doesn't really develop the point explicitly. So I will offer my own explanation of the point, which in any case would be for us the most helpful way of doing it, given the content of the following parts of this chapter.

There is actually something blurred about Berlin's distinction from the very start. "Liberty in this [negative] sense is principally concerned with the area of control, not with its source" (1958, p. 176), says Berlin. But is it fair to say about the philosophers to whom he attributes the negative conception of freedom that they were primarily concerned with the area but not the source of control? John Stuart Mill is, according to Berlin, the most celebrated of the champions of the negative conception – therefore a good place to investigate this point.

In Mill's view, "The only freedom which deserves the name, is that of pursuing our own good in our own way" (Mill quoted in Berlin 1956, p. 174). For Berlin this just means not being interfered with. There is no question about Mill having included in his conception of freedom the notion of not being interfered with by others. But is that all his conception of freedom includes?

There is another way of understanding what is at stake here. It is only reasonable to think that Mill is not concerned with the source of control if we think that he takes freedom from interference to be an end in itself and not valuable because of the source of control. But it is not an end in itself for Mill. The reason why being free from interference matters is because what you need is for yourself to be properly developed. So the important thing for Mill is that actions spring from one's self:

"Where, not the person's own character, but the traditions or customs of other people are the rule of conduct, there is wanting one of the principal ingredients of human happiness, and quite the chief ingredient of individual and social progress" (Mill 1991, p. 63).

So not being interfered with is not for Mill an end in itself, it is rather a means for respecting one's individuality. People should be left to do their own things in their own way not merely because it should be their own choice. It is rather that it should be their choice because their choices should take into account their own character and different people may have distinct characters: "Customs are made for customary circumstances, and customary characters; and his circumstances or his character may be uncustomary" (Mill 1991, p. 65).

The point here is to clarify how it is not the case that for Mill not being interfered with is an end in itself. What is striking about Mill is that his concern with freedom is much richer and more interesting than Berlin suggests. And that raises a question: Why is it the case that, when we

look at a central figure in Berlin's negative conception of freedom, we don't find an account which is merely concerned with not being interfered with by others? It is due to the conception of individual nature that Berlin has - I will come back to this issue later. For now let me just say that it is hard not to see Mill as accepting a view of human nature along the same lines as Rousseau's understanding of the notion of one's particular genius. Mill explicitly wants to offer us "a conception of humanity as having its nature bestowed on it for other purposes than merely to be abnegated" (Mill 1991, p. 69). There are two main points being made in this quote. One is that abnegating one's nature is a bad thing. Now, when this quote is looked at in the context of its surrounding text one has a clear sense of 'having one's nature abnegated' being understood by Mill in terms very similar to Rousseau's understanding of similar phrases. Like Rousseau Mill has a conception of human nature with quite a lot of substance to it, and that is reflected in the importance that he attributes to one's particular character:

"A person whose desires and impulses are his own - are the expression of his own nature ... is said to have a character" (Mill 1991, p. 67).

That brings us to the second point Mill makes with the previous quote. What makes it particularly striking in Berlin's classification of Mill within his negative conception of freedom is not only that Mill thinks that individual characters, which have real substance and are diverse, must be respected, but also that Mill's view is that our particular characters must be developed. For him:

Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing" (Mill 1991, p. 65).

In the context of this discussion John Skorupski's interpretation of Mill is particularly useful. I have offered above a broader understanding of Mill's notion of 'pursuing our own good in our own way' than Berlin's understanding of it. That was done by pointing out that in Mill's own view freedom from external interference is not an end in itself, but rather a means to the development of one's individual character. Skorupski also makes that point, but through the notion of happiness, when discussing what is for Mill the ultimate good for human beings: "Mill thinks that it consists in happiness"<sup>7</sup> (Skorupski 2006, p. 15). Skorupski points out that for Mill once that is the case human beings need liberty, as they need to be free in order to be happy. Skorupski explains why: "It is because [for Mill] they can develop themselves only when they are free, and because self-development is a condition of the highest forms of happiness" (Skorupski 2006, p. 18).

We can see why, as Skorupski points out, "[i]t is not surprising then that 'free development of individuality' and 'individual spontaneity' are key notions for Mill" (Skorupski 2006, p. 31). What is striking is that Mill is supposed to be the principal exemplar of Berlin's negative conception of freedom, but when one looks to Mill one finds that what he is actually interested in is something which in another context one would think it to be related with positive freedom - he thinks that true freedom centres around the proper development of one's particular character. In fact this notion is also at the core of the account of positive freedom that will be offered by this thesis.

That becomes particularly clear in the way in which Mill endorses the views of Wilhelm Von Humboldt, who is for Mill a philosopher whose

---

<sup>7</sup> For the sake of precision I want to quickly clarify that Skorupski attributes to Mill two 'forms' - as he calls them - of happiness. A lower form which roughly focus on seeking pleasure and avoiding pain, and a higher form consisting roughly in the development of the human being according to his own character. In Skorupski's account, for Mill the good for human beings is essentially associated with the higher form of happiness.

doctrine's meaning were understood only by a few outside of Germany. Mill quotes the following passage from Humboldt:

“‘[T]he end of man ... is the highest and most harmonious development of his powers to a complete and consistent whole'; that, therefore, the object 'towards which every human being must ceaselessly direct his efforts, ... is the individuality of power and development'” (Humboldt quoted in Mill 1991, p. 64).

The fact that Mill endorses a view like the above is deeply problematic from the point of view of Berlin's distinction between negative and positive freedom. When it comes to someone like Mill, in the end the ultimate thing that is wrong about being unfree is the source of control being out of place, and not merely the area of control as Berlin insists.

We can also find misconceptions related to philosophers that Berlin classifies as positive freedom proponents. Berlin traces the tradition of the positive conception of freedom back to Rousseau and Immanuel Kant. But what is interesting about it is that he thinks they have very similar views on freedom (which they don't as we shall see in the coming sessions), because he thinks they have very similar attitudes towards nature. According to Berlin the doctrine of positive freedom:

“... was deeply influenced both by Kant and by Rousseau in the eighteenth century. ... it is a form of secularised Protestant individualism, in which the place of God is taken by the conception of the rational life, and the place of the individual soul ... is replaced by the conception of the individual, endowed with reason, straining to be governed by reason and reason alone, and to depend upon nothing that might deflect or delude him by engaging his irrational nature” (Berlin 1958, p. 185).

To be fair, one can see why Berlin would find something in common between Rousseau and Kant's views. Kant thought that the problem with loss of freedom is being ruled by something external to us, an alien influence of some sort. And effectively there is a whole aspect of Rousseau's views on one's proper development which protects the child precisely from that. But to then conclude that a conception of freedom



derived from Rousseau, like Kant's, can be synthesized in terms of rational self-direction, that there is nothing else to it, seems to me to be missing the mark completely. Basically Berlin places Kant and Rousseau together because both of them have some version of subdivision within the self. So there are some things within the self that even in Rousseau are regarded as alien. But as we have seen in the first section of this chapter, there are also other things that ought to be preserved and developed. Berlin doesn't notice the fundamental difference in their approach. And that, again, raises a question: why, after having traced back the tradition to Rousseau, does Berlin hold such a narrow understanding of Rousseau's conception of freedom? This is a question that I have already placed in this section before, and it is now time to properly answer it.

Let's keep track of things. We have noted three problems with Berlin's account, problems that have already been (in one way or another) individually noted in the literature. They are: there isn't very much of a distinction between Berlin's positive and negative conceptions of freedom; the positive conception isn't really that positive; and there is a slightly odd approach to history. Now the thing that I want to draw attention to is that when we take them all into account we can actually see that the issue is not just that there are these individual problems. Once one sees those problems the obvious question is, why is it like that?

Let's think of something the thinking of which would insure that you would have a negative conception of positive freedom, which of course would also result in no positive-negative type of distinction at all. We have seen that positive freedom is oneself being the source of one's own action. Now, if you thought there was nothing to being the source of one's own action other than nobody else being the source of one's action then you would have a negative conception. So why would one think that there was nothing more to 'being the source of one's own

actions' other than nobody else being the source? You would think that if you had no fuller account of what it would be for an action to come from the self. Now, you would have a fuller account of what it would be for an action to come from the self if you had a rich conception of the self with a range of interesting very particular idiosyncratic features, something like Rousseau's genius. If Berlin had a substantial conception of the self, a non-blank conception of the self, he could have said with Erich Fromm that: "... the full realization of positive freedom ... is based upon the uniqueness and individuality of man" (Fromm 1997, p. x). So the best explanation of the positive not being very positive and there not being much difference between the positive and the negative is that Berlin has the blank conception of the self, because if he had a richer conception of the self there would be an alternative account of what it is for oneself to be the source of one's actions, and he could build a substantive positive conception of freedom from that.

So effectively one is forced into thinking of positive freedom as no more than negative freedom if one can't draw the idea that your actions are flowing from what is particular to you as a self. And the way in which one can't draw on that is by having a blank conception of the self.

MacCallum comes quite close to this point, though he doesn't further pursue it:

"It would be far better to insist that ... the differences, rather than being about what freedom is, are for example about what persons are, and about what can count as an obstacle to or interference with the freedom of persons so conceived<sup>8</sup>" (MacCallum 1967, p. 320).

---

<sup>8</sup> Erich Fromm makes a similar point in the preface of his work 'The Fear of Freedom', when introducing the relation between the notion of freedom and certain issues present in our modern society: "the meaning of freedom for modern man ... can be fully understood only on the basis of an analysis of the whole character structure of modern man" (1997, p. ix).

But the issue is not that Berlin has a blank conception of the self merely. It is that he is blind to any other. And the fact that Berlin is blind to any other conception of the self explains why he can't see the fundamental difference between Rousseau and Kant, why he can't see that Mill is not quite the straightforward person he believes, and why he doesn't talk about Humboldt. Because all those three things have the same explanation, that what is interesting about them is that they are all related to Berlin's failure to acknowledge the importance that they attribute to people's individuality.

Let me make a quick pause to make a point related to Humboldt. Though the relevance of Berlin leaving Humboldt out of the discussion will become clearer as we come to my proposed conception of positive freedom, for the moment let me just point out how odd it is for Berlin to leave Humboldt out of the discussion if one accepts Geuss' interpretation of Berlin. According to Geuss, "Berlin made a highly influential distinction between two concepts or two families of concepts of liberty, a distinction that can be seen as a generalisation from the concrete political conceptions that figure in the answers to Humboldt's two distinct questions<sup>9</sup>" (Geuss 2001, p.89). So Geuss actually traces the source of Berlin's distinction between negative and positive freedom directly to a distinction that Humboldt makes in the context of the political sense of freedom.

In explaining what has gone wrong with Berlin's view we have already seen that we get a genuinely positive conception of positive freedom if

---

<sup>9</sup> Humboldt's two questions are: (a) Who rules? That is, what structures exist in society for exercising political power — how is the government organized — and who actually controls these structures and how?; (b) To which 'objects' (i.e. to which spheres of human life) ought the governmental power to extend its activity and from which ought it to be excluded? According to Geuss, Berlin calls the 'positive' conception of freedom the answer to the first question (a): 'That society is freest that has the most fully developed structure of internal self-rule'. And he calls the 'negative' conception of freedom the answer to the second question (b): 'That society is freest in which the individuals are least, externally obstructed or interfered with by government' (Geuss 2001, p. 89).

we ground it in a rich conception of the self. That permits it to be the case that being oneself the source of one's actions means more than not being interfered with. What it does is it traces one's action to one's own distinctive character. So what we need now is to develop a conception of the self which makes sense of that.

With that in mind, in the following section I will rely on Rousseau, enriched by Mill, to start presenting a distinction between negative and positive freedom which is clearer and cleaner than Berlin's (the next section will be focused on negative freedom). On this view the negative aspects of freedom will all be centred around the notion of external influences. That will be understood in a way as to be consistent with both senses of the term used by Berlin. And the positive side will take into account not just that you are not ruled by other people, but most importantly that there is a positive source of development, a real trajectory which arises from your own nature. And this is exactly the opposition to the blank conception of the person held by Berlin.

So the whole idea of my positive conception will be to use the richer conception of the self with its own genius that we get from Rousseau, complemented by other philosophers, to give us a proper genuinely positive account of positive freedom. Such account will then involve the view that there is something that is particular to the person that needs to be developed. But let me make from the beginning a caveat about that view. It is obvious that the person's genius doesn't include every characteristic which they happen to have because otherwise there is no possibility of development in a good way. Yet it must include quite a rich amount, enough to give some sense of a real character. And there will be different kinds of things that will be reflections of their nature.

### 1.3. An Agent-Particularist Negative Freedom:

Once we reject the blank conception of the self and come into the richer-genius conception we can immediately see certain distinctive aspects of negative freedom that we want to preserve and in the same way certain distinctive aspects of positive freedom that we want to preserve. In this section I will present the features of negative freedom that are special for Agent Particularism, and will proceed later to what is especially important for Agent Particularism in terms of positive freedom.

In this Agent-Particularist negative freedom there are some things which I will keep from Berlin's account of freedom. I will keep his basic distinction between a negative and a positive freedom. I will also keep the essence of his negative freedom; yet that will be only one level of my negative freedom (composed of two levels), the one that understands "external influences" in terms of other people. In effect we could also say that I will preserve something from his positive freedom, because what Berlin calls positive freedom is actually understood in negative terms, i.e. freedom from external influences - but in this case external influences involve internal threats. The second level of negative freedom that I will present also differs from the first by understanding external influences in terms of internal threats.

Though I will keep Berlin's basic distinction of negative and positive freedom, unlike his which is a distinction between two irreconcilable notions mine involves a negative and a positive freedom that not only co-exist but also are two inter-dependent aspects of the same account of freedom. The two aspects will roughly be contrasted as follows: the negative seen as freedom from the cramping of one's particular genius, and the positive seen as freedom to develop for oneself one's particular genius. And we will see later how negative freedom allows the

possibility of positive freedom, while positive freedom informs the value of negative freedom.

### 1.3.1. The First Level of Negative Freedom:

There are distinctive aspects of negative freedom which become vivid and striking once we take on the Agent-Particularist conception of the self. They revolve around protecting the individual genius. Protection is needed because there is a special threat which becomes visible once we have this idea of a rich particular genius. It is basically a kind of inhibition which comes from the forcing of an individual genius to conform to the common, homogenous mode of behaviour found in society.

Since Mill and Rousseau also work with a richer-genius conception of the self and that is precisely what Berlin fails to see about them, he also fails to see that exactly that same threat is a major concern for them<sup>10</sup>. That is clearly reflected on Rousseau's account of a proper education, which is also structured around a negative and a positive education. In fact the kinds of issues that Rousseau is concerned with in education are precisely issues to do with making a person free - first by preserving the natural self from the harmonizing, unifying, pressures from the outside and secondly by letting it grow. So like the Agent-Particularist freedom Rousseau's negative education is centred on the notion of a particular genius, and the focus is in protecting it from external influences. Rousseau offers a negative account of education that prepares the ground for the positive side of education,

---

<sup>10</sup> Though for Berlin Mill is concerned exclusively with negative freedom while Rousseau is concerned exclusively with positive freedom, I will rely on notions from both of them for both my negative and my positive freedom.

which looked from the point of view of freedom is of course problematic for Berlin.

We are now ready to look at the threat anticipated in section 1.1. Rousseau suggests that we come to “esteem each individual but despise the multitude; let him see that all men wear pretty much the same mask, but let him also know that there are faces more beautiful than the mask covering them” (Rousseau 2010, p. 391). This points to a danger coming from the fact that people in some sense adopt very similar attitudes. And as Mill points out it is not as if one adopts, out of one’s own volition, the standard mode of behaviour:

“Society can and does execute its own mandates (...) there needs protection also against the tyranny of the prevailing opinion and feeling; against the tendency of society to impose, by other means than civil penalties, its own ideas and practices as rules of conduct on those who dissent from them” (Mill 1991, pp. 8-9).

Mill is calling attention to the pressure that society exerts on each of its individual members for a uniformity of ways of being. This gives us a new development that follows from Rousseau’s image of people wearing the same mask mentioned above. The problem is, as Mill shows us, not only that people wear the same mask but also that they exert a pressure for everybody else to also use it.

This point anticipates the sharp and fundamental contrast between my kind of Agent-Particularist morality and the morality that comes out of the blank conception of the self. As we have seen in the previous chapter the universalizability theorist wants everybody to behave exactly in the same way in relevantly similar circumstances. But from the perspective of the richer-genius conception of the self this is in fact a kind of unwanted inhibition. It is seen as constraining the individuality of the person by making them conform - become more like what everyone ought to be like. And what makes this contrast in views so

fundamental is that this inhibition coming from the universalizability principle is in a sense the very aim of the kind of morality that comes out of the blank conception of the self, whereas for the Agent Particularist it prevents the person from developing into a moral being. In other words, the inhibition coming from the universalizability principle is itself the threat and removes the very possibility of morality<sup>11</sup>.

The basic thought behind Berlin's view is that you should be able to do what you want provided it doesn't prevent other people from doing what they want. Now, what makes it the case that what is wrong in terms of freedom is stopping people from doing what they want, and the only situation in which it is alright to put what someone wants into question is when their pursuing it would stop other people from doing what they want?

There is a natural answer to this question. There is a general ethical point of view which would justify taking this attitude, and it is not easy to see another way of justifying it. In essence, the view is that wants are beyond evaluation, i.e. they are not the kind of thing that can be criticized rationally<sup>12</sup>. Human beings just happen to have the wants they have and live their lives by trying to get what they want. On this conception the only criticism that could be of a want is that possessing

---

<sup>11</sup> I will come back at this point when I discuss specifically the threat to freedom coming from the blank conception kind of morality.

<sup>12</sup> For a similar view see Bernard Williams' 'Internal and External Reasons'. Williams' view is that an action is rational if it satisfies some pre-existing desire of the agent: "A has a reason to [X] iff A has some desire the satisfaction of which will be served by his [X]-ing" (Williams 1981, p. 101). In this view there is no sense of a desire being rationally criticized. Having the desire in itself counts as a reason to do something. Desires can still be in some sense organized, in such a way that what is rational is for the agent not to satisfy some desire. But this would be because they have some other (more fundamental) desire which would remain unsatisfied by their satisfying the former desire. The view that desires aren't the kind of thing that can be rationally criticized can also be traced back to David Hume. Hume makes a clear distinction between reason and feelings and in effect between reason and morality in such a way that he sees moral judgements as responsive to the agent's feelings of approval or disapproval, which in their turn are seen as not responsive to reason: "... reason alone can never be a motive to any action of the will; and ... it can never oppose passion in the direction of the will" (Hume 2011, p. 359).



it is liable to lead to other people's wants being unsatisfied – a view which is the basis of Berlin's conception of negative freedom. The role of reason in this picture is essentially in enabling human beings to get what they want - there is no suggestion that it might be the task of reason to criticize our wants. This may sometimes involve the ordering of one's desires, and perhaps removing inconsistencies between divergent wants. So if we assume that living in society is one of the fundamental wants for us we would see that certain wants that people may have would make it very difficult for us to live in society. Thus in order to pursue our fundamental want of living in society any wants we may have that contradict it needs to make room for the want of living in society. So what would be needed, this view says, is a kind of negotiation in a community to find some kind of accepted agreement among members of such community. This is the point where laws, contracts, and more relevantly – morality, come into the picture.

When I ask myself 'what would justify Berlin's kind of view?', the only thing which I can see as able to justify it is the view presented above, centred on the idea that wants can't be criticized rationally. Berlin does not himself say this is the reason for it - he does not provide this justification, nor is this in general taken to be the justification. But it is hard to see what other justification there could be.

This is how we can understand, in Berlin's kind of account, the demand for universalizability. In Chapter One I have presented the view that for the universalizability theorist what justifies the demand for universalizability is a demand for consistency. What I am saying here is that in effect for someone like Berlin the demand for consistency that explains his insistence on universalizability is itself explained by a demand for something like the social accommodation of wants which are not themselves rationally accessible.

By contrast the basic thought behind the Agent-Particularist freedom is that one's particular genius needs to be protected from external influence so that it can develop itself of its own accord. Now, posing the same kind of question that we posed to Berlin - What makes it the case that what is wrong in terms of freedom is interfering with one's natural genius and in this way preventing its positive development?

Since I am working with a richer-genius conception of the individual self the idea that a human being 'ought to become a certain kind of person' is understood in terms of a proper development of the person according to their natural genius. The view is that each person ought to be a properly developed version of themselves, and they are only obliged to do something insofar as it is something that their properly developed version could do<sup>13</sup>.

This section on an Agent-Particularist negative freedom is as I said before all about protecting the particular genius. I will show later in the section how the value of negative freedom is informed by positive freedom, in such a way that the possibility of the positive development of the person is only available when they have negative freedom. Since that is the case, the justification of negative freedom on my account comes from the fact that unless you have positive freedom you will be prevented from becoming a moral being.

But before coming into morality per se let's understand this threat in more general terms. The source of Mill's concern with individuals being forced to conform is clear. For him the aim of those social pressures is to: "...fetter the development, and, if possible, prevent the formation, of any individuality not in harmony with its ways, and compel all characters to fashion themselves upon the model of its own" (Mill 1991, pp. 8-9). Like Rousseau, Mill's concern in protecting the particular

---

<sup>13</sup> I will explain this view in detail on Chapter Four.

genius comes essentially as a necessary step towards developing that genius to the fullest of its unique potential.

One way of describing the difference between this first level of negative freedom and the second one is this: we can understand the first level as preventing the internalization of externally inflicted ideas, behaviours and attitudes (everything that could be seen as not grown from the inside), and the second level as dealing with what has already been internalized. So what follows involves the idea of protecting the particular genius in terms of preventing the internalization of what is unwelcome.

When we wonder about how to protect the particular genius in the context of an education in terms of the initial stage of that process, we may conclude with Rousseau that:

“... the first education ought to be purely negative. It consists not at all in teaching virtue or truth but in securing the heart from vice and the mind from error. (...) Without vice, without habit, he would ... become in your hands the wisest of men; and in beginning by doing nothing, you would have worked an educational marvel” (Rousseau 2010, p. 226).

The problem with life in society is that it forces us into socially accepted modes of behaviour: “... there are so many contradictions between the rights of nature and our social laws that one must ... use a great deal of art to prevent social man from being totally artificial” (Rousseau 2010, p. 485). That is why a proper education of the person involves exposing ideas to the individual gradually, only in as much as they are able to evaluate them by themselves. The point of education at this first stage would be not in learning truths but rather activating one’s own capacity to find them out by oneself.

What is helpful for an Agent-Particularist conception of negative freedom in this kind of negative education is that by definition even if that stage of education is fundamentally concerned with itself it does

so while consciously preparing the ground for the development of the child's particular nature. Unlike Berlin's irreconcilable conceptions of negative and positive freedom, this allows us a conception of negative freedom that is in a sense the indispensable other-half of positive freedom.

Negative freedom aims at protecting the uniqueness of the genius, and so "... it is never as important that he learn as that he do nothing in spite of himself" (Rousseau 2010, p. 317). For Rousseau the education starts with preventing the child from developing the habit of relying on external opinions as opposed to developing their own: "What is more useful to him, he takes more seriously; never deviating from this way of evaluating, he grants nothing to opinion" (Rousseau 2010, p. 359).

Once we accept the richer-genius conception of the self there are two points involved in the threat coming from pressures towards uniformity which connect negative and positive freedom. The first is the need to respect one's individual interests, talents, inclinations, and as Mill puts it, the "... liberty of tastes and pursuits; of framing the plan of our life to suit our own character". People are different, and this is reflected in diverse and unique ways in different people. The difference that is integral to each person being the person they are needs to be respected. Once the person is forced to develop in directions that don't concur with their particular genius, in some important sense it is no longer a natural process, and the possibility of positive development of the genius and so positive freedom is prevented. The threat in this case comes from the unacknowledged importance of the particular genius:

"If it were felt that the free development of individuality is one of the leading essentials of well-being (...) there would be no danger that liberty should be undervalued (...). But the evil is, that individual spontaneity is hardly recognized by the common modes of thinking, as having any intrinsic worth, or deserving any regard on its own account" (Mill 1991, p. 63).

The second point that makes this connection between negative and positive freedom is the importance of practicing one's particular capacities in order to be able to develop them. As Rousseau points out: "(f)orced to learn by himself, he uses his reason and not another's; (...). From this constant exercise there ought to result a vigor of mind similar to the vigor given to bodies by work and fatigue" (Rousseau 2010, p. 358).

The two points are related, as it is by respecting one's natural inclinations and talents that their particular capacities will be able to be properly developed. The threat in conforming to the standard opinions and modes of behaviour involves therefore not only forcing the person to neglect their own individuality. That brings along with it a concern with a kind of cramping of one's capacities. Mill illustrates the point:

"... though the customs be both good as customs, and suitable to him, yet to conform to custom, ... does not educate or develop in him any of the qualities which are the distinctive endowment of a human being. The human faculties of perception, judgment, discriminate feeling, mental activity, and even moral preference, are exercised only in making a choice. (...) The faculties are called into no exercise by doing a thing merely because others do it, no more than by believing a thing only because others believe it. If the grounds of an opinion are not conclusive to the person's own reason, his reason cannot be strengthened, but is likely to be weakened, by his adopting it: and if the inducements to an act are not such as are consentaneous to his own feelings and character ... it is so much done towards rendering his feelings and character inert and torpid, instead of active and energetic" (Mill 1991, p. 65).

This first level of negative freedom understands "external influence" in terms of other people, and protecting the particular genius involves preventing the internalization of ways in which we come to repress our natural way of being in favour of meeting the social pressures that fall upon us. We neglect our genuine inclinations. We neglect our own capacities as we develop the habit of relying on external authorities rather than by gradually coming to understand and do things by ourselves. All of those acquired attitudes cause a sense of

disconnectedness with one's own genius, as it is put to the side in favour of us adopting the standard mode of behaviour: "... constrained by our habits, [our natural inclinations] are more or less corrupted by our opinions. Before this corruption they are what I call in us nature" (Rousseau 2010, p.161).

### 1.3.2. The Second Level of Negative Freedom:

Though at the first level of my Agent-Particularist negative freedom I kept much of Berlin's account of negative freedom, the distinct conceptions of the individual self that we work with makes it the case that Berlin's negative conception falls short of providing all the elements that an Agent-Particularist negative freedom needs. The aim of negative freedom for the Agent Particularist is in protecting the individual genius. Though not having to do what other people want you to do is necessary, it is not enough. Since the threat is to one's particular genius, the very idea of 'doing what you want' becomes problematic. If the 'you' in 'doing what you want' doesn't properly refer to your particular genius, doing what you want actually becomes a threat to the freedom of that genius (we will see how as I develop this sub-section) at an internal level.

The threat that comes from the pressures to conform actually encompasses the two senses of the notion of 'external influences' that Berlin uses. At the first level of negative freedom it involves the understanding of external influences in terms of other people. And at this level 'external influences' is understood in terms of influences that are alien to one's genius. In this case it refers to influences that are internalized by us, but that are in effect external to our nature.

The internal level of negative freedom also connects with positive freedom. By repressing the individual genius in favour of conforming to the standard way one prevents the possibility of the proper development of one's genius (as we have seen in the previous subsection). Yet we need to understand what is meant by the idea of one repressing one's individual genius in order to understand clearly how the threat operates at this second level.

Once it is the case that we are coming from a richer-genius conception of the self, we are not understanding a notion like one's "true self" in terms of some very narrow human capacity such as reason like Berlin is. If one takes on the blank conception of the self one is forced to seeing things in some kind of narrow way like this. In the case in which the "true self" is understood in terms of rationality, acting freely becomes acting out of rationality and in opposition to one's feelings. In Kant for instance one is free by resisting what he refers to as one's inclinations, and acting out of reason purely<sup>14</sup>. But once we have a richer conception of the self, there is a distinction to be made between the different kinds of inclinations that we have:

"... would it be reasoning well to conclude, from the fact that it is in man's nature to have passions, that all the passions that we feel in ourselves and see in others are natural? Their source is natural, it is true. But countless alien streams have swollen it. (...) Our natural passions are very limited. They are the instruments of our freedom; they tend to preserve us. All those which subject us and destroy us come from elsewhere" (Rousseau 2010, p. 362-63).

What we need is basically a distinction among our feelings between those feelings which flow out of one's genius and those which are alien to it. Once we have that distinction we can have a view in which the agent, when demanded to act, is able to "...distinguish between the

---

<sup>14</sup> I will look at Kant's view in detail later in the section.

inclinations which come from nature and those which come from opinion" (Rousseau 2010, p. 161).

Rousseau's distinction between the notions of 'love of oneself' (*amour-de-soi-même*) and 'self-love' (*amour-propre*) is particularly useful in this context. In his own words:

"We must not confuse [self-love]<sup>15</sup> with love of oneself. (...) Love of oneself is a natural feeling 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. [Self-love] is only a relative sentiment, artificial, and born in Society, which inclines each 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 honor" (Rousseau 2010, pp. 742-43).

But how does this relate to negative freedom? By relating love of oneself with an original and natural inclination that flows from one's genius, and self-love with inclinations coming from internalized social pressures that are alien to the genius:

The source of our passions, the origin and the principle of all the others, the only one born with man ... is [love of oneself] - a primitive, innate passion, which is anterior to every other, and of which all others are in a sense only modifications. (...) most of these modifications have alien causes without which they would never have come to pass; and these same modifications, far from being advantageous for us, are harmful. They alter the primary goal and are at odds with their own principle. It is then that man finds himself outside of nature and sets himself in contradiction with himself (Rousseau 2010, pp. 362-63).

Love of oneself involves natural and necessary inclinations directed towards one's preservation and wellbeing. Whereas self-love is related to a kind of self-image that we want to project and defend, and involves

---

<sup>15</sup> Different translators give different treatments to the expressions "*amour-propre*" and "*amour-de-soi-même*". In order to avoid confusions I will keep a standard treatment of those expressions throughout, and so will substitute the original expression used by the translator for my suggestion in brackets whenever it is necessary. I will use "love of oneself" whenever I am referring to the natural feeling that reflects one's genius, and will use "self-love" to refer to the feelings that are external to one's true self and in this sense artificial. By doing it in this way I believe I will be doing proper justice to the meanings of the original expressions in French.



inclinations directed towards being ranked as high as possible by other people within society: “[Love of oneself], which regards only ourselves, is contented when our true needs are satisfied. But [self-love], which makes comparisons, is never content and never could be, because this sentiment, preferring ourselves to others, also demands others to prefer us to themselves, which is impossible” (Rousseau 2010, p. 364).

The need for negative freedom here is clear. We have a genuine feeling that causes us to care for ourselves. And the threat is that through the influence of causes that are alien to one’s genius that original feeling towards self-care suffers modifications in the direction of something like the care for one’s social self-image:

“... how much this universal desire for reputation, honours, and preferment, which devours us all ... excites and multiplies the passions (...) this furor to distinguish oneself which keeps us almost always outside of ourselves...” (Rousseau 1754, p. 96).

So we have a basic and innate inclination – love of oneself – *that refers to the care for one’s wellbeing*, and a basic modification of that natural inclination – self-love – which is born out of something like a desire for social reputation. From that basic structure Rousseau devises the whole structure of our feelings in such a way that “... the gentle and affectionate passions are born of [love of oneself], and ... the hateful and irascible passions are born of self-love” (Rousseau 2010, p. 364).

In order to make things clear for my Agent-Particularist negative freedom I will make a distinction between our feelings in the following way: I will refer to the feelings that are originated in the particular genius and so are related to Rousseau’s basic inclination for the love of oneself as ‘sentiments’; and will refer to the feelings originated in the internalization of the social pressures to conform to the model

which are related to Rousseau's self-love as 'passions'<sup>16</sup>. Negative freedom at this second level will involve freedom from one's passions: "Make me free by protecting me against those of my passions which do violence to me, prevent me from being their slave" (Rousseau 2010, p.495).

The image of being enslaved by one's passions brings us back to the question of the source of control. When one's passions take over and are in control the agent becomes passive in some important sense. They are blindly led into action by their passions. Having negative freedom at this level can therefore be understood as refraining from acting under the control of one's passions; acting instead out of one's natural character: "... I sense perfectly within myself when I do what I wanted to do or when all I am doing is giving way to my passions" (Rousseau 2010, p. 441).

From the point of view of the protection of the particular genius, the threat is especially present in a situation in which one accustoms oneself to behave according to social conventions and the opinions of others. Overtime we develop the habit "... to show oneself as different from what one, in fact, [is]. Being and appearing [become] two entirely different things..." (Rousseau 1754, p. 90). Once we accept the view that the possibility of positive freedom can only occur if one behaves out of their natural genius (for reasons already explained), the habit of 'showing oneself as different from what one in fact is' becomes a serious threat to positive freedom: "... habit gives us a second nature

---

<sup>16</sup> Rousseau also uses both terms – sentiments and passions. Though the general sense that he gives to the terms is roughly the same that I am giving to them here I still want to say that it is a distinction that I am using in my own way. Not only because I am putting it into the context of Agent Particularism, but also and mainly because sometimes Rousseau's use of those terms is confusing in such a way that it is not really clear that the distinction he has in mind is as sharp as the one I am making.

that we substitute for the first to such an extent that none of us knows this first nature any more<sup>17</sup>" (Rousseau 2010, p.294).

I will now move into a special case of the threat to individual freedom that we are dealing with: the pressures to conform one's behaviour to a given socially accepted view of morality. The main reason (though there are others as we will see) for me to treat morality as a special case of the pressures to conform is that this will allow us to see clearly how views of morality that come out of the blank conception of the self, even when proposed with the best of intentions, are actually a very serious threat to the very possibility of morality as conceived from the richer-genius conception of the self.

### 1.3.3. The Threat Coming From the Blank Conception's Morality:

Before going into contrasting my kind of Agent Particularism with moralities that come out of the blank conception, there is something to be said about negative freedom in terms of pressures to conform to an externally imposed morality. The reason is there are certain distinctive features of this kind of threat. So let's look at what is it that is particularly bad about them.

The first of these features comes from the fact that moral pressures are seen as having a certain privileged legitimacy, and this supposed legitimacy brings extra concern in terms of the threat to freedom. This idea anticipates a point that will be fully developed when we get to our

---

<sup>17</sup> Rousseau uses the notion of a 'second nature' in the specific context of the child's appetite. He makes the point that the child is forced to eat at the moment that others decide (as opposed to when he is actually hungry), and eat what they decide he should need (as opposed to eating what his body naturally asks for). Over time that externally imposed pattern becomes his second nature, suppressing the child's natural appetite. I generalize the meaning of 'second nature', applying it to every situation in which one's behavior is being controlled by one's passions as opposed to flowing naturally out of one's original and particular nature.

discussion on Kant later in this section. As we will see for Kant moral demands present themselves as having authority of a special kind, which the agent must listen to. The authority is essentially the authority of the categorical imperative. In this view the whole point of a moral demand is that you must do what it says independently of what you want to do. There is an unconditionality of the demand, and for that reason you are supposed to overcome what you might want. That makes it a worrying kind of way in which one might insist on conformity for two reasons: firstly this kind of pressure for uniformity presents itself as not to be representing any interest. In Kant's words: "[The categorical imperative] precisely for the idea of universal legislation, [is] founded on no interest" (Kant 2012, p. 44); and secondly it anticipates a reluctance the agent might feel to conform with it, i.e. the agent might feel disinclined to do it and yet the moral demand comes loaded with the idea that this disinclination must not be taken into account.

The idea here is that the insidious path and the forcefulness of moral pressures for uniformity makes them particularly dangerous. When we put that into the context of the Agent-Particularist conception of negative freedom we can see that: a) from the perspective of the first level of negative freedom, i.e. pressures coming from others, there is potentially an increase in the aggressiveness with which the pressure is done. The reason for this is that from within a blank conception of the self, in general, "[t]he practical principle which guides [people] to their opinions on the regulation of human conduct, is the feeling in each person's mind that everybody should be required to act as he ... would like them to act" (Mill 1991, p. 10), and once one has the supposed legitimacy of being morally right they in many occasions may abandon the restraint and prudence that they usually have with explicitly interfering with someone else's life; and b) from the perspective of the internal level of negative freedom passions coming

from moral pressures to conform such as guilt, given the particular status that morality has in society, tend to have a stronger grip on the agent.

There is also another kind of distinctive feature of the threat coming from the blank conception of the self's morality. It reveals itself when we bring to the table the idea that the value of negative freedom is informed by positive freedom. From a richer-genius conception of the self there is a need to protect one's genius from external interference (negative freedom). That need is justified by the requirement to develop into an ideal version of oneself (positive Freedom). And there is a treat coming from the fact that when the agent is operating from an internalized (yet externally imposed) system of morality, the agent's proper development is prevented because that artificial morality stops them from practicing their own genuine moral sensitivity: "Unruly passions inspire evil actions, but evil precepts corrupt the reason itself and cut off the possibility of a return to virtue" (Rousseau 1968, p. 80).

This point puts Agent Particularism in sharp, in some sense dramatic, contrast with moralities that come from the blank conception of the self. In order to illustrate what I mean I will show how this point works in the case of two types of moral conception that accept the universalizability principle: Thomas Scanlon's and Kant's. I will start with the former.

#### 1.3.3.1. An Agent-Particularist Critique of Scanlon's Contractarianism:

I first introduced Scanlon in Chapter One, while discussing the universalizability theorist's motivation for insisting on a requirement of

consistency in morality even though their view is that there are no truths in morality. The reason is that someone like Scanlon sees morality as nothing but a system that we invented to help guide society in a good direction. Scanlon is working with a blank conception of the self and therefore thinks that for that we need principles that anybody rational could be brought to accept without being forced. But as I showed in Chapter One it is perfectly possible to keep consistency in morality without such principles. I would like to now look in more detail at Scanlon's blank conception kind of morality.

Since Scanlon is operating with a blank conception of the self it is no surprise to find that his moral account assumes the universalizability thesis as I understand it:

"... let G be the set of factors, whatever they may be, in virtue of which Jane takes herself to have reason to help her neighbor. Since she accepts the judgment that, given G, she has reason to help her neighbor, Jane is also committed to the view that anyone else who stands in the relation described by G to someone in need of help has reason to provide it. This is an instance of what I will call the universality of reason judgments" (Scanlon 1998, p. 73).

In other words, everyone ought to behave in the same way in relevantly similar circumstances. The central idea in Scanlon's moral account is the view that moral judgements are:

"... judgments about what would be permitted by principles that could not reasonably be rejected, by people who were moved to find principles for the general regulation of behavior that others, similarly motivated, could not reasonably reject" (Scanlon 1998, p. 4).

Scanlon's view is that when we find ourselves in situations in which we are under some moral obligation what we need to do is to act according to principles that no one could rationally reject – provided they are also concerned with acting according to principles that no one could rationally reject. Scanlon essentially thinks that morality involves us striving towards finding some kind of agreement within a community

in terms of which reasons would justify the judgment that each given action is wrong or right. And the important thing is that given the universalizability principle this means striving for rules that apply to everyone within that community equally, which is in effect a striving towards a uniformity of behaviour:

“... an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement” (Scanlon 1998, p. 153).

Scanlon insists that his account includes the idea of unforced general agreement. But since he is working with a blank conception of the self he fails to acknowledge that the problem is not in people finding agreement, but in this agreement being about a uniformity of behaviour. Because of his blank conception he does not see the insidious character of the threat. Rousseau and Mill, as we saw in detail earlier, have a different view.

Scanlon thinks that his proposed aim of striving towards rational agreement is necessary for us to live with each other in good terms. Ironically, for Mill such kind of agreement is precisely the threat:

“There is a limit to the legitimate interference of collective opinion with individual independence: and to find that limit, and maintain it against encroachment, is as indispensable to a good condition of human affairs, as protection against political despotism” (Mill 1991, pp. 8-9).

We saw in the previous subsection problems that the agent’s particular genius faces when confronted with a morality that accepts the universalizability thesis. Though those problems can involve for the most part a subtle kind of pressure, this pressure can be more relentless and penetrating than less subtle ones. Scanlon’s kind of account tends to leave no room for diversity of behaviours. It is relevant that Scanlon not merely accepts the universalizability principle,

but he also is explicit in his view that a certain concern with other people's conformity (or lack of it) to the socially accepted standard model of moral behaviour is an integral part of a community's moral life. What is important for us is that he justifies this concern precisely with the universalizability principle:

"... there is fundamentally no question of why we should be concerned with the reasons that other people have. We must be so concerned, insofar as we take ourselves to have any reasons at all, since any judgment about our own reasons entails claims about the reasons that others have or would have in certain circumstances" (Scanlon 1998, p. 73).

Again, Scanlon would want to insist that his account aims for an unforced general agreement, stressing the idea of the unforced. In fact, he goes as far as saying that:

"... the reason we have to want to be able to justify our actions to others on grounds that they (if similarly motivated) could not reasonably reject, must be distinguished from the reasons we often have for wanting to be able to justify our actions to others on grounds that they actually do or will accept. ... the appeal of actual agreement cannot be the motivational basis of morality..." (Scanlon 1998, p. 74).

So Scanlon distinguishes between the idea of acting on principles that no one could reasonably reject from the idea of acting on the basis of actual agreement reached by people. And he explains his concern in so doing in terms of leaving room for disagreement. That reinforces the view that his aim is for an unforced agreement. Yet, his own description of the process of selection of the features of a situation that are to be taken as morally relevant, which in their turn determine the reasons that will justify an action, probably makes my point better than anything that I could say at this point:

"There are good reasons to expect people's judgments about reasons to be framed in terms that others around them not only could but do actually understand and use. Since we acquire the concepts involved in such judgments chiefly by imitating others, it is to be expected that in our process of selecting, from among the range of features and distinctions which might be noticed, those to which reason-giving significance is to be attached, we will



generally settle on ones that others around us already recognize and see as important" (Scanlon 1998, p. 74).

Scanlon's account of guilt illustrates well how his account of morality represents a threat to the particular genius in both levels of negative freedom. In terms of the first level it makes the external pressure towards the uniformity of behaviour legitimate. He wants to say that the attitude that people have towards someone that doesn't behave according to what is socially agreed shows how guilt also involves that same kind of reproach. And in the middle of discussing this point he says:

"To see the special force of the kind of self-reproach that guilt ... involves ... consider first the significance, for other people, of the moral criticism ... . If an action is blameworthy, then the agent has either failed to take account of or knowingly acted contrary to a reason that should, according to any principles that no one could reasonably reject, have counted against his action. ... [This is] what makes it appropriate for the person who was wronged to feel resentment rather than merely anger and dismay. Similarly, it is this violation of the requirements of justifiability to others that makes it appropriate for a third party to react with' indignation rather than merely dismay or pity for the victim" (Scanlon 1998, p. 271).

On an internal level reproach of the same kind is due according to Scanlon as in both cases the issue is exactly the same, i.e. the lack of compliance with the socially agreed principle:

"Guilt requires negative self-evaluation of a particular kind, which I will call self-reproach ... taking one's rational self-governance to have been faulty, and recognizing that some judgment-sensitive attitude must be modified or taken back" (Scanlon 1998, p. 270).

Scanlon wants to justify his demand for the social agreement of moral principles in people's "... value as rational creatures" (Scanlon 1998, p. 270). I will come to Kant in a moment, but will anticipate here that Kant's demand for respecting people as ends rather than means also comes from their value as rational creatures. But unlike Kant who

grounds his demand for respecting the moral law in rationality itself, Scanlon explains the demand involved in his kind of morality in terms of the "... importance of standing in a certain relation to others" (Scanlon 1998, pp. 177-8). As I pointed out in Chapter One, the motivation behind Scanlon's moral account is his feeling that we need some kind of system that would ensure that society is guided in a good way: "... when we look carefully at the sense of loss occasioned by charges of injustice and immorality we see it as reflecting our awareness of the importance for us of being 'in unity with our fellow creatures'" (Scanlon 1998, p. 163).

The core notion for Scanlon is actually that of an unforced agreement, not rationality as Kant has it. Scanlon thinks that moral judgments should conform to principles that other people couldn't reasonably reject if they are also trying to conform to principles that others couldn't reasonably reject. This is how he introduces rationality to his account. But the point about the offering of reasons is to tell us about the manner in which we reach unforced agreements rather than it being about rationality.

Now let's ask ourselves what would justify the view that morality is about reaching an unforced agreement with others, based on principles that they couldn't reasonably reject? As in Berlin's case, the only justification I can see for Scanlon's moral account is the view that desires can't be rationally criticized. To recapitulate I have said that for Berlin the demand for consistency that explains his insistence on universalizability is itself explained by a demand for the social accommodation of wants which can't themselves be rationally criticized. But at the same time those wants can't all be encouraged or accommodated if we are to live together. This brings us to a demand to accommodate different interests in some kind of negotiation in a community - Scanlon goes as far as explicitly accepting that according to his account "... it would be unreasonable to give the interests of

others no weight in deciding which principles to accept" (Scanlon 1998, p. 192). Morality is then presented as the solution for the conflicting wants that different people have, as a code that we can live by and that will generate rules of conduct which apply to everybody no matter who they are. Scanlon's kind of conception of morality, involving the kind of justification that he gives for insisting in universalizability, is only itself a reflection of a thought about desires as being arbitrary in the relevant sense, i.e. not having any justification in themselves.

Scanlon says that "[he] originally identified the motivational basis [of his account] as a desire to act in a way that can be justified to others..." (Scanlon 1998, p. 7). But he formally renounced that view, claiming that the motivation for acting on principles that no one could reasonably reject can be wholly explained in terms of reasons. The problem is that even though Scanlon has formally renounced his early view of desire, the position actually still relies on it. According to Scanlon the aim of his kind of morality is in "... finding principles that others, *insofar as they too have this aim*, could not reasonably reject" (Scanlon 1998, p. 191). Now the question is, why does he need to add the thought that 'insofar as they too have this aim'? Adding this point means that we have to make it a condition that others have the same goal of acting on these principles. Why do we need that condition? Because we can't require that the goal itself is rational; requiring that the goal itself is rational would be like saying that desires are rational, but the whole point that makes morality necessary in this kind of view is that they are not. What that extra clause does is make it the case that whether it is rational for one to be acting on principles that no one could rationally reject depends on what they want to do. The goal of acting on principles no one could rationally reject is not a goal that they can have reason to have, as you can't in general have reasons for goals on this view. So even the goal of acting on principles that no one could rationally reject has no rational justification. All that can be done is

make the morality valid only among the people that already have the goal of consenting to those principles. There is nothing in the account to say that anyone should consent.

#### 1.3.3.2. An Agent-Particularist Critique of Kant's Account of Freedom:

At the beginning of the sub-section on the second level of my account of Agent-Particularist negative freedom I briefly distinguished my view from Kant's by pointing out the different treatment that the notion of inclination receives in the two views. I want to explore this point in more detail now. According to my Agent Particularism it is absolutely essential that we make a distinction between different kinds of inclinations. While sentiments are genuine and direct expressions of one's self, passions are alien influences. Since in order to properly develop one's particular character one ought to practise it, and one can only practise it if one acts according to one's sentiments, it is essential that one behaves according to one's sentiments. Therefore for my Agent Particularism repressing those sentiments invites a disaster.

But Kant is working with a blank conception of the self, and so as Thomas Hill puts it he "... urges us not to value a person's individuality but rather something which he has in common with others ..." (Hill 1985, p. 43). For Kant as we shall see in detail later what is essential to being an individual person is not only the same for every one, it is also a very narrow capacity of the human being, i.e. reason. In fact the same capacity is what is essential not merely for all persons, but rather for all rational creatures. For this reason Kant insists that:

"... the ground of an obligation, must carry with it absolute necessity; ... not just hold for human beings only, as if other rational beings did not have to heed it ...; [it] must not be sought in the nature of the human being, or in the

circumstances of the world in which he is placed, but a priori solely in concepts of pure reason" (Kant 2012, p. 5).

The same capacity, reason, is what is essential to every rational being, and absolutely everything else about them is regarded as alien to their true self. So if in Kant inclinations are seen as alien forces, one's inclinations include not only what is external to their genius, but rather everything about a person's character should also be excluded as mere inclination:

"Understanding, wit, judgment, and whatever else the talents of the mind may be called, or confidence, resolve, and persistency of intent, as qualities of temperament, are no doubt in many respects good and desirable; but they can also be extremely evil and harmful if the will that is to make use of these gifts of nature, and whose distinctive constitution is therefore called character, is not good" (Kant 2012, p. 9).

Kant thinks that everything in the empirical world is determined by nature, a chain of causes and effects put in motion by causal laws of nature. Even living beings that are not gifted with rationality such as animals are merely passively following their instincts. As we exist in the empirical world and are affected by sensation we could have had the same fate, being a mere passive instrument of our inclinations.

But we happen to have reason, which happens to be the only thing that is essential to each one of us. So, Kant thinks, what it takes for a person to act free from the influence of alien forces is that their actions are not determined by inclinations. Otherwise, as Christine Korsgaard explains, they are just being one more passive element of the blind chain of causes and effects determined by nature:

"... inclinations, in the world of phenomena, are completely determined by natural forces, by the nexus of causal laws. So such a will becomes a mere conduit for natural forces. The person who acts from [inclination] is in a sense not actively willing at all, but simply allowing her self to be controlled by the passive part of her nature, which in turn is controlled by all of nature" (Korsgaard 1989, p. 32).

Kant makes a contrast between actions done out of inclination and actions that are independent from the influence of inclinations. Since actions done out of inclination always aim at a certain end the worth of those actions is conditioned by the attainment of that end. But the worth of the action that is independent from inclination (the action of the rational being that acts morally) is not conditioned in the same way, i.e. it has absolute worth:

"A good will is good not because of what it effects, or accomplishes, not because of its fitness to attain some intended end, but good just by its willing, i.e. in itself; and, considered by itself, it is to be esteemed beyond compare much higher than anything that could ever be brought about by it in favor of some inclination, and indeed, if you will, the sum of all inclinations. ... the absolute worth of a mere will..." (Kant 2012, p. 10).

Kant associates what he calls hypothetical imperatives with actions to which he attributes a conditional worth, while he associates what he calls categorical imperatives with actions to which he attributes an absolute worth. Since, according to Kant, the actions of the free will of the rational being are supposed to be independent from inclinations they are supposed to be motivated by categorical imperatives as opposed to hypothetical ones. And if the actions of the free will are categorical imperatives they are not determined by causal laws of nature. This brings us to a point that I will come back to in more detail at the end of my criticism on Kant. For the moment what is important is that this gives us a clearer sense of Kant's connection between the action that is free and the action that is good. Instead of the actions of the free rational will being conditioned by the attainment of some specific end previously determined by nature, and in that sense making no difference to the world as nature determined it (such as happens by actions governed by hypothetical imperatives), their actions have the power to as it were change the sequence of events and in that way

make a difference to the world (such as happens by actions governed by categorical imperatives). According to Korsgaard Kant's view is that:

"[I]f you will in accordance with the moral law, you ... actually contribute - we might say to the rational, as opposed to the merely natural, ordering of the sensible world. The choice of the moral maxim ... [is] a choice to use your active powers to make a difference in the world" (Korsgaard 1989, p. 33).

Korsgaard introduces in the quote above a new notion from the Kantian account, namely the moral law. So let's look at how exactly this notion comes in. As we will see in more detail later Kant thinks that the rational being is at least in some sense free. As rational they are bound to act according to some freely adopted maxim, which can be of two kinds. It can be a maxim grounded in some inclination, conditioned by the attainment of the end which is the object of the inclination – a hypothetical imperative with no intrinsic worth. And the only other kind of maxim one can adopt when acting is a maxim that has an intrinsic value independent of anyone recognizing that value, which according to Kant is the moral maxim. This maxim, Kant thinks, is grounded not on any subjective end but rather on an objective moral law – which turns out to be the categorical imperative.

Since Kant's version of one's true self comes down to rationality, which is the same for everyone, he wants to further say that the moral maxim is grounded in some law of rationality itself. Because of his blank conception of the self, Kant can't help but to conceive this law as taking the universalizable form:

"But what kind of law can that possibly be, the representation of which even without regard for the effect expected from it - must determine the will for it to be called good absolutely and without limitation? Since I have robbed the will of all impulses that could arise for it from following some particular law, nothing remains but as such the universal conformity of actions with law, which alone is to serve the will as its principle, i.e. I ought never to proceed except in such a way that I could also will that my maxim should become a universal law" (Kant 2012, p. 17).

The demand for universalizability in Kant takes a shape of its own. He formulates his version of the universalizability principle, with a few variations thorough his writings, as: "act only according to that maxim through which you can at the same time will that it become a universal law" (Kant 2012, p. 34).

Like the traditional universalizability theorist, Kant also brings in universalizability with a concern for consistency in mind, as Henry Allison implies:

"... in claiming that one's reason for acting in a certain way is a 'good' in the sense of justifying reason, one is, implicitly at least, assuming its appropriateness for all rational beings. The intuition behind this is simply that if reason R justifies my x-ing in circumstances C, then it must also justify the x-ing of any other agent in similar circumstances" (Allison 1990, pp. 204-5).

So Kant's formulation of the universalizability principle does entail that who the agent is makes no difference. But the basic challenge that Kant has in mind when he starts introducing the formula of the universal law involves something more than that as well. When it comes to the actual application of the law Kant wants to also suggest a kind of inconsistency found in not acting in an universalizable way coming from a basic challenge that is something like "what if everybody did that?". One of Kant's examples is the case of promise keeping: "Let the question be, e.g., may I not, when I am in trouble, make a promise with the intention not to keep it?" (Kant 2012, p. 17).

So let's imagine a situation in which you decide to break a promise. But you decide to do so knowing that you are going to be an exception. You will be able to break the promise and get your benefit out of it without threatening the general institution of promising. The way in which the standard universalizability reading takes this is 'what if someone else did that?', 'what if it was not you but somebody else,



what would you think?'. The question behind this point is something like 'what is so special about you that it is ok for you to break the promise but not for them?'. But actually what Kant is concerned with is something like: 'what if everybody break promises at the same time?'. So it is not about what if other people were like you in breaking promises when they are the only one who is doing it, but rather what if everyone did it at the same time? And the answer is the whole institution of promise keeping would collapse. The idea here is that when I ask myself whether if 'breaking my promise' could become a universal law – thus applying to everyone with no exceptions:

"I soon become aware that I could indeed will the lie, but by no means a universal law to lie; for according to such a law there would actually be no promise at all, (...) my maxim, as soon as it were made a universal law, would have to destroy itself" (Kant 2012, p.18).

What Kant thinks this shows is that effectively you are just riding on other people's back in a case like this:

"... someone who has it in mind to make a lying promise to others will see at once that he wants to make use of another human being merely as a means who does not at the same time contain in himself the end" (Kant 2012, p. 42).

This is a kind of anticipation of a point that we will come to in detail later, revolving around the idea of treating other people merely as means. For the moment what is important is noting that according to Kant's view if you break promises you are just treating all those other people who are keeping their promises just as a device for maintaining a system that you now take the benefit of. Kant's point suggests that you are basically exploiting their conformity to the moral law. This is the key thought that gives Kant's version of the universalizability principle its distinctive shape.

There is still another feature of Kant's version of universalizability that we need to look at. Kant's kind of morality is particularly threatening from the perspective of Agent Particularism. The agent is supposed to detach themselves from everything that is particular to them, including genuine inclinations that tell them to do otherwise – making this kind of external pressure for uniformity harder to resist. On top of that the demands coming from Kant's moral law are also supposed to be unconditional, admitting of no exceptions:

“... the concept of morality ... is so extensive in its significance that it must hold not merely for human beings but for all rational beings as such, not merely under contingent conditions and with exceptions, but with absolute necessity” (Kant 2012, p. 23).

Kant believes that the call of reason is simply an absolute call, one to be followed if you can. And as it happens human beings can. The motive behind following the call of reason on this view is just the fact that it is the demand of reason. There is no further motive than that. And the demand of reason is absolute on any creature that can hear it. So for instance the reason why it applies to us rather than other animals is that we can hear it whereas other animals can't. Kant is here relying on a principle of the form '*ought* implies *can*', i.e. you can only be under an obligation to do something if you can do it: “... duty commands nothing but what we can do” (Kant 2001, p. 92). Now once Kant has accepted this principle an issue emerges: it is unclear, to say the least, what grounds his conception of what people can do. When we put together the thoughts that moral demands are absolute to anyone who can hear them, and that these are universal demands that apply equally to everybody, we get as a result the idea that everybody can do the same thing effectively. And that is exactly what is not obvious, because each person has their own particular genius. The problem is that Kant starts from an assumption of the blank conception of the self without offering any kind of justification for this view.

We can at this point see a clear contrast between Agent Particularism and Kant's ethics. Both views accept the *ought*-implies-*can* principle - I introduced my Agent-Particularist application of the *ought*-implies-*can* principle principle and its result in the final section of Chapter One, and will briefly restate the point in order to clarify the contrast I have in mind. Kant thinks that everybody can do the same thing, because everybody is essentially the same - his universalizability principle works from this (ungrounded) assumption. For Agent Particularism the person's particular genius restricts the kinds of things they can do while still being that person. So when the Agent Particularist plugs in the principle that '*ought* implies *can*', they can then see that what someone can be required to do has to fall within the range of things they can do while being that person. And it is obviously possible for something to fall within the range of things that one person can do and not within the range of things that another person can do - which justifies the different moral demands on them in relevantly similar circumstances. As I showed in Chapter One there is no problem of consistency in this view as Kant (and the universalist in general) supposes there would be: it is still the case that reasons apply equally whenever we have the same relevant features involved in two different cases.

In fact there are problems with Kant's conception which go beyond its merely being a (ungrounded) version of the blank conception of the self. The first of these problems involves his understanding of what it is to be a person. Specifically I will concern myself with the question 'what kind of account of being an individual person can be given by Kant's theory?'. The second problem involves Kant's account of autonomy, and in this case I will concern myself with the question 'who or what is in fact autonomous in Kant's theory?'. And the third problem involves his account of freedom. As we will see later, these first two problems bring us to the consequence that in effect Kant's theory

leaves the agent with neither negative nor positive freedom. So let's take a look at these problems in detail.

Kant believes that everything in the empirical world works according to causal laws of some kind. And the essence of Kantian freedom is in the idea of being independent from the influence of causal laws of nature, independent from the determination by the empirical/sensuous world (in a contrast for example with other animals that are in a sense passively obeying their instincts). But human beings too, as part of the empirical world, are susceptible through their inclinations to the influence of the same causal laws of nature that non-rational creatures are; yet as rational beings endowed with a will they have something that non-rational creatures lack, i.e. the capacity to act according to rational principles. Given this capacity, human beings have the possibility of acting independently of causal laws of nature:

"Every thing in nature works according to laws. Only a rational being has the capacity to act according to the representation of laws, i.e. according to principles, or a will" (Kant 2012, p. 26).

We need to be clear about what is implied by this basic point in Kant's understanding of the freedom of human beings. In particular we need to be clear about the very serious implications involved in Kant's idea of 'freedom from causal laws of nature'. This idea involves the notions of 'true self' and 'alien influences' in such a way that alien is understood as anything which is determined by empirical law. Now since his view is that everything apart from the will (the true self) is determined by empirical law, and desires and character are seen by Kant as merely empirical determinants of one's behaviour, desires and character are understood as alien.

I will come back to explain this point in detail in a moment, but first I want to bring attention to another implication involved in Kant's idea of freedom from empirical laws. Kant's position depends upon a kind

of restriction on the scope of reason so that desires can't be within the scope of reason. As a reminder, this is the same kind of view that I have attributed to Berlin in my criticism of his account of freedom. This point anticipates a clear point of distinction between my account and Kant's, which I will properly introduce in the coming subsection. Essentially while Kant thinks desires can't be rational my view is that they will always be rational in some sense – sometimes they will involve good reasons (they will be properly rational) and sometimes they will involve bad reasons (they will be irrational). This is connected to the broader issue of our different treatments of the notion of inclinations – while Kant makes no distinction between different kinds of inclinations I distinguish between our own genuine inclinations and inclinations that are alien to one's particular genius. I will explain this point carefully later; I will now come back to the point I made in the previous paragraph: the grave implications of Kant's conception of freedom from empirical laws, in particular Kant's understanding of desires and character as alien influences.

At the beginning of my criticism on Kant I have noted that Kant rules out as mere inclination absolutely everything about a person's particular character<sup>18</sup>. As we will see at the end of this section, if inclinations included just desires we would already be entitled to claim with Hill that there is an "unfortunate [sense], in which Kant regarded even the agent's own desires as 'alien'" (Hill 1985, p. 109). But the problem in fact gets much worse, because 'inclinations' includes not only their desires, but also things like their character, talents, deepest

---

<sup>18</sup> This move brings up problems of its own. In the next chapter I will show how Agent Particularism is a kind of virtue ethics, and in that context it is relevant that the things which Kant says are of no relevance include the virtues effectively. The thought is that character is irrelevant, but virtues are in fact virtues of character – what you have to do is to develop a character of a certain kind, and Kant rules out everything about someone's character as mere inclination.

aspirations, personality and particular style. Below we have yet another of Kant's formulations of this point:

"Skill and diligence in work have a market price; wit, lively imagination, and humor have a fancy price; by contrast, fidelity in promising and benevolence from principles (not from instinct) have an inner worth" (Kant 2012, p. 47).

A moment's reflection on what exactly is ruled out by being placed on the side of inclination by Kant reveals that everything which we might think is distinctive of any individual person will count as being on the inclination side. All these features of character that Kant rejects are supposed to be merely contingent features of human beings, features that they just happen to find themselves with in the circumstances. Now since everything that might distinguish a particular person has now been put on the side of inclination, is seen as alien, we lose the idea of individual personal identity in human beings. So the drastic consequence of the Kantian freedom from empirical law is that though there is within this picture some sense of human identity, there is no sense at all of what it is to be a particular person. I will come back to this point at a later stage of this criticism on Kant.

My other question is, who or what is in fact autonomous in Kant's theory? Let's now turn to this problem. For Kant one's empirical, sensuous, physical self that engages with the empirical/sensuous world is just contingent and thus not really essential to them. But as I have mentioned at the beginning of this section this doesn't mean Kant thinks there is nothing essential to human beings. What is essential to them in Kant's view, their true self, is the rational will:

"[A] human being presumes for himself a will that lets nothing belonging to his desires and inclinations be put on its account (...) pure reason independently of sensibility ... is the actual self" (Kant 2012, p. 66).

For Kant, saying that human beings are free is the same as saying that their rational will is free. And he thinks that reason can cause human beings to act; so while in non-rational creatures their instincts cause them to act, in human beings besides inclinations that can possibly cause them to act (I will clarify soon what Kant means by inclinations causing people to act) they also have reason as a possible cause of their actions:

"A will is a kind of causality of living beings in so far as they are rational, and freedom would be that property of such a causality, as it can be efficient independently of alien causes determining it; just as natural necessity is the property of the causality of all non-rational beings to be determined to activity by the influence of alien causes" (Kant, 2012, p.56).

The human will then, because it is rational, has the capacity to cause the human being to action, and if it is free it will do so independently of any kind of influence coming from the empirical world. The key point about this idea is, as Hill indicates, in who/what is the source/cause of the action:

"[T]o attribute an event to a person, thing, or prior happening as its cause (source, or author) we need some appropriate connection between the event and the alleged source, something to warrant saying the event occurred in some sense because of the source. Often the connection is an observable regularity between types of events, and then the 'because' is an empirical causal one; but in the case of human action, on Kant's view, the connection is between the event and a person's beliefs and policy commitments, and then the 'because' must be of a different type" (Hill 1985, p. 108).

As we have seen Kant sees the will as a kind of causality in rational beings. But Kant wants to make a distinction between two different kinds of 'causality', even though both senses refer equally to the source of the action. The first sense, the one he attributes to empirical causes, implies the idea of a physical/instinctual impulse as irresistible to creatures whose action are determined by those causal laws (like the instincts of other animals). This is a kind of causality that involves a passivity from the agent. So in this sense the (non-rational) agent is

moved into action by instinct (which it can't but obey), which in its turn has been directly determined by some causal law of nature. Here, the source of the action is understood as alien to the agent (i.e. external to their self as reason). This kind of causality does not apply to human beings.

The second sense of causality, the one Kant attributes to rationality, doesn't involve in his view being 'moved into action' like the first one. In this second case the rational agent keeps themselves active, which for Kant means that they are not being caused to act by some external force but rather are causing themselves to act in a certain way. But this doesn't mean that the source of the action will always be reason (the true self) – and this is where the distinction between the free and the unfree action comes in. His view is that when the rational will acts freely it acts out of 'pure reason', which means that the ground of the action is just rationality and nothing else (nothing empirical related to the person or the world). Kant claims that the rational will is in fact free in this way, which turns out to mean that the person is autonomous. But when we look closely to what he has in mind when he says that the person is autonomous, the claim that the grounds of the free action is pure reason has drastic consequences for his conception of autonomy. Those consequences are what I want to bring out now.

At the same time that he claims that rational beings are in fact free Kant doesn't deny they sometimes make choices in order to fulfil their desires, but he insists that even in those cases they are not merely being passively lead by inclination. Korsgaard explains the point in the following way: "You may of course choose to act on a desire, but insofar as you take the act to be yours, you think you have made it your maxim to act on this desire" (Korsgaard 1989, p. 26).



This issue brings us back to Kant's distinction between hypothetical and categorical imperatives. Kant thinks that the rational being (with a will) can't but act as if they were free, which implies that even if in effect their action has been determined (in the second sense of causality presented earlier) by inclination they will still be seen by Kant as acting rationally, even if just with a kind of imperfect reasoning. According to Hill Kant's view is that:

"[W]hen a will is "determined" by inclination ... this means the agent's policy or guiding idea was some hypothetical imperative concerning the means to satisfy the inclination. (...) the agent's chosen policy makes a certain causal connection, or strictly his belief in a certain causal connection, be a decisive (or "determining") factor for what he does. His full rationale (not a causal event) is: 'I shall do whatever is necessary as a means to satisfy my inclination B; A is a necessary means to satisfy B; hence I shall do A'. The agent has let the causal law, or strictly his idea of the causal law, between that sort of means and end be the dominant or 'determining' factor in his choice" (Hill 1985, p. 109).

But even if the rational being is free in the sense that they are never really determined (in the sense that non-rational creatures are) by any kind of force coming from the empirical world, when they act from inclination (according to some hypothetical imperative) the act is not properly free in Kant's view, as Allison explains: "the agent's choice would be ineluctably heteronomous, since it would be limited to the determination of the best means for the attainment of some end implanted by nature" (Allison 1990, p. 207). Even though rational beings are free to choose which kind of imperative to follow (hypothetical or categorical ones), they can only act autonomously according to Kant if they do not follow hypothetical imperatives. Hill explains that:

"it seems clear that Kant's conception of freedom, even negatively defined, encompasses not only capacity to will without the willing being explainable by causal laws and prior events; crucially, freedom also includes the ability to will, or act for reasons, where the agent's rationale is not a hypothetical imperative indicating the means to satisfy an inclination. Without this stipulation the argument that negatively free wills necessarily have autonomy would fall flat: for autonomy ... implies a capacity/disposition to follow

principles other than desire-based hypothetical imperatives" (Hill 1985, p. 110).

The problem with following hypothetical imperatives is that according to Kant one's inclinations are not free – they are just feelings, character traits, and talents one happens to have. For him it is true that human beings are formally free to act or not according to their inclinations, but if they do act according to the inclinations they are not acting freely. The sense in which there is a free decision involved in hypothetical imperatives is this: if one thinks that they ought to do 'A' but they actually want to do 'B', and they go with what they want, what they have done is freely done in the sense that they have chosen to go along with the inclination. At the same time, what they end up doing in the situation above is what they would do if they weren't free, i.e. it's the type of action of a non-free being. In order to clarify this point let's look at a couple of cases.

Suppose one is at a pub with their friends. Suppose further that it is clear (for some reason) that one ought not to have another drink yet they decide that they want another drink and just go along with that inclination, i.e. they do have another drink. The question in this case is: are they freely choosing to have another drink? And the answer is: before they act they are free, but at the moment that they acted they are no longer free. The idea here is that the agent is abandoning their free choice and letting the physiology take over. The agent had the choice of acting freely; the one free action that they took was to decide not to act freely and when they actually act, when they have the drink, they are not acting freely. They are stepping down from their position of freedom: there was in some sense the choice there but the moment they acted, they had to be acting on some law, and the law they were acting on was an empirical law.

The basic problem in terms of freedom in the above case is that the motivation or source of the action is traceable to the empirical world (to anything other than pure rationality). Exactly the same issue may come up even if one's action accords with the moral law. Let's consider a different kind of example in order to see what Kant has in mind with this (Kant 2012, p. 13). Suppose there is a shopkeeper that knows that it is in their interest as a shopkeeper to give people the right change (in the long run customers will realize they can trust them, which will be good for business). If the shopkeeper gives the right change for that reason they are simply following a hypothetical imperative (and so wouldn't be acting morally). Now let's suppose they recognize that if they did it for that reason they wouldn't be acting morally, and decide to do it just because it is the right thing to do. What the shopkeeper is doing in this second version is the same thing they would have done if they were following their inclinations, yet they are not doing it out of inclination – they are motivated solely by respect for the moral law.

The key thing in here is what motivates the rational will to act, i.e. what is the source of the action. Kant is so seriously concerned with this issue because, as Korsgaard explains, he "claims that it is impossible for a human being not to be moved at all by incentives" (Korsgaard 1989, p. 29). His view is that if we were perfectly rational beings such as is the case with God, we wouldn't be tempted by inclinations and doing what is right would just be what we always do. But since we are imperfectly rational beings we can only be moved into action if we take an interest in the object of our action: "An interest is that by which reason becomes practical, i.e. a cause that determines the will" (Kant 2012, p. 68). This is why for Kant what matters in terms of freedom is what is the source of the interest or incentive that moves the agent into action. Korsgaard explains that:

"Kant uses the term "incentive" (Triebfeder) to describe the relation of the free person to the candidate reasons among which she chooses. An incentive is something that makes an action interesting to you, that makes it a live

option. Desires and inclinations are incentives; so is respect for the moral law" (Korsgaard 1989, p. 29).

Kant sees one's true self as just the capacity of reason – excluding as we saw everything about one's particular character. He also thinks that when the rational will acts freely it acts out of 'pure reason', meaning that the incentive or motivation for the act is nothing empirical but rather a kind of principle of reason. So what we end up with in Kantian freedom is a view in which reason acts freely when it acts only out of a principle of reason, or in other words the free act is one in which reason follows only laws that it gives to itself. And that is precisely Kant's conception of autonomy as I will now show.

As I have explained before Kant sees the will as a kind of causality of rational beings, and freedom as a property of the will that makes it possible for it to act independently of the influence of external causes. He then claims that although this conception of freedom is merely negative,

"there flows from it a positive concept of freedom. ... freedom, though it is not a property of the will according to natural laws, is not lawless because of that at all, but must rather be a causality according to immutable laws, but of a special kind .... Natural necessity was a heteronomy of efficient causes; for every effect was possible only according to the law that something else determines the efficient cause to causality; what else, then, can freedom of the will be, but autonomy, i.e. the property of the will of being a law to itself? But the proposition: the will is in all actions a law to itself; designates only the principle of acting on no maxim other than that which can also have itself as its object as a universal law. But this is just the formula of the categorical imperative and the principle of morality ..." (Kant 2012, pp. 56-7).

Autonomy comes in in Kant as a kind of consequence of the fact of our negative freedom. We are negatively free from the determination of alien forces coming from the empirical world, and as an inevitable consequence we are autonomous. Now the reason why autonomy is supposed to flow 'automatically' from negative freedom (or more precisely, from keeping one's negative freedom) is that though as

negatively free we are free from the determination of causal laws of nature, we can't not follow a law according to Kant. According to Hill's interpretation: "behavior cannot be attributed to the will of an agent ... unless it is supposed that the agent was acting for a reason, or guided by 'the idea of a principle (or law)'. Thus it is part of the concept of a will that it cannot be 'lawless'" (Hill 1985, p. 108).

The view here is that rational beings (with a will) can't avoid acting for reasons, and reasons that they believe are good ones (at least in the sense of allowing them to achieve some end that they have). And acting for reasons is understood in Kant precisely in terms of acting according to laws. It is in this sense that he claims the will can't be 'lawless', even though they are free (or more precisely, even though their freedom must be presupposed). But the point is not just that the will can't be lawless, it is also that it must act according to laws of a very specific kind if they are to remain free. The free will must follow causal laws of rationality. And in Kant's view it so happens that the only possible candidate is his moral (universal) law, as he sees it as the only kind of law an autonomous rational creature (in contrast with heteronomous - creatures that follow causal laws of nature or laws given by alien forces) would give to itself. Hill says that:

"... we can see how that argument from negative freedom to autonomy must go. Negatively free rational wills can act for reasons without being motivated by desires and hypothetical imperatives. But, as they are not "lawless," ... the agent must have, or be committed to, principles he acknowledges as rational even though they are not of the hypothetical imperative sort. Because the agent is negatively free, acceptance of such principles cannot be causally determined. ... The only alternative, it seems, is that the principles reflect some necessary features of rational agency itself independently of its special contexts. If we assume (with Kant) that one's nature as a rational will is in some sense one's "true" self ... then we could conclude further that the rational principles in question are 'one's own' or 'given to oneself by oneself' ..." (Hill, 1985, pp. 111-2).

So basically in Kant's view the agent's rational will follows laws given to it by rationality, and that is understood by Kant as autonomy. Now

I ask my reader: who or what is autonomous according to Kant's account of morality? Autonomy means that the agent only follows their own law, and Kant insists that the free agent is doing just that. But we have seen that the situation in which Kant sees autonomy is one in which the agent is following laws given to them by rationality itself 'independently of its special contexts', independently of everything that is particular to the person, or independently of everything about the person except for their rational capacity. It is then hard to see what could be autonomous in Kant's account of freedom other than rational agency itself. In other words, in Kant's account the free person is nothing but (it must be nothing but) a kind of personification of rationality. In fact this becomes evident in Kant himself if we look carefully at how he presents his view. According to him, when it comes to a moral action, "reason all by itself determines conduct" (Kant 2012, p. 39). So who or what is actually autonomous according to his view of the free agent? Kant makes it clear enough: "Reason must regard itself as the author of its principles independently of alien influences" (Kant 2012, p. 54). So in the end what is autonomous in Kant's account is rational agency itself, and this is the first of the three problems I have introduced at the beginning of this subsection on Kant.

The second problem is the other side of the same problem. Given that the only autonomous thing in Kant's account is rational agency itself, we don't have any idea of the particular person being autonomous. In fact we don't have an idea of what it is to be a particular person at all. According to Hill, for instance, Kant's understanding of one's rational will as one's true self "raises deep questions about personal identity" (Hill 1985, p. 112). The deep questions about personal identity it raises are basically that there is nothing to say about personal identity at all, as the only thing that is the person is common to every person.

I have introduced this problem earlier, when I explained what Kant excludes as mere inclination (which includes someone's desires,

character, talents, deepest aspirations, personality and particular style), and we can now see the problem even more clearly. In the end it is simply unclear what account can be given of what it is to be a person. Even from a neutral point of view one can see there is something problematic in Kant's view because there is nothing which constitutes personal identity anymore.

This issue points to a general problem with any kind of blank conception of the self, which I want to briefly explain. From the perspective of the blank conception there will be nothing qualitative which is essential to any individual person, and in this sense there is no self at all. There may be some sense in which one can identify one person as opposed to another space-temporally, through them being the occupier of a certain body. But even if they are the occupier of a given body the characteristics of that body seem to provide nothing which is essential to them.

This is the general case for the blank conception of the self. In the specific case of Kant there is a further difficulty. I have said earlier that there is a sense in Kant in which the agent always has negative freedom, even if sometimes they may act like a non-free being would act. According to Korsgaard Kant calls this notion 'spontaneity of the will', and offers this as part of the explanation as to why the will is not influenced by causality. What is relevant for the current discussion is the other part of this explanation offered by Kant:

"... Kant associates the will's spontaneity with the fact that it does not exist under temporal conditions and so is uninfluenced by causality..." (Korsgaard 1989, p. 29)<sup>19</sup>.

From the above quote we can see that in Kant there is a difficulty about the will having space-temporal location, and with this there is difficulty

---

<sup>19</sup> For a similar point made in a different context see (Hill 1985, p. 107).

even in identifying the individual person. It may be that in some way Kant could stipulate some kind of contingent relation to a particular body, but in any case the point presses even harder in an account like his.

If the fact that in Kant there is nothing which constitutes personal identity is already problematic from a neutral point of view, from the point of view of Agent Particularism this second problem and its implications are even more dramatic. Far from resulting in free persons Kant's account rather removes completely from the agent the very possibility of having even negative freedom. At the same time that the agent is pressed towards ridding themselves of everything that makes them the person they are, they are also supposed to obey unconditionally a law that can only be understood as external to their particular genius. What I mean is this: the law, which is supposed to be a demand of pure rationality, is external on Kant's own account to everything which is individual to a person and marks them out as distinct from other people. Given that, the only grip we can have on any individual self is going to be one to which the law is external. The actual adoption of the Kantian outlook would result in a total surrender of one's particular character to external forces.

This means that from the perspective of the individual self (the character who has these feelings, distinctive traits, particular attributes) the law has to seem external. From any perspective that acknowledges the individuality of the person the law will seem alien. Nonetheless Kant insists that the agent is supposed to bow down absolutely to such law. There isn't as I said even negative freedom (there is no room for protecting one's particular genius), let alone developing oneself properly, i.e. according to one's particular genius. So the drastic result of the problem of rational agency as the only thing that is autonomous, and the lack of a sense of personal identity in Kant is, in effect, the impossibility of the agent being a moral person.



The third problem we are investigating in this subsection is that Kant's understanding of autonomy raises a question about his positive freedom; in fact it implies two issues with Kant's account of positive freedom. The first issue is very similar to a problem I have already explained with Berlin's account of positive freedom. In essence Kant understands negative freedom as freedom from being determined by empirical causes, while he understands positive freedom in terms of having autonomy. Since Kant's view is that all action is determined by some law, and we just have a choice between whether it is your own law or an alien law, to have autonomy just means that one follows one's own law as opposed to externally given ones. But following one's own law as opposed to externally given ones is precisely what negative freedom is. Within a view like Kant's there is nothing of real substance involved in following one's own law that isn't at the same time involved in not following any externally given laws, i.e. there is no real substance to Kant's positive freedom. As with Berlin this happens with Kant, I think, because he is working with a blank conception of the self – if he were working with a richer-genius conception of the self he could derive his positive sense of freedom from the positive development of the particular genius of the person.

The second issue that this third problem raises is connected with the first two problems that I have explained above. The only thing which is autonomous in Kant's account is rational agency itself, which means that we don't have any idea of the particular person being autonomous. Since that is the case the person is not following their own law, but an externally given one. Now the situation in which one follows an externally given law is one in which one does not have negative freedom. But one can only have positive freedom if they have negative freedom. Since the agent, in Kant's account, has no room for having negative freedom they are prevented from the very possibility of having positive freedom. Therefore the second issue that Kant's

account of positive freedom raises is that in effect his supposedly free agent has no positive freedom at all.

There is one last criticism I would like to lay on Kant, about how his account of morality and his view on human desires are connected. The problems mentioned so far are the most threatening ones from the point of view of Agent Particularism, but what makes this problem interesting is that it is a point in common between the three accounts we have looked at in detail in this chapter that work with blank conceptions of the self, namely Berlin's, Scanlon's and Kant's. Since I have already discussed this issue in Berlin and Scanlon I will move straight into Kant.

As we have seen for Kant human beings can only be moved into action with a motivation or incentive. So when I say that Kant's autonomous agent is like a personification of rationality what I have in mind cannot be a view in which there are all the feelings to one side and reason to the other. The one feeling that Kant does allow is respect for the moral (universal) law: "All moral interest, so called, consists solely in respect for the law" (Kant 2012, p. 17).

Kant's view is that things in the empirical world are the object of inclinations and desires, whereas the law is an object of respect: "[R]espect is a feeling ... not ... received by influence, but one self-wrought by a rational concept and therefore specifically different from all feelings of the former kind, which come down to inclination or fear" (Kant 2012, p. 17). For Kant not only is respect an interest, but also there is no higher interest for a human being. This issue brings us back to Kant's vision of the free agent as one that contributes to the rational (as opposed to mere natural) ordering of the empirical world. Korsgaard further clarifies the thought:

"Kant thinks of the idea of our intelligible existence as being, roughly speaking, the motivating thought of morality, and so what makes morality possible. In the Religion, Kant tells us that one who honors the moral law cannot avoid

thinking about what sort of world he would create under the guidance of practical reason, and that the answer is determined by the moral idea of the Highest Good (6: 5/5). In the second Critique, Kant says in one place that our intelligible existence gives us a "higher vocation" (5:98/91). This vocation is help to make the world a rational place, by contributing to the production of the Highest Good" (Korsgaard 1989, p. 33).

The issue behind this is that this "higher vocation" of rational beings is what explains the fact that just respect for the law is supposed to motivate us to action more than any inclination (or the sum of all inclination as Kant says). What is relevant for us is Kant's specific sense in which rationality is able to bring the highest good into the world. Kant says:

"The principle of humanity and of every rational nature as such, as an end in itself (which is the supreme limiting condition of the freedom of actions of every human being)... " (Kant 2012, p. 43).

The reason Kant claims that rational creatures must be treated as ends in themselves and not just as a means is that they are capable of acting rationally. And what this means is that they are capable of limiting their actions in the same way - it is by limiting their actions to Kant's universalizable moral principle that human beings are supposed to make the world a better place. Kant's view is not that our wants are bad, it is rather that they have no moral worth. They only become bad when they conflict with other people's wants - at this point rationality needs to come in to mediate the conflict. Unlike Berlin and Scanlon, Kant doesn't see morality as a kind of system that we came up with to prevent human beings from becoming very nasty with one another. He sees morality as something that exists objectively in the world. Nonetheless the function that morality has in those three accounts turns out to be quite similar. In the three cases wants are seen as not open to rational criticism, and there is a system (morality) that takes

care of filtering out the wants that could be threatening to human life as a society.

#### 1.3.4. A Natural Development of Agent Particularism:

The central contrast I have been making throughout this thesis is between the blank conception of the self and the richer-genius conception of the self. I have provided a sketch of a richer-genius conception of the self, and showed that the blank conception of the self can't really be made any proper sense of. I have illustrated this point through the views of Berlin, Scanlon and Kant (which is the only one that offers some explanation about how a blank conception of the self might work), while showing what is wrong with their views. I expect that by this point it is clear to the reader that given the deep inconsistencies involved in the blank conception of the self, we must adopt a richer-genius conception of the self.

The richer-genius conception of the self is the backbone of Agent Particularism. In Chapter One I have presented Agent Particularism as a kind of rejection of universalizability. As I explained, the elementary thought of Agent Particularism is that what is right for each person to do depends upon who that person is, which in its turn depends upon features of that person that are not shared by other people – such as their character, style, and so on. So some notion like that of one's particular genius is essential to Agent Particularism. That in itself requires the idea of an ideal improvement of oneself, otherwise Agent Particularism wouldn't be a plausible rejection of universalizability. There are two points involved in this.

I am rejecting the universalizability theorist's suggestion that if x is the right thing for one person to do in a certain situation, x is also the right

thing for everybody else to do in a relevantly similar situation. So I am proposing the view that what is right for each person to do depends upon facts about that person's actual character, which in its turn requires the idea of an ideal development of oneself if we are to make good Agent Particularism. Firstly because I also want to reject the suggestion that what is right for someone to do might be determined by some of their failings. If the idea of the ideal development of one's true self is not built into Agent Particularism the position comes close to looking like an excuse for people being bad people, getting them out of all kinds of criticism based on ideas like whichever features they possess currently determine what can be demanded from them.

The second sense in which unless we have the notion of the ideal development of the person Agent Particularism wouldn't be a plausible rejection of universalism is explained by the fact that without this notion we end up with a form of subjectivism. The problem underlying this issue has to do with one of the universalist's main motivations for insisting on the universalizability principle. As I have explained in Chapter One:

"... the thesis of universalizability ... rests on a concern about the evils of arbitrariness. Such a concern, rather than any considerations of logic or semantics, is what truly justifies the aforementioned principle and thesis" (Kramer 2005, pp. 173-4).

It is required of any Agent-Particularist position, if it is to be a plausible rejection of universalizability, that it gives an appropriate response to the challenge of the 'evils of arbitrariness'. The way to respond to this properly is by allowing room for constraints of reason in morality, which the elementary Agent-Particularist position I formulated in Chapter One does. Basically I make room for constraints of reason as I ground what is essential to the person in features of their actual character, and then avoid the first problem I explained above with the idea of the ideal

development of oneself. Someone like Winch, as I explained in Chapter One, tries to make room for rational constraints through a different route. He justifies the view that what is right for different people to do in relevantly similar situations might differ with the notion of what is morally possible for each person, and he grounds that in something like the moral outlook they have at the time of the action.

Now, if what justifies the kind of moral constraints that are applicable to the theory is something like one's moral outlook it in the end is shaped by people's values, beliefs and so on, i.e. something subjective about them. Even if there is some kind of progression or development of the person it can only be a progression of something about their subjective content.

This is the sense in which this kind of view can't dissipate the universalist's worry with the evils of arbitrariness. Since within a blank conception of the self we can't have some objective features of the person grounding the right kinds of values to have there is something arbitrary undeletable from the account, coming from the fact that what is supposed to justify the relevant differences between people is something wholly subjective.

With the two above senses in mind we can see that Agent Particularism essentially involves the idea of the ideal development of the person. Since this idea is essential to Agent Particularism, a conception of negative freedom similar to the one I have offered here is also essential to it. This is because, as I have explained earlier, the proper development of the person depends on the freedom to be oneself, i.e. to act out of one's true self (this is what I meant when I said that the value of negative freedom is given by positive freedom). There is another way in which the kind of conception of negative freedom which I have offered is essential to Agent Particularism: once we have accepted the richer-genius conception of the self we can see, as I have

explained, a threat to the freedom of that genius coming from social pressures for uniformity. Negative freedom is therefore essential to Agent Particularism as a way of protecting the true self from those external pressures.

I will now transition into positive freedom. In terms of what is formally required for an Agent-Particularist account of positive freedom the main thought is that the ideal development of the person is one that happens organically in accord with one's particular genius. In other words, it is essential to Agent Particularism that one's positive development flows out of one's particular genius.

Once you have a richer-genius conception of the self you have a much more substantial contrast between negative and positive freedom. Negative freedom will be a necessary condition for positive freedom, but positive freedom will involve the freedom to develop yourself in the direction of the ideal version of yourself. This means that in order to be free to develop in the appropriate way the agent needs things like removing various kinds of impediments, and working on their own character in the right kind of way.

The thing is that for someone coming from the blank conception, such as Berlin and Kant, positive freedom ends up being something like negative freedom reformulated into different words. There is a very thin account of positive freedom and there isn't much that can be done in terms of that. Whereas if you are coming from the richer-genius conception of the self there is a substantive account of positive freedom to be had, and something needs to be said about that.

This anticipates a thought I will develop in the section on positive freedom. For the moment I want to point out that what the Agent Particularist needs is a conception of what is involved in developing oneself in such a way as to have one's own particular genius become something nearer the ideal version of themselves. Naturally anybody

thinking about how they are going to do that will think of different ways of fleshing out the kind of Agent-Particularist skeleton that I have presented above. Any particular version of this will describe a certain kind of psychological process in terms of what needs to happen to prevent the person developing the kind of fake second nature I mentioned earlier. So that will involve a process of a certain character, and we will see that philosophers such as Humboldt, Mill, and Emerson - who are developing Rousseau's tradition of the richer-genius conception of the self and so share the concerns involved in the Agent-Particularist negative freedom - are precisely describing this process.

Before going into positive freedom I would like to offer the basis for something of a similar kind as the process described by Humboldt, Mill and Emerson. There is a very natural way of developing Agent Particularism which has interesting connections with Kant and it is worth exploring that route. The reason is that there is something natural about Kant's view that he misunderstood, and I want to show where Kant misunderstood it by over-radicalizing certain distinctions. I will point to certain notions in Kant that work so naturally as developments of Agent Particularism that Kant might himself have been tempted towards seeing them in this way. Exploring this route also has the additional advantage of helping us to further flesh out the bare bones of Agent Particularism. My point in this sub-section is not in giving an exhaustive account of how some of Kant's notions can be developed in an Agent-Particularist way. Rather I simply want to give a sense of how that can quite naturally be the case. For the most part I will organize this discussion as an illustration of what the threat to freedom at the internal level might look like, making the necessary contrasts with Kant in the relevant places.

There is a simple move that makes some of Kant's notions remarkably fitting for Agent Particularism, and at the same time points to one of Kant's key misunderstandings. The move is the following: imagining



that whenever Kant is talking about alien forces inside the person what he has in mind for 'alien forces' is what the Agent Particularist has in mind, i.e. forces that are external to one's particular genius. And the misunderstanding this move unveils comes from Kant's conception of the notion of inclinations.

This issue brings us back to the contrast between my Agent-Particularist treatment of inclinations and Kant's treatment of them. Kant as we have seen excludes as mere inclination and therefore as alien (among other things) all feelings except for one – the feeling of respect for the moral law. Meanwhile it is essential to Agent Particularism that a distinction is made between two basic kinds of feelings: the feelings I refer to as sentiments – feelings that are original expressions of one's particular genius; and the feelings I refer to as passions – feelings that are alien to one's true self. I have already explained in detail earlier what is the Agent-Particularist understanding of the distinction between sentiments and passions. What I want to offer here is a natural development of this distinction, through the concept of the second nature that I have introduced earlier in the chapter.

But before getting into that there is a clarification which I want to make. I want to be explicit about the fact that I am picking up on points that I find relevant in Kant and applying them in my own way. This means to clarify that I am not at the same time claiming that Kant would himself agree with the way in which I am treating those points. He might have been tempted by those views, for the reasons that I will present below. Yet I don't need to prove this for my purposes in this thesis.

Still it is significant to note how, when seen from the perspective of a richer-genius conception of the self, the points I will explore now feel like a natural development of some of the points I have picked up from

Rousseau in order to develop my Agent-Particularist freedom. In fact Suzan Neiman helps us to see that this may be no accident. According to her "... Rousseau was Kant's guiding star ..." (Neiman 2016, p. 34). The influence that Rousseau had on Kant shouldn't be taken lightly. Neiman shares a story according to which Kant would religiously follow his routine that included a morning walk, and only two events had such a strong impact on Kant that caused him to forget about taking his walk. One was when Kant learned about the French revolution, and the other was coming across Rousseau's work:

"Jean-Jacques Rousseau left him spellbound. It wasn't an easy experience. Kant later wrote that he had to read Rousseau's sentences several times in order to understand them, so stirred was he by the beauty of their prose. The experience was liberating, as we saw in the note that said it was Rousseau who changed his life and taught him his true calling<sup>20</sup>" (Neiman 2016, p. 33).

What is relevant for us is not merely that Rousseau was a major influence on Kant's work, but rather the kind of ideas from Rousseau that influenced Kant <sup>21</sup>. According to Neiman it was precisely Rousseau's views on human development, or as she puts it his "... vision of growing up [that] most inspired Immanuel Kant..." (Neiman 2016, p. 14). Rousseau's account of education (which as I pointed out earlier contains his view on freedom and human nature) was a major influence in Kant's work: "Rousseau's Emile is the text Kant took for granted in most everything he wrote about humankind" (Neiman 2016, p. 35).

---

<sup>20</sup> Kant had such a great admiration for Rousseau's ideas that the only piece of art he had in his house was a portrait of Rousseau (Neiman 2016, p. 34).

<sup>21</sup> Some ideas that on my view could most naturally be seen as notions that Kant picked up from Rousseau would be: a) freedom involves reason; b) autonomous action is following a law that you give to yourself; c) the threat to freedom that the influence of passions represent; d) the notion of treating people as an end – which is precisely coming from the idea of respecting people's geniuses.

The most remarkable thing is that apparently Kant came close to acknowledging Rousseau's (and consequently Agent Particularism's) understanding of human nature:

"I am by nature an inquirer. (...) There was a time when I believed that this is what confers real dignity upon human life, and I despised the common people who knew nothing. Rousseau set me right. This imagined advantage vanishes, and I learned to honour human nature" (Kant quoted in Neiman 2016, p. 28).

Again, I am not trying to prove the point that Kant consciously developed in his own way certain points that he picked up from Rousseau's account of education, as proving that is not necessary for Agent Particularism. What I want to make clear is that I will intentionally shape these points that I am picking up from Kant as a natural development of the Agent-Particularist's views on freedom, some of which are based on Rousseau's notions – and doing that is not as far-fetched as some people might initially believe.

As I have explained earlier the different treatments of the notion of inclinations is a key difference between Agent Particularism and Kant's view. What I want to offer now in terms of a natural development of Agent Particularism – though not compulsory to an Agent-Particularist position – involves the link between this difference in treatment of the notion of inclinations and the difference in our accounts of reason. The idea involves keeping Kant's contrast between free actions as rational and unfree actions as not properly rational. Unlike Kant (and Berlin and Scanlon) this view does not understand desires as not open to rational criticism; on the contrary rationally criticizing one's desire is, on this account, essential both in terms of negative and positive freedom.

Kant's view is that properly rational action is free from the influence of inclinations, treating feelings – except for the feeling of respect for the law – as external to one's true rational self. Agent Particularism on the

other hand makes a distinction between different kinds of feelings. It treats sentiments as genuine expressions of one's particular genius and in this sense as one's legitimate feelings, while it treats passions as external to one's particular genius.

The Agent Particularist will take desires to be open to rational evaluation, i.e. we can distinguish between desires that are rational and desires which are not. A desire will be rational if it is a desire which one ought to have, if one has reason to have such desire. Now, on the Agent-Particularist position the contrast between sentiments and passions is that sentiments are reflections of one's true self. And it is also the case that effectively what one ought to be is their true self, which guarantees that desires which are reflections of one's true self are the desires one ought to have. Therefore sentiments are rational. Since Agent Particularism is a theory according to which what one ought to be is their true self, sentiments must be the right kind of desires to have.

So while Kant excludes sentiments (as inclinations) because he sees them as external to one's true self – which is in his view pure rationality – this Agent-Particularist development takes them as genuine expressions of one's self. This means that while for Kant sentiments are to be rejected (together with everything else he includes as mere inclination), for Agent Particularism they can't be rejected (they are an integral part of you being who you ought to be). Given the requirement to develop into the ideal version of oneself not only sentiments are not to be rejected but also they are to be pursued and fulfilled.

I want to turn now to the distinction between one's nature and the alien second nature, as part of this natural development of Agent Particularism. The basic contrast is between whether the source of the action is one's true self or alien influences, and for Kant that means the source as pure reason or inclinations (with everything that the

concept involves). On the Agent-Particularist alternative the true self is rather identified with one's particular genius and so that is the appropriate source of the actions of a free human being. And a natural way of developing so as to illustrate this point is in terms of looking at what the agent's concern is in relation to other people's opinions.

If the concern is to become the best possible version of themselves it is unproblematic as the source of one's action. So there is room for instance for receiving advice from someone you trust very much, as long as that advice is properly processed by the agent who may accept it entirely, in part or not at all. If the kind of concern one has in terms of other people's opinions involves a concern for having other people's approval and recognition for the sake of being highly esteemed within society, that would then be seen as an alien influence. This is the wrong source of one's actions according to the Agent Particularist, and as such it is precisely the kind of thing inner negative freedom requires independence from.

What is important in terms of the distinction I made above is that it illustrates the intrinsic irrationality of passions. As I explained when I claimed that sentiments are rational, when we are talking about rational desires what we are talking about desires that one has reason to have. So when we are talking about irrational desires what we are talking about are desires that one has reason to avoid having.

A sentiment is clearly a rational desire; once it is the case that one has reason to be one's true self one obviously has reason to have the desires that are expressions of one's true self. It's not, in the same way, immediately obvious that passions are immediately irrational in the sense that they are desires that one has reason to avoid having. The reason is that it is not immediately obvious that passions will go against one's true self. What is nonetheless immediately obvious is this: because passions are desires for the preservation of something

other than one's true self they are desires that one has no reason to have. On top of that, in so far as pursuing the preservation of the alien self damages the development of the true self (as I have explained) passions are irrational - they are precisely the kind of desires one has reason to avoid having.

Now I want to turn to one of Kant's notions that is in fact essential to the Agent-Particularist position. It is the notion of 'absolute worth'. There is within the Agent-Particularist position a demand to not force people to be something other than they are, and it is because of the absolute worth of someone's particular genius that it is wrong to force them to act in a way that they couldn't act while still being the same person (likewise that is the reason why it is wrong for them to be other than they are). Now what confers the absolute worth to the particular genius is the fact that even if developing oneself according to one's particular genius is not sufficient for them being moral, it is required for the reasons that I explained when I presented the view that positive freedom justifies negative freedom.

When we think of the true self in terms of one's particular and rich genius as Rousseau does, the idea that human beings have absolute worth acquires real substance. Instead of wanting to ground people's absolute worth in rational agency, itself detached from everything about them as a particular person, Agent Particularism grounds people's absolute worth in the unique characteristics that composes their particular genius, which makes them the person they are. This issue becomes very clear in Kant's application of the notion of treating people as an end and not merely as a means to get what you want. When we look at it from the point of view of a richer-genius conception of the self, we can see that though Kant uses the notion of treating people as ends, ironically (from the point of view of the richer-genius conception) his account issues in a lack of both self-respect and respect for others, because Kant fails to acknowledge what in fact demands

respect in people – their particular genius, and the (moral) need to develop it properly.

This discussion on the notion of 'absolute worth' anticipates a point which I will develop in Chapter Four, which is on the notion of human dignity. In the beginning of this chapter I presented Rousseau's view that human beings are originally good, in that the original impulses of nature are always good and so people are originally good in their particular natures. I also explained that what is good in each person is the actual unique content of that person's particular nature. This is a background thought for the Agent-Particularist conception of the notion of absolute worth.

When the Agent Particularist accepts the idea of the absolute worth of a person, they have in mind the view that in order to become moral at least part of what is required is that they develop into the ideal version of themselves. This is what justifies the Agent-Particularist demand for freedom. And this is also what gives the peculiar content to the Agent-Particularist conception of the notion of human dignity. In accepting the idea that people are originally good the Agent Particularist accepts two views: the first is that we can't make sense of an agent not having a self which has an ideal development; and the second is that we can't make sense of an ideal development of a person not being a good thing. Which, as I will explain in Chapter Four, issues in the view that respecting people's dignity is a matter of letting them be themselves even if one disagrees with the way they are (properly acknowledging that they are also required as they act to respect other people's particular nature). This involves acknowledging that even if one may say they don't like them, they are not open to moral disapproval. That is what it is to respect the dignity of a person just as a person.

## 2. An Agent-Particularist Account of Positive Freedom:

As I said in the previous subsection the main thing that is essential to an Agent-Particularist conception of positive freedom is that the ideal development of the person is understood in terms of an organic development that occurs in accord with one's particular genius. What else can be said in terms of an Agent-Particularist conception of positive freedom?

The conceptions of human nature and freedom to be found in Rousseau's account of education have been a major source for the Agent-Particularist conception of negative freedom I have presented. So a really natural way of developing an Agent-Particularist positive freedom is by drawing from Rousseau's own positive development of those points.

Once one is coming from the richer-genius conception of the self one ends up with a substantial process involved in the exercise of positive freedom, and that space needs to be filled in. And one can see that there is a whole tradition in the history of philosophy that has picked up on the Rousseauian richer-genius conception of the self trend (and some of the implications he draws from it). Now, a certain kind of remark that one finds in the philosophers in this tradition that might at first look like some kind of psychological speculation or moralizing is in fact a description of what is required in the exercise of positive freedom. These remarks have a structural role in a larger project which is precisely of the character I have been describing in this chapter. This is what we will see in the particular cases of Humboldt, Mill and Emerson.

What I mean by this is that there is a formal position involved in the richer-genius conception of the self, in which what needs to be done is



to make sense of the idea of a contrast between sentiment and passion. Since the view is that sentiments are genuine feelings of the person while passions are imposed in large measure by other people, it implies a certain kind of way of being true to oneself. A whole kind of way of describing challenges to oneself and to one's integrity become important and natural as we have seen in the conception of negative freedom. And the substantial process of developing into something nearer the ideal version of one's self, i.e. positive freedom, becomes required – and so a conception of what is involved in positive freedom also becomes required. Any particular version of this will describe a certain kind of psychological process, and when one reads these philosophers saying things very similar to each other what one can see is that what they are describing is exactly this process.

So this section will bring out aspects of Humboldt, Mill, and Emerson that don't normally feature largely on philosophical discussions of theoretical philosophers. Given that this section will in effect find a place for a whole side of philosophy that contemporary theoretical philosophy on the whole has not given space to.

## 2.1. Negative and Positive Freedom:

In the section on an Agent-Particularist conception of negative freedom I have explained that the value of negative freedom is informed by positive freedom. So when asked 'why is it good to have negative freedom?', the Agent-Particularist answer is: because without negative freedom the human being can't have positive freedom. And the problem with not having positive freedom is, as we have seen, that one wouldn't in such a case become a properly developed version of themselves. I have already explained in the section on negative

freedom how this idea works, but want to flesh out the key points here given the deep connection between negative and positive freedom. That in itself would already justify having a short discussion on this connection here, but there is still another reason for me to do it. There is a sense in which the positive development of certain aspects of the person starts with developing negative freedom in terms of those aspects. But before going into the key points of the connection between negative and positive freedom there is a clarification that I would like to make.

Though negative and positive freedom are so deeply connected, there is a clear way of distinguishing between them. We can see this contrast if we ask ourselves, 'What is it to exercise negative freedom?', and 'What it is to exercise positive freedom?'. To exercise negative freedom is simply to act on a law<sup>22</sup> of your own, or not under alien determination, as I pointed out. To exercise positive freedom is a process – one exercises positive freedom through developing themselves so as to become the best possible version of themselves. So we can distinguish between negative and positive freedom by seeing that the exercising of negative freedom is an act, and the exercising of positive freedom is a process.

Having made this distinction let me now turn to fleshing out the key aspects of the connection between negative and positive freedom. Essentially the threat to freedom comes, as we have seen before, from pressures coming from society towards uniformity. Mill helps us to see how this connects to positive freedom:

"... peculiarity of taste, eccentricity of conduct, are shunned equally with crimes: until by dint of not following their own nature, they have no nature to follow..." (Mill 1991, p. 68).

---

<sup>22</sup> For the sake of clarity I am for the moment riding with Kant here in the use of the term "laws" although in fact when we get to the Agent-Particularist virtue ethics which I will develop in the next chapter it will become clear that the notion of "laws" does not apply to Agent Particularism.

The problem with giving in to such social pressures towards uniformity is that these are influences that are external to one's nature and which have the effect of leaving one out of touch with oneself. In Emerson's poetic words: "The objection to conforming to usages that have become dead to you is, that it scatters your force. It loses your time and blurs the impression of your character" (Emerson 1841).

To be clear my view is not, of course, that the particular nature of the agent that doesn't have negative freedom ceases to exist. Rather what I have in mind is that they will come to identify themselves more and more with the alien second nature, and as a consequence their true nature will remain dormant, underdeveloped. From this perspective of the agent that identifies themselves with second nature as opposed to their true nature Mill's remark becomes useful to us:

"I do not mean that they choose what is customary, in preference to what suits their own inclination. It does not occur to them to have any inclination, except for what is customary" (Mill 1991, p. 68).

The key thing in the relation between negative and positive freedom can be understood in the following way. One needs to protect one's particular genius from both social pressures towards uniformity and the tyranny of passions. The reason for that is that without such negative freedom one will not live their life, will not have their experiences, as the human agent that they are. Humboldt unpacks this point in the following way:

"Whatever man is inclined to, without the free exercise of his own choice, or whatever only implies instruction and guidance, does not enter into his very being, but still remains alien to his true nature, and is, indeed, effected by him, not so much with human agency, as with the mere exactness of mechanical routine" (Humboldt 2014, pp. 29-30).

What makes acting mechanically as opposed to acting with human agency a problem is the consequences of behaving in such a way. Mill, who also concerns himself with this problem, sees one of these consequences in the following way:

“... their human capacities are withered and starved: they become incapable of any strong wishes or native pleasures, and are generally without either opinions or feelings of home growth, or properly their own” (Mill 1991, p. 68).

Let’s imagine the agent’s particular genius as a platform (a kind of foundation) for a moment. We would have a sense in which what negative freedom provides is the possibility of the platform being a platform, from which positive freedom would then be able to raise itself to a standing position. The image of the condition of muscles in the human body also gives a good analogy to illustrate this point. If one doesn’t in some way exercise their body, and thus doesn’t put their muscles into exertion, the muscles can go as far as to atrophy from lack of use. If one puts their muscles into constant exercise on the other hand they strengthen and develop. In a similar way by behaving autonomously, which implies behaving out of one’s particular genius, one develops according to that genius. The key thought behind this is in the idea that by putting into practise one’s actual capacities one develops them.

## 2.2. The Process of Positive Freedom:

In this section I will present my Agent Particularist account of positive freedom. I have already said that positive freedom is about developing into an ideal version of oneself (I will unfold the point in this section). At the outset of the section I would like to raise a question: why does what I call positive freedom count as a form of freedom precisely?

In order to answer this question it is helpful to think of the concept as having longer roots back in an older society where freedom is not so much freedom from the imposition of government, or prison or anything of the like. What we have in this context is freedom in contrast to being a slave. It is relevant to note that in his account of education Rousseau relies on the contrast between the free person and the slave. In explaining why there is a threat to the freedom of the child Rousseau says:

"All our wisdom consists in servile prejudices. All our practices are only subjection, impediment and constraint. Civil man is born, lives, and dies in slavery" (Rousseau 2010, p. 167).

Once we have the contrast of freedom to be with slavery we can ask, what is distinctive of someone who is free? There is with the contrast between freedom and slavery a conception of what a free person - the person who is not a slave - does. There is within the tradition a strong tendency to think that what is characteristic of a free person is a kind of natural flowering. The key thought is in the idea of developing oneself, becoming one's own person rather than being constantly enslaved by other people's conception of how one should be. This is what is characteristic of a person who is free rather than a slave. And that is precisely the kind of idea I will develop in this section. This in fact anticipates a point which I will come back to later in the section.

As I explained at the beginning of this section a range of remarks to be found in Humboldt, Mill and Emerson should be seen as a way of filling the space that is there to be filled as a description of the process of the agent coming as close as possible to the ideal version of their own selves. I would now like to explain what I meant better.

Once we have the idea that the exercise of positive freedom is a process rather than an act we can see that the process itself requires certain kinds of virtues of character which other acts, like action

towards other people, may not require. Now, that is the focus of a whole range of remarks to be found in Humboldt, Mill and Emerson. In other words, if you have a richer-genius conception of the self there is a fundamental difference between negative and positive freedom: the exercise of negative freedom is an act, the exercise of positive freedom is a process – the development of oneself – and because that is the case it requires virtues of character that no individual acts require. It is then therefore not surprising to find these philosophers who are in the Rousseauian tradition precisely praising these virtues of character. In this sense a contribution of this section as I have said before is making sense of certain aspects of a whole tradition in philosophy that hasn't been in general properly acknowledged.

This anticipates something from the next chapter, on an Agent-Particularist virtue ethics. Virtue ethicists are naturally thought of as Agent Particularists because their view in general is that what one has to do is to become a virtuous person. What that means is that what one has to do is become an ideal version of oneself. This means that the idea of the ideal development of the agent is common to Agent Particularism and traditional virtue ethics. But in general virtue ethicists are also universalists and in effect think that everyone has the same ideal version, whereas for the Agent Particularist the ideal version of different people will differ according to the differences in their particular geniuses.

Before going into Humboldt, Mill and Emerson's descriptions of the virtues of character required for positive freedom there is something I would like to make explicit. The basic Agent-Particularist position involves the view that what is required of each person depends upon who that person is, which depends upon features of their actual character. That in itself requires the idea of an ideal improvement of oneself as I have shown. Now that I will start the part on positive freedom properly speaking it is opportune to pose ourselves a

question: What else this essential position, the very idea of Agent Particularism, requires?

This question brings us to three (at least) possible positions that could be taken as developments of the essential position, which I would like to briefly look at now. Basically the first position is that what is right for one to do depends upon who the person is (which requires the idea of the ideal development of their true self). The second position involves the view that there is nothing involved in one doing what they should do, or that would be good for them to do, other than them doing what the ideally developed version of their true self would do. And the third and more radical view involves the idea that what one ought to ultimately do is simply be oneself. Let's look at these three views more closely.

I want to start with the more radical view, which is that what one ought ultimately to do is just be oneself. The idea behind it is that the only motive that someone has for acting is always to do with their developing into the ideal version of themselves. I want to start from this view because I want to rule it out as inconsistent from the start. The problem with it is that it leaves no room for someone being required to do something if that doesn't generate their development.

Suppose there is situation which you would normally see as a moral one, in which you would be required to do some simple action, a not particularly demanding one. Because it is not particularly demanding it wouldn't really involve any further development of the self – it would be a good thing for you to do, but just wouldn't contribute to your development in any way. Does it make sense to say that they are not really required to do it? In order to make the point more dramatic let's imagine someone which has already developed into the ideal version of themselves. Does it make sense to say, because there is no more development to be had, that they are not required to do all the things

that they would be required to do if those things were to cause their development? Of course it wouldn't, therefore we can rule out this as a possible development of the basic Agent-Particularist position.

Another position would be that there is nothing involved in one doing what they ought to do other than them doing what the ideally developed version of their true self would do. The thought is that one is currently required to do everything that the ideal version of themselves would do, irrespective of what point they find themselves in the process of their development. There are two implications to holding this view. First it makes no distinction between the requirements on one now and the requirements that one would be under once they have developed into the ideal version of themselves. Secondly it doesn't distinguish, even when one has reached the stage of the ideal development, between what they are required to do and what is not required but would be optimal for them to do.

The third position which I want to look at is centred around the view that the person can't be required to do things which they in some sense couldn't do while still being the same person. So while in the previous position the ideally developed version of themselves determines all of what they ought to do, in this position who they are provides a constraint on what they ought to do. It doesn't imply that everything which an ideal version of themselves would do is required of them. This thought can be unpacked into two distinct issues. One is that some things that the ideal person would do might not be required but just optimal, which means there is a possible contrast with the ideal person between what is required and what is optimal.

The second issue involves a contrast between what is required of one now and what would be required of the ideal version of one's self now. The relevant question here is, in what sense the person is now required to do what the ideal version of themselves would be required to do?



There is some sense in which one ought to become the person who would be required to do that, and not everybody is required to become that person, but even so one might not now be required to do it. The reason is that it might be the case that one is not yet in a state of being capable of doing what the ideal version of themselves would do. In order to illustrate this point let's suppose a situation in which the person knows what they should become but they haven't got there yet. There are some things which they in a certain sense ought to do, in that they ought to be someone who is capable of doing that, but they are not now capable of doing it. In such a case they can't be required to do it.

I will come back to this issue in Chapter Four, when I introduce the Agent-Particularist understanding of the notion of supererogation. Now I want to highlight certain natural ways of developing the basic Agent-Particularist position which describe the virtues of character required in the process of becoming who one ought to become. I will do that by showing where we would get to if we were to follow the main leads to be found in Rousseau, Humboldt, Mill and Emerson.

#### 2.2.1. Strength of Character:

The view of the autonomous agent as strong (as opposed to the agent that lacks autonomy which is weak) is something echoed by Rousseau, Mill, Emerson and Humboldt. Rousseau for instance describes this contrast in the following way:

"The wicked man fears and flees himself. He cheers himself up by rushing outside of himself. (...) By contrast, the serenity of the just man is internal ... he bears its source in himself" (Rousseau 2010, p. 451).

A sense of 'the wicked man' as Rousseau uses it could be illustrated in the following way. Though a rational being, their capacity to engage their own reason becomes fragile overtime, and they will most times happily borrow another's instead. Qua rational being, they are weak in two senses. First in the sense that they blindly follow other people's views as opposed to engaging their own reason (irrespective of whether or not those views are rational). Second in the sense that their reason becomes enslaved by alien passions.

The autonomous agent on the other hand reassures themselves in their own absolute worth. They have no reason to follow other people's opinion if those people don't offer them a good reason for changing their mind. They simply focus on doing what is properly rational. Second nature emerges in them too, and so do its accompanying passions, yet they know how to differentiate a good reason from the emotional grip of passions. By practising this commitment to their own genius and reason (which implies practicing being themselves) they are developing strength of character. A natural thought is that it is precisely this foundation which is required to support their proper positive development, allowing them to become moral.

I have explained earlier that the exercise of negative freedom (autonomy) is an act and the exercise of positive freedom a process. I also said that part of the process of exercising positive freedom is developing negative freedom (as a kind of foundation). The strength of character I am referring to here relates to negative freedom in the sense of it as part of the process of developing positive freedom, rather than merely as an act. At the same time it is precisely by practising autonomy in the different occasions that such foundation is built. Something like this view is at the heart of Emerson's conception of self-reliance and is reflected in the following quote:

"Insist on yourself; never imitate. Your own gift you can present every moment with the cumulative force of a whole life's cultivation; but of the

adopted talent of another, you have only an extemporaneous, half possession” (Emerson 1841).

If we continue to follow Emerson’s track, ‘insisting on yourself’ includes the view that one should insist on being themselves even when one can not foresee where this will take them. The thought behind the importance of this attitude is the idea of the development of one’s particular genius as a kind of dynamic process (as opposed to a view in which one’s nature is something pre-fixed, and one’s job is to as it were bring that pre-fixed thing out). Emerson makes the point in the following terms:

“No man yet knows what it is, nor can, till that person has exhibited it” (Emerson 1841).

Since there isn’t something like the exact map of one’s development Emerson offers an alternative solution:

“A man should learn to detect and watch that gleam of light which flashes across his mind from within...” (Emerson 1841).

There are different possible ways of developing a notion like strength of character as a virtue of character required for developing into the ideal version of oneself, and here I have illustrated a few. What is essential not only in this but in any virtue of character involved in the Agent-Particularist positive freedom is its link with morality, as Rousseau in different occasions points to:

“All wickedness comes from weakness. The child is wicked only because he is weak. Make him strong; he will be good” (Rousseau 2010, p.161).

### 2.2.2. Developing as Fully as Possible in Accord with the Potential of One's Particular Genius:

#### 2.2.2.1. Realizing One's Potential:

The thought that justifies the Agent-Particularist positive freedom is that what a person is obliged to do is dependent on what an ideally developed version of that person would do. From this thought flows the notion of a standard moral demand on human beings to develop into the ideally developed version of themselves. This is where the notion of 'potential' comes in. More specifically this is where a demand to realize one's potential as far as possible comes in. Given the richer-genius conception of the self that the philosophers in this tradition operate with, implicit in the notion of people's potential is the idea that each person's potential is unique. I want to remind the reader at this point that though the motive of developing into the ideal version is an essential part of becoming a moral person it isn't the only motive one has. I am making this remark in order to explicitly clarify that this view is not the point made in this section.

As was implicit during the discussion on negative freedom, realizing one's unique potential is not among the highest priorities in society. On the contrary the uniformity found in society tends towards 'appearing' rather than 'being', and that undermines the necessary basis for one's potential to flourish. Mill expresses his version of this concern in the following way:

"If the grounds of an opinion are not conclusive to the person's own reason, his reason cannot be strengthened, but is likely to be weakened, by his adopting it: and if the inducements to an act are not such as are consentaneous to his own feelings and character ... it is so much done towards rendering his feelings and character inert and torpid, instead of active and energetic" (Mill 1991, p. 65).

Setting aside the religious tone that Mill uses, the passage below gives us the essence of the sense in which developing our potential is something which is morally required of us:

"Many persons, no doubt, sincerely think that human beings thus cramped and dwarfed, are as their Maker designed them to be; just as many have thought that trees are a much finer thing when clipped into pollards, or cut out into figures of animals, than as nature made them. ... it is more consistent with that faith to believe, that this Being gave all human faculties that they might be cultivated and unfolded, not rooted out and consumed, and that he takes delight in every nearer approach made by his creatures to the ideal conception embodied in them, every increase in any of their capabilities of comprehension, of action, or of enjoyment" (Mill 1991, p. 69).

When we consider the idea of realizing one's potential in the context of a positive freedom derived from a richer-genius conception of the self, three directions in which this idea could be developed naturally come to mind. One, taken from Mill, involves the idea that each individual capacity requires practice and a maturing process to develop, as the following quote suggests:

"The human faculties of perception, judgement, discriminative feeling, mental activity, and even moral preference, are exercised only in making a choice. ... The mental and moral, like the muscular powers, are improved only by being used" (Mill 1991, p. 65).

A second way of fleshing out a richer-genius based conception of 'realizing one's potential' involves, besides the idea that one develops their capacities by putting them into practice, the idea that it is important to develop in all directions that are appropriate for their particular genius, as Mill's quote below suggests:

Human nature is ... a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing" (Mill 1991, p. 65).

The contrast here is between developing in all directions that are appropriate for one's particular genius, and developing in certain directions but remaining underdeveloped in other aspects. The main relevant implication in this distinction for the context of our discussion is that the partly underdeveloped person of the second kind of case is subject to a lack of balance or harmony between their capacities. This would issue in a kind of lack of consistency in their character. In order to avoid this problem Humboldt suggests that one's proper development requires that one gradually bring those different capacities into a state of working in harmony:

"... unite the separate faculties of his nature, often singly exercised; by bringing into spontaneous co-operation, ... and endeavoring to increase and diversify the powers with which he works, by harmoniously combining them" (Humboldt 2014, p. 12).

If we consider that Humboldt accepts the idea that the development happens through practice, we can see the following quote as a way of assimilating something like the three directions I have just presented:

"The true end of Man, or that which is prescribed by the eternal and immutable dictates of reason, and not suggested by vague and transient desires, is the highest and most harmonious development of his powers to a complete and consistent whole" (Humboldt 2014, p. 11).

Now, once one acknowledges how rich the idea of realizing one's potential actually is and acknowledges the undermining consequences of the threat coming from the pressures towards uniformity of thought and behaviour, a natural thought comes to mind. It is the thought that when one is coming from a richer-genius conception of the self the idea of 'realizing one's potential as far as possible' carries with it a view of the excellence of agents which might go beyond what would be expected for the blank-conception theorist. Besides the fact that the view starts from the view of a self who is rich in its unique character,

in order to grasp the kind of excellence we would have in mind in this context we need to also account for the cumulative effect of the development. Humboldt makes a point in this same spirit:

"It is impossible to estimate a man's advance towards the Good and the Beautiful, when his unremitting endeavors are directed to this one engrossing object, the development of his inner life" (Humboldt 2014, p. 28).

#### 2.2.2.2. Developing One's Uniqueness:

Agent particularism accepts the idea that even if developing into the ideal version of oneself is not sufficient for becoming moral, it is at least an integral part of everything involved in becoming moral. It is from this perspective that we can understand the remarks on this section on the development of one's uniqueness as a kind of virtue of character. In the quote below we can see Mill making the link between individuality and one's ideal development:

"Having said that individuality is the same thing with development, and that it is only the cultivation of individuality which produces or can produce, well-developed human beings, I might here close the argument: for what more or better can be said of any condition of human affairs, than that it brings human beings themselves nearer to the best thing they can be? or what worse can be said of any obstruction to good, than that it prevents this?" (Mill 1991, p. 71).

I made earlier a point about the remarks that these philosophers make being not just incidentally moralizing, but rather play a structural role in a broader project similar to the Agent-Particularist conception of freedom I am presenting in this chapter. The above quote illustrates this point, in that it clearly works with the same kinds of notions I have been working with in this chapter, such as the idea of the demand to develop into the ideal version of oneself and the idea of the importance of one's particular genius. If we follow Humboldt, we can have a

connection not only of the ideas of individuality with the ideal development, but also a connection of those two and the notion of absolute worth (applied in precisely the kind of way Agent Particularism needs it to be applied):

"... that inherent greatness which is comprised in the life of the individual, and perishes along with him,—the bloom of fancy, the depth of thought the strength of will, the perfect oneness of the entire being, which alone confer true worth on human nature. ... this essential worth of human nature, of its powers and their consistent development..." (Humboldt 2014, p. 8).

Humboldt therefore sees individuality, and its proper development, as being at the core of the notion of the absolute worth of the person. And he sees the positive process of one's development according to their particular genius as a process guided by reason, another idea which is integral to Agent Particularism:

"... reason cannot desire for man any other condition than that in which each individual ... enjoys the most absolute freedom of developing himself by his own energies in his perfect individuality" (Humboldt 2014, pp. 17-8).

Once one is starting from a richer-genius conception of the self the notion of individuality becomes naturally a very important one. But when Humboldt (Mill and Emerson) talk about developing one's individuality as a kind of virtue of character they mean something specific with the term. The idea involved in individuality here involves notions such as 'originality', as can be seen in the quote below:

"This individual vigor, then, and manifold diversity, combine themselves in originality; and hence, that on which the consummate grandeur of our nature ultimately depends, — that towards which every human being must ceaselessly direct his efforts ... is the individuality of power and development" (Humboldt 2014, p. 13).

The thought, if one follows this track, would then be that among the things that are required in order to develop into the ideal version of



oneself we would find the requirement to develop one's uniqueness and originality. I would like to call attention to the fact that it is at this point that we really have the idea, which I introduced when explaining the contrast between the free person and the slave, of someone being their own person, being authentic in that sense. This is the point which seems to be characteristic of the idea of freedom, which is a kind of uninhibited flowering.

The thing about this thought about uniqueness and originality in Humboldt's view is that he associates it with a sort of individual power (as illustrated in the quote above), a personal energy that is required for the process of developing into the ideal version of oneself. We can actually find the same sort of view in Mill, Emerson, and even Rousseau (though in his case more rarely), but in Humboldt it takes a central role. In Humboldt's words, this idea of one's individual power (or energy) and its proper cultivation:

"... appears to me to be the first and chiefest of human virtues" (Humboldt 2014, p. 100).

Humboldt gives another interesting fleshing out for the idea of individuality. Rather than having the individuality in a kind of necessary opposition to what is externally presented to the person, he actually thinks it can be beneficial to the development of one's individuality provided it is properly appropriated:

"Now, whatever man receives externally, is only as the grain of seed. It is his own active energy alone that can convert the germ of the fairest growth, into a full and precious blessing for himself. It leads to beneficial issues only when it is full of vital power and essentially individual" (Humboldt 2014, p. 15).

If one follows Mill's track there is a way of fleshing out the idea of the need to develop one's uniqueness from the perspective of the threat to

individual freedom presented in the section on negative freedom. Mill points out that:

"... individual spontaneity is hardly recognized by the common modes of thinking, as having any intrinsic worth, or deserving any regard on its own account" (Mill 1991, p. 63).

Such lack of acknowledgment of the importance of developing one's individuality becomes particularly problematic for two reasons which I have explained in the section on negative freedom. The first comes from the pressures towards uniformity of thought and behaviour. The idea here would be that the lack of acknowledgment in question makes room for some of that pressure. The second reason, which makes the connection with positive freedom, is the need to develop one's uniqueness as necessary for one's ideal development. It is in this spirit that Mill claims:

"That so few now dare to be eccentric, marks the chief danger of the time" (Mill 1991, p. 75).

Mill points to a kind of inconsistency about the resistance found in society towards the development of the person's uniqueness:

"... surprising as it may be ... to find so high a value attached to individuality, the question, one must nevertheless think, can only be one of degree. No one's idea of excellence in conduct is that people should do absolutely nothing but copy one another. No one would assert that people ought not to put into their mode of life, and into the conduct of their concerns, any impress whatever of their own judgement, or of their own individual character" (Mill 1991, p. 63).

But such apparent acceptance, according to Mill, only happens in theory. In the practice of what developing one's originality involves, it is far from being really accepted in society:

"I insist thus emphatically on the importance of genius, and the necessity of allowing it to unfold itself freely both in thought and in practice, being well

aware that no one will deny the position in theory, but knowing also that almost every one, in reality, is totally indifferent to it. People think genius a fine thing if it enables a man to write an exciting poem, or paint a picture. But in its true sense, that of originality in thought and action, though no one says that it is not a thing to be admired, nearly all, at heart, think that they can do very well without it" (Mill 1991, p. 72).

The relevance of developing one's originality and uniqueness is acknowledged not only by Humboldt and Mill, but also by Emerson as can be seen in the following quote:

"The power which resides in him is new in nature, and none but he knows what that is which he can do, nor does he know until he has tried. Not for nothing one face, one character, one fact, makes much impression on him, and another none" (Emerson 1841).

Emerson's perspective on the importance of originality takes us back to the idea that this kind of personal power is developed by being put into practice, a thought constantly acknowledged by Mill. In the sense of allowing room for that to happen Mill introduces the notion of one's 'plan of life':

"He who lets the world or his own portion of it, choose his plan of life for him, has no need of any other faculty than the ape-like one of imitation. He who chooses his plan for himself, employs all his faculties. He must use observation to see, reasoning and judgment to foresee, activity to gather materials for decision, discrimination to decide, and when he has decided, firmness and self-control to hold to his deliberate decision" (Mill 1991, p. 65).

Given Mill's richer-genius conception of the self, the idea of one choosing and living in accord with their own plan of life is understood in terms of a plan of life that reflects their own particular genius. This is what the quote below suggests:

"To give any fair play to the nature of each, it is essential that different persons should be allowed to live different lives" (Mill 1991, p. 70).

Having a plan of life – which is something like the overall shape that one's practical life takes – that reflects one's particular genius gives in Mill the appropriate conditions for one's development into the ideal version of themselves. The analogy with the growing conditions of plants in the quote below has precisely this sense:

"If it were only that people have diversities of taste, that is reason enough for not attempting to shape them all after one model. But different persons also require different conditions for their spiritual development; and can no more exist healthily in the same moral, than all the variety of plants can in the same physical, atmosphere and climate. The same things which are helps to one person towards the cultivation of his higher nature, are hindrances to another. The same mode of life is a healthy excitement to one, keeping all his faculties of action and enjoyment in their best order, while to another it is a distracting burthen, which suspends or crushes all internal life. Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable" (Mill 1991, pp. 75-6).

## **Agent Particularism and Virtue Ethics**

### 1. Introduction:

Before we go on to the serious business of explaining the Agent-Particularist conception of human dignity it may be a good idea to explicitly situate Agent Particularism in relation to a more traditional virtue ethics. The aim here will be simply to mark up what is similar and what is different in the two views. We will see that some of the same sorts of arguments apply to both.

In the discussion of positive freedom in Chapter Two I introduced my Agent-Particularist form of virtue ethics, as I explained that once one has a richer-genius conception of the self the process of positive freedom requires virtues of character that, along side with the virtues of character more widely accepted by virtue ethicists, are essential to one's proper development. These are virtues of character which more directly involve the proper development of one's unique genius. I would like to now turn to a more explicit and broader discussion on a kind of virtue ethics which turns out to be necessary if one is coming from a richer-genius conception of the self.

At this outset of the chapter I would like to make a distinction between certain positions within ethics, as I will refer to them later and it is important that we are clear about what I am referring to in each occasion. The position I am proposing in this thesis is what I am calling Agent Particularism. Agent Particularism can be understood as the denial of the principle of universalizability of moral judgments. The universalizability principle states that who the agent is is irrelevant to determining what is the right thing for them to do.

Agent Particularism on the other hand accepts that who the person is is ethically relevant. This means that in relevantly similar circumstances the right thing to do might differ for different people. What justifies such differences, according to Agent Particularism, are differences in people's particular geniuses. In rejecting the universalizability principle Agent Particularism distinguishes itself from what I refer to in this thesis as Situation Particularism<sup>23</sup>, commonly known as moral particularism.

Situation Particularism can be understood as a rejection of what I will refer to in this chapter as Situation Universalism. Situation Universalism is the view according to which the business of ethics is to come up with universal principles of behaviour, be it one (like in Kant, which we looked at in detail in the previous chapter) or a set of universal principles. Situation Particularism rejects Situation Universalism by claiming that the right thing to do depends on the circumstances, in the sense that variations in the surrounding circumstances or contingent facts about the agent might make a moral difference. But for this point to count as a proper rejection of certain forms of Situation Universalism (the ones that do take into account variations of circumstances in order to state which universal principle is applicable in each occasion<sup>24</sup>) it requires a further thought. The key idea behind Situation Particularism that makes it a proper rejection of even this kind of Situation Universalism is that the differences in circumstances which might be morally relevant can't be codified. That basically means that it is not possible to fix in one or more principles all that may be a morally relevant variation on circumstances.

The particular kind of Agent Particularism which I am developing also rejects Situation Universalism. But unlike most forms of Situation Particularism its focus is not on the moral relevance of variations in

---

<sup>23</sup> I quickly introduced Situation Particularism on Chapter One.

<sup>24</sup> For an example of this kind of position see David Ross, "The Right and the Good".

circumstances. The Agent-Particularist claim I am holding is that not only variations on circumstances may be morally relevant, but also variations on people's natures may be morally relevant. What makes a sharp contrast between Agent Particularism and Situation Particularism is that a Situation Particularist might accept the universalizability principle – and in so doing deny Agent Particularism.

## 2. Agent Particularism as a Kind of Virtue Ethics:

As I explained in Chapter Two the notion of an ideal development of oneself is essential to Agent Particularism, otherwise it wouldn't be a plausible rejection of the universalizability principle. The notion of an ideal development of oneself is also the core idea of virtue ethics. According to virtue ethics what one has to do is to develop into an ideal version of oneself, because it is in this way that one becomes able to respond properly to the demands of circumstances on them. Therefore the notion of the ideal development of oneself is essential to both traditional virtue ethics and Agent Particularism.

But I want to make, from this beginning, a distinction between virtue ethics as traditionally understood and Agent-Particularist virtue ethics (a point I anticipated in Chapter Two). The traditional virtue ethicist that I have in mind is a situation particularist but also accepts some form of the universalizability thesis (I will explain this point in detail at the last section of this chapter). This means that traditional virtue ethics and Agent-Particularist virtue ethics find themselves on distinct sides when it comes to the universalizability principle.

As I said the notion of the ideal development of the agent is common to both Agent Particularism and traditional virtue ethics. But in general traditional virtue ethicists, because they are coming from a blank

conception of the self, are also universalists in the sense that they think the ideal version of each person will be the same. For the Agent Particularist on the other hand the ideal version of different people will differ according to differences in their particular geniuses.

There will be another important difference between the traditional kind of virtue ethics and Agent Particularism that derives from this. The traditional virtue ethicist thinks that there is the virtuous person, which is a kind of perfect human being. So it is not just that for the traditional virtue ethicist the ideal version of everyone is the same. Traditional virtue ethics wants each person to strive towards an ideal of perfection that isn't really reachable by real people. There is no such thing as the ideal person. I will explain this point in detail at the end of the chapter, and we will see then why there is something special about Agent-Particularist virtue ethics.

Having said that I want to turn to presenting certain features of a traditional approach to virtue ethics, namely McDowell's.

### 2.1. Agent Particularism and Uncodifiability:

In Chapter One as I said earlier I have rejected the principle of the universalizability of moral judgements. This rejection is the essence of the Agent-Particularist position. But the possibility is open to the Agent Particularist for also being at the same time a Situation Particularist, and that is the position which the particular kind of Agent Particularism which I am developing in this chapter adopts. Nonetheless my Agent Particularism will develop Situation Particularism in its own distinct way, as I will show in this section.



We should recall that the Agent Particularist holds that who one is makes a difference to what one should do. The situation Particularist holds that what one should do depends upon the surrounding circumstances. That is something that is important for the Agent Particularist also to accept. So even once it is fixed who one is there is still no simple rules for saying what one should do.

In chapter one I explained that there are two main roots in contemporary ethics for Situation Particularism, McDowell's and Dancy's. The kind of Situation Particularism I am aligning myself with is a McDowellian rather than the Dancy one, therefore in this thesis I won't go any deeper into Dancy's arguments. I will nonetheless look at McDowell's arguments in detail. McDowell's rejection of Situation Universalism takes him straight into virtue ethics. As I will explain in detail later, the key basis of McDowell's approach to virtue ethics is that it solves an epistemological problem which no other approach to ethics does. It doesn't solve a practical epistemological problem - namely how we work out what to do - but a problem of pure epistemology: how can there be moral knowledge given that there is no theory (once Situation Universalism is rejected)? Since my kind of Agent Particularism also rejects Situation Universalism it is faced with this same problem, which makes McDowell's version of virtue ethics particularly helpful for my kind of view.

For McDowell, virtue ethics is the view you are forced into if you accept two things. First, that virtue is a kind of knowledge (or wisdom). Second that ethics is uncodifiable, i.e. there can't be simple universal prescriptions which will provide a recipe for doing what is right. Given those two points the kind of knowledge which virtue will have to be must be a matter of being sensitive to situations (as we will see in the next sub-section).

McDowell's (and in general the Situation Particularist's) motivation for rejecting codifiability comes from the fact that life's situations are too various to be captured in generalizable principles, which he takes to be quite a clear point:

"... to an unprejudiced eye it should seem quite implausible that any reasonably adult moral outlook admits of any such codification" (McDowell 1979, p. 336).

The Situation Particularist's sort of motivation for uncodifiability is already applicable from the perspective of my Agent Particularism. It is a point of connection between my Agent Particularism and Situation Particularism. Yet as I will show later Agent Particularism accepts this kind of motivation for even deeper reasons, and adds to it a further one. So let's look at the Situation Particularist's motivation for uncodifiability in detail.

Following Aristotle, McDowell makes a straightforward point about it:

As Aristotle consistently says, the best generalizations about how one should behave hold only for the most part" (McDowell 1979, p. 336).

McDowell's point is not to deny that a sophisticated set of rules of behaviour could be assembled by clever philosophers. Neither is it to deny that refined sub-categories of rules could be put together - so if we have got a rule such as 'it is wrong to harm an innocent person', we add a clause like 'provided the situation causes no harm to the agent themselves' (I will come back to this point later). It is rather to say that universal principles of behaviour of whichever quality will eventually not help the agent in sorting out what is the right thing for them to do. Now, a moral theory that purports to guide one's actions in every occasion but in reality isn't able to do the needed job just can't be the one we rely upon. The reason is it would lead the agent to mistakes about the ethical demands of situations on them:

"If one attempted to reduce one's conception of what virtue requires to a set of rules, then, however subtle and thoughtful one was in drawing up the code, cases would inevitably turn up in which a mechanical application of the rules would strike one as wrong ..." (McDowell 1979, p. 336).

Before going deeper into the Situation-Particularist motivation for uncodifiability let's look in detail at what uncodifiability means. The codifiability thesis states that all true moral judgements can be codified into universal principles. Key in this view is the idea of a universal principle, which in ethics has a certain structure. Firstly, the principle is universal, i.e. it is expressed through a universal quantifier (such as 'all', 'every', 'any', etc). This means that it applies universally every time that certain conditions are met. We have for example a principle of the form: all Xs are Y. According to a principle of this form every time we have an X the appropriate judgement is that it is Y.

Another key element of such universal principles is that they are supposed to provide a bridge between descriptive, non-moral features of situations and moral judgements. One example is the principle 'You should always help the needy'. It is a principle with this basic structure: everything which has this non-moral property - in this case being needy - has this moral property, namely being such that you should help them. Another example is the principle 'Killing is always wrong'. 'Killing' describes a situation non-morally. We can see if someone is alive or dead without that involving any value, i.e. independently of engaging in morality. According to Situation Universalism that non-moral property always issues in a moral property, i.e. it being wrong.

Richard Holton presents the following basic formula of universal principles: "Under descriptive conditions D the correct moral verdict is V" (Holton 2002, p. 192). And he explains it:

"... if D were a complete description of a situation (including the claim that it is a complete description!), and if V were the correct verdict in that situation,

then the corresponding principle would be universally true: whenever D were realized, V would be the correct verdict" (Holton 2002, p. 192).

Holton wants to embrace Situation Particularism (which means he rejects the codifiability thesis) at the same time that he does not deny the existence of universal principles. He accepts the view that whenever there are descriptive conditions D the appropriate moral judgment is V. But his acceptance of the existence of universal principles doesn't mean he accepts any form of principle. He rather accepts principles of the basic structure he gives only in the case of D being a complete description of a situation. This means that according to Holton there are, *formally speaking*, universal principles (I will come back to this point in a moment).

The existence of a formal principle with such structure is something which I have already accepted in Chapter One. What I mean is that I accepted the view that we ought to judge two relevantly similar people in relevantly similar circumstances in the same way. This is one way of formulating the supervenience thesis, which I endorsed and explained in Chapter One. In fact this is precisely what makes room for consistency in moral judgments. As Holton points out, in general the Situation Particularist accepts supervenience:

"... particularists typically endorse the supervenience of the moral on the descriptive. That is, they accept the highly plausible thesis that any two situations that are identical in their descriptive properties will be identical in their moral properties" (Holton 2002, p. 192).

Since Holton accepts some form of universal principles his denial of codifiability obviously doesn't happen through his denying the existence of principles altogether. In order to clarify what his rejection of codifiability consists in let's look at a distinction he makes:

- a) "[ $\exists X$ : X is a finite set of true moral principles][ $\forall y$ : y is a moral verdict] y is entailed by X and the non-moral truths;

and

b)  $[\forall y: y \text{ is a moral verdict}][\exists X: X \text{ is a finite set of true moral principles}]$   
 $y$  is entailed by  $X$  and the non-moral truths" (Holton 2002, p. 195).

We can distinguish between "a)" and "b)" through a parallel distinction between the ideas that 'there is somewhere where all the roads lead' and that 'every road leads somewhere'<sup>25</sup>. Basically the claim in "a)" is that there is a specific and finite set of universal principles from which all true moral judgements can be derived. In this case the task of the moral agent would be to first learn the principles and then go on to apply them in life's situations. The principles one has learned would then serve as a guide to the agent's actions. This is the claim that the codifiability thesis wants.

The claim in "b)" is that for every true moral judgment there is a finite set of universal principles from which it can be derived. It involves no idea of learning the principles first and then applying them to the situation. This claim has to do with the justification of actions and not, as the Situation Universalist has it, with determining which action the agent should take. This is the sense in which my Agent Particularism accepts formally the existence of universal principles. This is also the sense in which Holton endorses such principles:

"Principles will not provide us with anything like a decision procedure for telling whether an action is right. ... principles play an important justificatory role" (Holton 2002, p. 196).

The distinction above introduces a new feature to the codifiability thesis. The mere formal existence of such principles isn't enough for codifiability. In Holton's view the full version of the codifiability claim also includes the idea that there is a finite set of universal principles

---

<sup>25</sup> This distinction was taken from Professor Michael Morris' handout for the course "Epistemology" held in 2017 at the University of Sussex.

from which every true moral judgement can be derived. In order to reject codifiability, Holton observes, what is needed is:

“... denying that any set of principles can serve to capture ethics” (Holton 2002, p. 192).

Though there are formally speaking universal principles, they are of a very specific sort. The sole sense in which it is possible for there to be formally speaking a principle of the form “Everything which has this non-moral feature has that moral feature” is if the non-moral feature is defined in completely global terms. That is all that is required by the supervenience of the moral on the non-moral, which means just that there can’t be a moral difference in the total situation without a non-moral difference. Holton’s way of doing it is by adding to each moral argument a premise that plays this role, which he refers to as ‘That’s it’:

“There are no further relevant moral principles and non-moral facts; i.e. there is no true moral principle and set of *true* non-moral sentences which supersede those which appear in this argument” (Holton 2002, p. 199).

In the context of an argument ‘That’s it’ always makes a claim that is specific to that argument. It says there are no other relevant features; all that is relevant is already stated in the argument. To come back at a previous example, let’s imagine that one meets a needy person and there is a principle that says one should help the needy and ‘That’s it’. The key thing that ‘That’s it’ does is it says that nothing about one’s present circumstance matters apart from that this is a needy person. There is nothing else that is relevant. There could be some other feature that is relevant like helping the needy in this case would put someone else in a cruel situation, for whatever reason. But ‘That’s it’ excludes any such kind of possibility. This is what it means for the non-moral features of a situation to be defined in completely global terms.

When we look at Holton's 'That's it' carefully we notice an aspect of it which is a point of contrast with the codifiability thesis. What codifiability requires is a way of getting from non-moral descriptive features of situations into moral judgements. But 'That's it' states that there are no further relevant moral principles and non-moral facts, which causes an issue here. We can see the point clearly when Holton tries to produce the kind of principle he accepts:

"P1 This is a killing

P2  $\forall x ((x \text{ is a killing} \ \& \ \text{That's it}) \rightarrow \text{you shouldn't do } x)$

P3 That's it

C You shouldn't do this" (Holton 2002, p. 199).

Holton presents in P2 the principle "For every X if X is a killing and 'That's it', then you shouldn't do X". But the thing is that 'That's it' isn't a non-moral claim. It says nothing else is relevant but Holton means morally relevant. That is not a principle of the relevant sort required by codifiability.

Holton's principle is illuminating because it shows that in trying to provide a principle which is plausible we have to bring in the moral perspective. That is precisely the point that McDowell makes in his rejection of codifiability. It is not clear that you can judge what is morally relevant from a non-moral perspective. What is required for supervenience is that if everything were the same non-morally then it would be the same morally, but that is not enough to guarantee that there is a usable principle:

"Supervenience requires only that one be able to find differences expressible in terms of the level supervened upon whenever one wants to make different judgements in terms of the supervening level. It does not follow from the satisfaction of this requirement that the set of items to which a supervening term is correctly applied need constitute a kind recognizable as such at the level supervened upon" (McDowell 2002, p. 202).

Given supervenience there must be some principle of the form "everything which is X is Y" where "X" is a non-moral term and "Y" is a moral term. Formally speaking there must be some kind of global principle, but not a graspable one because we don't know which descriptive features have to be the same. That is not something which can be determined non-morally. The view that morality is codifiable is not just the view that there are universal principles of the kind I have explained earlier. It also involves accepting that you can grasp the non-moral features that are supposed to give substance to the principle in order to arrive at a moral judgment. In effect what McDowell rejects is the possibility of there being any principles in which descriptive features are collected together on the left in a graspable way with no view to morality at all, and on the right you get some moral judgement. It is a rejection of any principle which might give one a practical recipe for acting.

The reason is that such principles would presuppose that a non-moral perspective is decisive in determining whether something is good or bad for instance. And McDowell's resistance to codifiability comes from the view that given that the concerns of the moral perspective and of the non-moral perspective are so different it would be extraordinary if the non-moral perspective delivered the right answer for the moral perspective. In this sense supervenience isn't enough to give us codifiability. McDowell continues with the explanation:

"... however long a list we give of items to which a supervening term applies, described in terms of the level supervened upon, there may be no way, expressible at the level supervened upon, of grouping just such items together. Hence there need be no possibility of mastering, in a way that would enable one to go on to new cases, a term that is to function at the level supervened upon, but is to group together exactly the terms to which competent user's would apply the supervening term. Understanding why just those things belong together may essentially require understanding the supervening term" (McDowell 2002, p. 202).



There is a sense in which the moral features follow from the non-moral features of situations, but there is no principled way of selecting the non-moral features that are relevant. To illustrate the point let's go back to the principle 'All killing is bad'. Being alive and being dead are categories that make sense independently of engaging in morality.

Now, suppose one tries to qualify the principle, and instead of saying all killing is bad one says 'all killing which X is bad' (we can substitute the X for different further conditions, such as 'isn't in self-defence'). And we can imagine one then trying to lay out a whole series of complicated further conditions. But why are those further conditions there? There is no way of grouping them together meaningfully unless they are understood just as the further conditions which were relevant for making this moral judgment. One can complicate the principle, but when we think about it we can see that there is no limit to the complications. The only way of stopping the complication is to say something like 'we are just talking about the features which are morally relevant'. And the problem is that this is something one cannot get at non-morally.

Codifiability aims for a single set of principles from which all moral judgements follow, and the principles have to follow this particular structure that you can use them as a recipe to get from a non-moral conception of a situation to a moral judgement. Holton is right in thinking that this is what codifiability requires, and he is also right in pointing out that formally speaking it doesn't rule out it being the case that for every moral judgement there is a principle from which it follows. But the interesting thing is that when Holton tries to produce the principle he provides further evidence to a deep reason as to why one might reject codifiability. A deep reason that depends on the fundamental difference between the moral and the non-moral perspectives.

What this all means is that McDowell's rejection of codifiability is reinforced by Holton's 'That's it'. It gives further support to McDowell's argument that one is not going to be able to single out, to make the right discriminations from a non-moral perspective. At the moment in which one tries to get a plausible principle they are forced to bringing in the moral perspective. The thought is that it needs moral judgement to determine what is morally relevant, what McDowell calls the salient facts about the situation:

"If there is more than one concern which might impinge on the situation, there is more than one fact about the situation which the agent might, say, dwell on, in such a way as to summon an appropriate concern into operation. It is by virtue of his seeing this particular fact rather than that one as the salient fact about the situation that he is moved to act by this concern rather than that one" (McDowell 1979, p. 344).

The agent's attention will be called to different facts of situations, and there will need to be a way of determining which among those facts are the morally relevant ones in each case. McDowell sees virtue as a kind of sensitivity as we will see in detail in the next subsection. And the virtuous person is the one that possesses the appropriate sensitivity to the relevant facts. The key feature of the virtuous agent is precisely that they can tell what is morally relevant, in virtue of being sensitive to the right features of a situation. The right features of the situation become salient to them. It is that perception of salience which is the distinctive characteristic of the virtuous person. Developing the appropriate sensitivity involves developing the capacity to see things from a moral perspective:

"... we can equate the conceptual equipment which forms the framework of anything recognizable as a moral outlook with a capacity to be impressed by certain aspects of reality" (McDowell 1979, p. 347).

As I mentioned earlier the McDowellian argument for uncodifiability is already compelling. But why should uncodifiability seem particularly

compelling if one is an Agent Particularist? Codifiability is already very implausible with the view that moral demands vary in the situations. It becomes even less plausible if they vary with the person as well. It looks more unlikely that we are going to have any kind of coincidence that justifies codifiability.

We can illustrate this point with an argument against codifiability made by David Wiggins<sup>26</sup>. Ethics reflects what matters to us as human beings. Unless we think that the world is made for human beings we shouldn't expect there to be a description of the world as it is independently of us to match the way it makes demands on us. But the world was not made for human beings, i.e. its divisions and categories are not shaped to suit our concerns. Since the world was not made for us it would be miraculous if descriptions designed to suit the world matched what matters to us. The world doesn't select what is morally relevant, which is what codifiability would entail.

The argument becomes even stronger when it is not just about our concerns as human beings, but also about the concerns of us individually. What matters for each one of us would be dependent upon their individual nature. If it would be miraculous for descriptions designed to fit the world to match what matters for us as a species, how much more miraculous would it be if there were not general moral truths applied to all human beings but an appropriate set of principles for each person? What we would have to suppose is that what is the universal law for one person is quite different from what is the universal law for another. Having codifiability in these conditions would mean having a kind of recipe that is designed for each person built into the structure of reality.

A similar kind of point can be made in terms of McDowell's argument for uncodifiability. The key thing in McDowell's argument is the idea

---

<sup>26</sup> Reported in conversation with Professor Michael Morris.

that the concerns which are involved in morality are different from those which are involved in non-moral description of situations. For this reason one can't from a non-moral perspective make a moral judgment - I explained this point while distinguishing between Holton and McDowell. What we can see from Holton is that at the moment in which we try to come up with a principle we must have moral judgement as a way of selecting from the features of a situation which ones are morally relevant (this is what prevents the principle from being of the kind which is relevant for codifiability).

This point takes a character of its own when seen from the perspective of Agent Particularism. For McDowell virtue is a kind of sensitivity. Developing virtue involves developing the moral outlook that makes the agent sensitive to the relevant features of a situation. But McDowell seems to see one same kind of ideal development for all agents. It issues in one kind of moral perspective for all agents irrespective of the differences between them. For the Agent Particularist each person's ideal development varies according to their nature. The moral perspective of each agent will also vary according to their nature. If a codification is already implausible if one thinks in terms of a general moral perspective, it makes even less sense if there is not just one moral perspective but as many moral perspectives as there are people.

The points I have just made about Wiggins and McDowell are distinct from one another. Wiggins thinks there is the way the world is and then there are human concerns. McDowell rather has a different set of concerns: what is available from a non-moral perspective, and the moral perspective. But the structure of the argument in both cases is the same. The point in Wiggins is that we shouldn't expect things, from the perspective of the world, to match what matters to human beings. And it would be even more miraculous for the perspective of the world to match what is morally relevant for each person when that would be different from person to person. In the case of McDowell one wouldn't

expect non-moral concerns to pick up what was important for morality, and in the same way even more so one wouldn't expect non-moral concerns to pick up what was important just for each particular individual morally.

The general point which is behind both of these arguments becomes even more vivid when we think in terms of moral development. Let's suppose for a moment that developing into a virtuous person, i.e. developing into someone that is properly responsive to the demands that situations make on them is a matter of learning a set of absolutely general laws. Once we are coming from a richer-genius conception of the self this gives us a view of people's moral upbringing according to which each person would have to learn a set of absolutely general rules that may differ radically from everybody else's set of rules. Now, it is really implausible that coming to be virtuous could ever be a matter of learning absolutely general laws in these circumstances.

This point brings out the particular character that the rejection of Situation Universalism takes when seen from the perspective of Agent Particularism. If an Agent Particularist has even stronger reason to accept Situation Particularism, any attempt to enforce a Situation Universalism will be not just philosophically mistaken, but also an imposition on the person. The implausibility of there being a set of universal rules suitable for each person means that the insistence on a universal set of rules turns out to be a demand for conformity. Insisting on universal principles in this context is an insistence on denying the freedom of each person to develop in their own way. As I said the demands of Situation Particularism have a different kind of character from an Agent-Particularist point of view. The rejection of Situation Universalism becomes the defence of the negative freedom of the agent, the rejection of an imposition.

It is worth noting that this view goes in line with the existentialist tradition in its way of treating universalist rule-following morality as a restriction on people's freedom. For instance, as Steven Crowell points out, both Søren Kierkegaard and Frederick Nietzsche emphasized individuality and saw universalism as a coercion on that individuality:

"Both were convinced that this singularity, what is most my own, "me," could be meaningfully reflected upon while yet, precisely because of its singularity, remain invisible to traditional philosophy, with its emphasis either on what follows unerring objective laws of nature or else conforms to the universal standards of moral reason" (Crowell 2016).

As I have shown in the previous chapter for the Agent Particularist the value of negative freedom comes from positive freedom. So negative freedom is a necessary step towards the ideal development of oneself. Since the Agent Particularist is coming from a richer-genius conception of the self, developing the perceptual capacity of the virtuous person comes out as a re-description of the positive development of one's own ideal self which I described in Chapter Two. So becoming moral takes on the character of the development of positive freedom.

## 2.2. Virtue as Sensitivity to the Demands of Situations:

At the end of Chapter Two, in the section on positive freedom, I presented certain virtues of character that are peculiar to the Agent-Particularist project, or at least are applicable if one holds a richer-genius conception of the self. I now want to use McDowell in order to complement the view I have presented in Chapter Two. Besides those virtues of character required in the process of positive freedom Agent Particularism, when developed as a kind of virtue ethics, also includes the virtues of character that allow the agent to develop a sensitivity to situations.

As I said earlier McDowell's account of virtue as sensitivity and as knowledge gives us the solution to an epistemological problem raised in the previous subsection: how can what is involved in ethics be knowledge given that there is no theory? It also complements my Agent-Particularist view on people's personal development.

The key thing is in the justification of moral claims. The model I have rejected grounds the status of moral claims as counting as instances of knowledge in the application of principles. Moral development would be according to this model a matter of learning principles. But it is absurd to think of moral development in terms of learning a distinct set of universal laws for each person. So the justification one has for making moral claims has to be of a different character from the sort that the Situation Universalist thinks. McDowell's account of the virtuous person offers precisely that.

For McDowell, virtue ethics is the natural view to take once one accepts both uncodifiability and that virtue is a kind of knowledge. Given those two points, the kind of knowledge which virtue will have to be must be a matter of being sensitive to situations:

"The deliverances of a reliable sensitivity are cases of knowledge" (McDowell 1979, p. 332).

For McDowell the knowledge status of moral claims is justified by the deliverances of a perceptual capacity. Such a capacity is built up over time as one develops virtues. A moral statement counts as a knowledgeable statement because it is the deliverance of a knowledgeable capacity. The identification of virtue with knowledge comes from the idea that knowing something involves getting it right. According to McDowell the virtues can be identified with knowledge as they are precisely:

"... states of character whose possessor arrives at right answers to a certain range of questions about how to behave" (McDowell 1979, p. 331).

For McDowell virtues are a kind of perceptual capacity that allow the agent to grasp the demands of the world on them. According to him to have a virtue is to have:

"... a reliable sensitivity to a certain sort of requirement which situations impose on behavior" (McDowell 1979, pp. 331-2).

We can illustrate the point with specific virtues. If we take courage to be a virtue, to have it would be to be sensitive to instances in which the situation demands a kind of courageous act from the agent. In the same way:

"A kind person can be relied on to behave kindly when that is what the situation requires" (McDowell 1979, p. 331).

So to have a virtue like kindness is to have a sensitivity to certain features of situations that singles them out as requiring kindness as a response. And a kind person knows what it is to be confronted by those features. But for this model to work as justification of moral statements it must be the case that the perceptual capacity arrives at right answers. It needs to be able to single out the morally relevant facts regardless of which specific sensitivity is demanded in each occasion. Therefore the development of the perceptual capacity to be sensitive to the morally relevant facts of situations involves developing all of the virtues. To keep McDowell's example, acting kindly if that doesn't involve also being fair in the situation isn't a way of responding properly to the demands of the situation:

"... we cannot disentangle genuine possession of kindness from the sensitivity which constitutes fairness. ... we use the concepts of the particular virtues to mark similarities and dissimilarities among the manifestations of a single sensitivity which is what virtue, in general, is: an ability to recognize requirements which situations impose on one's behavior" (McDowell 1979, p. 333).



I would like to note at this point that in my Agent-Particularist virtue ethics the development into a virtuous person takes a shape of its own. It involves developing according to one's particular genius. And if the perceptual capacity is to give right answers to the question of what is the right thing to do it must acknowledge the particularity of each person.

This anticipates a point which I will further develop in the last section of this chapter. The Agent Particularist way of educating yourself involves trying to become the ideal version of yourself. The way you do that is by considering what kind of person would it be best for there to be in a whole range of situations, and the concepts of the traditional virtues pick out something like the kind of person you want there to be in the different situations. The traditional virtue ethicist wants you to make yourself into the person that we would want there to be in every situation. And so they imagine that the ideally developed virtuous person will possess all the virtues and ideally to their limit. But the Agent Particularist recognizes that not everyone can do everything, and that makes a difference to the way in which the Agent Particularist understands the unity of the virtues thesis (introduced in McDowell's quote above). For Agent Particularism although the development of each virtue can't be independent of the development of the others, what it is to develop each one may be different for different people (there are, for instance, differences in degree). As I said I will come back to this point in the later.

Agent Particularism rejects Situation Universalism and with it codifiability. That brings a challenge to the view, which is explaining how moral judgments can count as instances of knowledge. The natural solution is the one which McDowell suggests, that is, to think of knowledge as the deliverances of a perceptual capacity. This gives us

a view of moral knowledge according to which it comes from the possession of a perceptual capacity whose deliverances are not justified by inference from general principles. What we need to do in this case is to make sense of accepting that the deliverances of this capacity can be justified without it being possible to provide a kind of independent certification of that justification.

It can seem to be frightening to have a kind of judgement which is justified but whose justification isn't externally certifiable. This is nonetheless what there is. In order to make the point that there is no need for concern about it McDowell draws a parallel between Wittgenstein and rule-following. In the parallel the idea that the justification of knowledge must happen through inference from universal principles comes under attack. McDowell explains the point:

"We tend to picture the understanding of the instruction "Add 2" - command of the rule for extending the series 2,4,6,8, ... - as a psychological mechanism which, aside from lapses of attention and so forth, churns out the appropriate behaviour with the sort of reliability which a physical mechanism, say a piece of clockwork, might have. ... the evidence for the presence of the pictured state is always compatible with the supposition that, on some future occasion for its exercise, the behaviour elicited by the occasion will diverge from what we would count as correct. Wittgenstein dramatizes this with the example of the man who continues the series, after 1000, with 1004, 1008.... If a possibility of the 1004, 1008, ... type were to be realized (and we could not bring the person to concede that he had simply made a mistake), that would show that the behaviour hitherto was not guided by the psychological conformation which we were picturing as guiding it. The pictured state, then, always transcends the grounds on which it is allegedly postulated" (McDowell 1979, p. 337).

In following a rule we tend to be confident in our success in the application of the rule based on the assumption that a psychological mechanism that functions like the gear of a machine is underneath guiding our judgment. But McDowell makes the point that there is nothing underneath grounding the judgment. There isn't in the case of mathematics any more than there is in ethics. If that is the case, what is the ground of our confidence that we will get it right? McDowell

quotes Stanley Cavell, in reference to the competent use of words, in order to answer that:

"We learn and teach words in certain contexts, and then we are expected, and expect others, to be able to project them into further contexts. Nothing insures that this projection will take place (in particular, not the grasping of universals nor the grasping of books of rules), just as nothing insures that we will make, and understand, the same projections. That on the whole we do is a matter of our sharing routes of interest and feeling, modes of response, senses of humour and of significance and of fulfilment, of what is outrageous, of what is similar to what else, what a rebuke, what forgiveness, of when an utterance is an assertion, when an appeal, when an explanation - all the whirl of organism Wittgenstein calls "forms of life." Human speech and activity, sanity and community, rest upon nothing more, but nothing less, than this. It is a vision as simple as it is difficult, and as difficult as it is (and because it is) terrifying" (McDowell 1979, pp. 338-9).

There is nothing underneath guiding our competent use of words. In just the same way as there is this sense of giddiness – a sense of lack of support - in terms of our knowledge about languages that is also the case in terms of knowledge in general. This causes a sense of vertigo, after all we might consider: What is it depending on? But McDowell's point is that the most basic forms of knowledge always have to have this character. And if that is right in general there is nothing to be especially worried about in the case of moral judgements. McDowell goes on with his explanation:

"The terror of which Cavell speaks at the end of this marvellous passage is a sort of vertigo, induced by the thought that there is nothing but shared forms of life to keep us, as it were, on the rails. We are inclined to think that that is an insufficient foundation for a conviction that when we, say, extend a number series, we really are, at each stage, doing the same thing as before. ... we feel we have lost the objectivity of (in our case) mathematics (and similarly in other cases). We recoil from this vertigo into the idea that we are kept on the rails by our grasp of rules. ... This composite idea is not the perception of some truth, but a consoling myth, elicited from us by our inability to endure the vertigo" (McDowell 1979, p. 339).

The point becomes even more relevant if one is an Agent Particularist, and the pressures towards adopting the McDowell kind of view are accentuated in this case. The sense of vertigo becomes clearer since

not only there isn't external certification of the judgment but also each person will have to make judgments by themselves, without the possibility of something like copying other people.

### 3. The Traditional Virtue Ethicist as a Universalizability Theorist:

Agent Particularism accepts the idea of the ideal development of the person. Traditional virtue ethics also does. But as I explained earlier there is a basic distinction between the two views. While Agent Particularism works with a richer-genius conception of the self, traditional virtue ethics in general works with a blank conception of the self. For this reason, even though the notion of the ideal development of the person is common to both views they understand it in different ways.

According to Agent Particularism people are particular in what is essential to them, and that difference ought to be respected. The ideal development of the person happens in accord with their particular genius. That means that the ideal version of each person will also be particular to them. For traditional virtue ethics on the other hand the ideal version of each person will be the same. This means that traditional virtue ethics implicitly accepts universalizability. We can for instance see it implied in the following quote from Rosalind Hursthouse:

"If we think someone did something 'because he thought it was right' but find him condemning someone else for doing the very same sort of thing for no better reason than that he, uncomfortably, is on the receiving end, we say he couldn't have done the first thing for that reason" (Hursthouse 1999, p. 134).

Of course being uncomfortably on the receiving end is not a good reason for making a different judgement. But Hursthouse seems to

think this is the only kind of reason one could have for thinking something would be right for one person to do but not for another person. What this means is that Hursthouse is in effect saying that if I judge that 'X' is the right thing for me to do in a situation I would have to necessarily make the same judgement if someone else were to do 'X' to me in a relevantly similar situation. This is just another way of saying that in a relevantly similar situation 'X' is the right thing for everybody to do, i.e. the universalizability thesis. Though I said that in general traditional virtue ethics accepts universalizability implicitly some virtue ethicists are quite explicit about it. Raja Halwani is an example:

"Practical wisdom is the faculty that enables the agent to discern what is right to do, given the circumstances. ... any person, in sufficiently and relevantly similar circumstances ... would be justified in acting in the same way. And this, it seems to me, gives us the core and essential idea found in universalizability" (Halwani 2003, p. 61).

For him:

"... virtue ethics accommodates universalizability, for it is important to see that it can retain universalizability despite its own emphasis on the particular" (Halwani 2003, p. 61).

Agent Particularism can take hold of the general idea of virtue ethics but is distinct from it. McDowell - as other virtue ethicists - has as key in his moral conception the notion of the virtuous person. Implicitly that does issue in something like universalizability. According to him:

"... no one virtue can be fully possessed except by a possessor of all of them, that is, a possessor of virtue in general" (McDowell 1979, p. 333).

The above quote introduces McDowell's thesis of the unity of the virtues, a thesis which Agent Particularism understands in a distinct way (I anticipated this point earlier). According to McDowell's understanding

of this thesis the virtuous person is a kind of standard model of an ideally developed human being. He imagines that the ideally developed virtuous person will possess all the virtues and ideally to their limit. Since according to McDowell everyone ought to strive to become that model this implies that everyone is supposed to develop equally and fully in all the directions that the individual virtues point to. The problem is that this means everybody ought to be the same person – someone who is like the ideal model that McDowell calls the virtuous person.

At the moment one accepts this notion they have accepted some form of universalizability without accepting general rules. Even if what has to be done has to be decided situation by situation it is the same thing that has to be done by everybody (I will come back to this in a moment). If in this view everyone ought to act like McDowell's virtuous person, everybody ought to have the same ideal progression becoming exactly the same kind of person. That is what the notion of the virtuous person issues in.

Effectively in rejecting the unity of virtue thesis as formulated by the traditional virtue ethicist, Agent Particularism avoids a difficulty which the traditional virtue ethicist faces. Traditional virtue ethics lacks an explanation for the fact that there are the different virtues. If what each person really need to become is a single ideal character why not just concern ourselves with how that character would behave? Why do we need to know how the brave person would behave, how the wise person would behave, or how the self controlled person would behave?

Why do we need to know how these different kinds of people would behave? Traditional virtue ethics lacks an explanation for this but Agent Particularism is able to provide a plausible explanation. Agent Particularism sees these as kinds of ways in which each of us will develop. As I anticipated earlier, on my view the individual virtues are

more specific features of character some of which will be closer to a person than others. So it is a kind of extra virtue of the Agent-Particularist account that it explains why we have the different virtues.

In the traditional virtue ethicist's view everybody ought to do what the imagined virtuous person would do:

"An action is right iff it is what a virtuous agent would characteristically (i.e. acting in character) do in the circumstances" (Hursthouse 1999, p. 28).

In the quote above Hursthouse hints at a worry about conceiving the virtuous person. By concerning himself with what the virtuous person would do the virtue ethicist reasons as if there were a single character who is the virtuous person. And that character is some kind of version of the perfect human being:

"Possessing a virtue is a matter of degree. To possess such a disposition fully is to possess full or perfect virtue..." (Hursthouse and Pettigrove 2016).

Virtue ethicists imagine how an ideal virtuous person would behave, and insists that everybody ought to strive to behave like that ideal:

"The thesis I aspire to defend ... is that the perfectly virtuous agent, when she acts virtuously, sets the standard ... against which we assess the extent to which the less than perfectly virtuous do the same" (Hursthouse 1999, p. 145).

Traditional virtue ethicists have to believe in some kind of ideal person, but there is no such person. So even if it doesn't involve following rules the traditional form of virtue theory accepts some form of universalizability, because in the end on the view it is the same that is required from everyone. Now, since it is required that people behave like such ideal character it is also required that they strive to become that character:

"... our appeal to the ideally virtuous person makes sense because it is an appeal made to not just what the virtuous person would do but to the kind of person that the virtuous person is" (Athanasoulis 2013, p. 71).

According to the virtue ethicist that is so because it is by becoming the virtuous person that one knows what to do:

"Occasion by occasion, one knows what to do, if one does, not by applying universal principles but by being a certain kind of person..." (McDowell 1979, p. 347).

So on that sort of theory it is required that everyone should be the same, which seems horrible. As I pointed out in Chapter Two a world in which everyone is the same isn't the world we would want to live in. Agent Particularism on the other hand involves letting people be who they are in two ways. One doesn't make a moral criticism of someone else because they aren't other than the person they are. At the same time one accepts who they are themselves.



## **An Agent-Particularist Conception of Human Dignity**

### 1. Introduction:

What does it mean to respect a person? What grounds the requirement for us to respect ourselves and other people? What grants absolute worth to a human being? These are some of the questions to which a peculiar kind of answer can be derived from the Agent-Particularist position developed up to this point of the thesis.

In Chapter One I have offered a rejection of the universalizability thesis as it is normally understood. In Chapter Two I offered a re-understanding of the distinction between negative-positive freedom and a reshaping of the tradition on freedom. In Chapter Three I offered a kind of virtue ethics that allowed me to locate the discussion of the thesis with respect to a certain strand in modern moral philosophy and to clarify the epistemology involved in moral knowledge. What we saw in those chapters is the preliminary material necessary for the development of this fourth and last chapter of the thesis.

The ideas developed in the preceding chapters are what were necessary to explaining what is really striking about pursuing the idea of a richer-genius conception of the self properly. As we will see taking that idea seriously gives one a very distinctive conception of what human dignity is and why humanity has a call upon us. What the material presented in the preceding chapters does is provide us a way of giving substance to the idea of an ethics of human dignity.

In this chapter I will explicitly present the Agent-Particularist conception of human dignity. I will first present some of the key features of an Agent-Particularist conception of dignity. I will then

present Kant's conception of dignity in order to use it as a contrast with the Agent-Particularist conception. I will finish the chapter by presenting the essence of an Agent-Particularist conception of dignity.

## 2. Some Key Elements of an Agent-Particularist Conception of Human Dignity:

This thesis offers an account of human dignity, so we want to know why precisely should it be seen as an account of dignity? When we look at the etymology of the word dignity<sup>27</sup> we see that it comes from the Latin *dignitās*, which comes from the adjective *dignus*, which means worthy - deserving of respect. An account of human dignity is focused around what deserves respect in people. And the view that I am offering is precisely that each person is worthy of respect. In this thesis – especially in Chapter Two – I have developed a conception of respect for each person, and a justification for such demand for respect.

The view I am developing in this chapter comes close to a kind of conceptual analysis of human dignity in terms of each person being worthy of respect. And I am offering a whole theory which is shaped to allow that possibility. I am offering a conception of respect for each person in which the key idea is in letting them be. In the particular view I am developing that means letting each person be who they are. So the characterization of what it is to respect someone, is to let them be who they really are rather than impose anything on them or subjugate them to anything other than themselves.

In Chapter One I presented the view that what someone is obliged to do depends on who they are. This means that there are some things

---

<sup>27</sup> Source: Oxford English Dictionary.

which would be all right for one person to do and not all right for another person to do in the very same situations. According to Agent Particularism it is not a criticism of the person for whom it is all right to do that that they do that and not what the other person would do. Agent Particularism acknowledges therefore that there is a limit to how much someone can be criticized. In essence the structural feature of the position is that there is no further moral criticism that we can make of the person once they have developed themselves into their ideal version.

This feature of Agent Particularism puts the position in sharp contrast with the universalizability theorist's view. Whereas the universalizability theorist will criticize someone for failing to meet the universal ideal standard, which allegedly applies to everyone, that is a kind of criticism Agent Particularism doesn't make. The universalizability theorist has in mind the kind of person we might want there to be in a certain situation and thinks that everybody ought to be that person, and they then criticize someone for not behaving as they would ideally like someone to behave. But for Agent Particularism that is an illegitimate demand and an illegitimate criticism, because it is wrong to criticize this person for not doing what the universalizability theorist thinks they ought to do if it is not demanded of them under the ideal development of their particular genius.

The universalizability theorist imagines that there is an ideal type of person, a sort of perfect human being that everybody ought to be, and insists that people should be like that. But the perfect human being doesn't exist. The universalizability theorist fails to acknowledge that there is no such person as the ideal person (I will come back to this point later). Because of that they want to insist that the person that behaves as we would ideally want someone to behave in a certain situation is a morally better person than whoever behaves differently from what we would ideally want.

Agent Particularism rejects this view. The position consists in accepting that this person is not worse than the person that behaves as we would ideally want even though they don't do that, they can't do it while being the same person, and therefore they can't be required to do what in a certain sense ideally speaking we would want someone in that situation to do. They are not in any way a worse person.

It is integral to the Agent-Particularist position that it respects people's individuality. Agent Particularism itself means that there can't inevitably be a moral criticism of someone because they don't do in a given situation what we would ideally want someone to do. They just aren't that person, and that doesn't make them a worse person. The thing is that for Agent Particularism there isn't a way in which someone could be compared to another person and found less good because they are a different person. And that simply is respecting the individuality of people. It does not require people to be other than they are. It requires them to reform themselves to become the ideal version of themselves but it does not require them to be different people. In order to help clarify the point let's see how it could be illustrated with an example.

Let's imagine that there is a group of fancily dressed people, with an arrogant attitude, humiliating someone because they are poor and black. Let's further imagine that there is someone else close by, occupying the position of a bystander. What we would want there in that position ideally is someone who could stand up to the bullies and show how that kind of behaviour is unacceptable. So ideally we would want a kind of person who is courageous, confident and bold to be there. If the person who is actually there is a 7 year old child it is no criticism that they can't do what this bold adult could do. Similarly if what we have there is someone who is meek and diffident it is no criticism of them that they don't do what the person with the bravado and confidence could do. This person simply couldn't do that. The

person we would want to be there is the person with the kind of bravado, as that would protect the vulnerable person. But Agent Particularism is committed to it not being a criticism of the diffident person that they aren't that person - they aren't a worse person for not being that person.

This example shows the contrast between Agent Particularism and the universalizability theorist. The central thing about Agent Particularism is that we are not going to make moral criticisms where the universalizability theorist's view would make moral criticisms, and we are not going to compare morally people who the universalizability theorist would compare morally.

In the context of this example, the defining feature of Agent Particularism is that the mild and diffident person in this case is no worse as a person than the confident person. In essence the Agent-Particularist definition of respect is a matter of letting people be who they are. Accepting that the diffident person in the example is not a morally worse person is to accept the individual for who they are, not for how well they live after some general standard. Agent Particularism has a sense of what is worthy about human beings in which it is about what is worthy about individual human beings. It is about the dignity of each human being. We let each human being be who they are, and it is not regarded as a fault in them that they aren't another kind of person.

We can see in an example like this that Agent Particularism inevitably respects human individuality. Now if Agent Particularism is true, since it is distinctive of Agent Particularism that it respects human individuality, then it is the view that everyone ought to hold. So if Agent Particularism is true everyone ought to respect people's individuality. The claim that we ought to respect people's particular genius just follows from the truth of Agent Particularism.

The Agent-Particularist claim is that we ought to respect people's individuality and ought to allow them to become the ideal version of themselves. At this point it might be asked, but what is the source of this ought? There are two questions that I want to separate when we are thinking about the source of the ought. On the one hand one might have in mind the view that the 'ought' here expresses a reason of one kind which needs justification in terms of reasons of another kind. For example, the idea might be that the 'ought' might express a moral demand which then needs justification on some non-moral grounds. There is an idea behind this that says that this kind of reason (moral reasons) is not basic. This kind of view is held, among others, by Christine Korsgaard who grounds morality in human nature:

"... morality is grounded in human nature, and ... moral properties are projections of human dispositions" (Korsgaard 2005, p. 91).

So when we say that you ought to become the best version of yourself and other people ought to let you be yourself, people like Korsgaard would insist on a demand that this be justified by something more basic – in her case grounding it in human nature. But it is not obvious that morality has that sort of source at all. Morality can reasonably be regarded as basic. Therefore, this is a demand that my Agent Particularism need not accept.

The other reading of the phrase 'source of the ought' asks for the justification of this particular Agent-Particularist claim. In my view the 'ought' involved here is a moral 'ought' so we need to look for the kind of justification which is required for a moral justification. I just said that doesn't need to be a justification in terms of human nature or prudence or anything like that. According to Agent Particularism becoming properly developed is what is required for you to become a moral person. This is the basis for morality on the Agent-Particularist

view. The thought is that morality is a demand on everyone and (as I explained in chapters Two and Three) you only become open to morality through developing in the direction of the ideal version of yourself. This means that you ought to become the best version of yourself. And then the thought is just that it must be immoral for anyone to prevent someone from becoming a properly moral person. So it's a moral ought and the basis of this demand is that it must be immoral to prevent someone from becoming a moral person.

At this point I want to clarify something. Agent Particularism allows that some cases will have that appearance that I am suggesting in the example above - that it is required for the person of confidence to act in a certain way (such as intervening) but not required for the mild and diffident person. But that doesn't mean that every case where you have this kind of difference needs to be treated similarly. The reason is that respecting people's individuality doesn't mean using whatever trait they happen to have in a given moment as an excuse to get them off the hook. Agent Particularism includes the demand for people to develop into their ideal self. At times some act may be challenging for the agent and yet they are required to do it as that is something that their ideally developed version could do. Let's look at how this would work in another example.

We can imagine for instance a situation in which someone needs to overcome, say, fear in order to do what is required of them. Let's imagine that an angry son and his elderly mother are arguing on the street. The discussion heats up and the son pushes his mother who loses balance and falls, hurting herself perhaps seriously. The son, still angry, seems to be glad about what happened and doesn't do anything to help her. Suppose there is a bystander - with a tendency to avoid confrontations - that sees the whole scene and knows they ought to help the elderly woman, but they are afraid that the angry son will do something to them if they try to intervene in the situation. If helping

the elderly woman in this case is something that the ideally developed version of them could do, though they are afraid they still ought to find the courage to help the woman even if that means somehow overcoming the fear of perhaps having to confront the angry son.

The point I want to make with this example is that I am not suggesting every time there is a situation which is challenging, given one's current traits, it is OK for them to not act on it (I first introduced this point at the end of Chapter One). I am also not suggesting that a person should never be expected to act as we would ideally want them to act. If under the ideally developed version of themselves they can do something, even if it is challenging, they still ought to do it (I will make this claim more precise later).

The point that Agent Particularism makes is simply that it is not always the same obligations for everybody. It still allows that there may be some things that no one should do and some things that everybody should do. If one happens to come across an abandoned baby for instance, it seems plausible that something that everybody should do is to try to find some care for that baby. That doesn't mean suddenly becoming its parent, but it does mean not letting a helpless baby simply die out of lack of care.

It is one thing to say that it isn't right to take one person to be better than another because they do what we would ideally want someone to do in a situation. But it doesn't follow from that that we can't like one more than the other. Agent Particularism makes a distinction between moral disapproval and dislike. It holds that anything we dislike about somebody else even when they have improved themselves to their ideal limit is merely dislike and can't be properly understood as moral disapproval – it is not well-founded moral disapproval. It is merely a matter of taste.



There is a temptation to think that one's taste is well grounded and that we like people because they are worth liking and good rather than it merely being the case that we like them. Agent Particularism accepts that we can disapprove of somebody morally and dislike them because they are morally mistaken. But it is a fundamental feature of the view that those will be cases in which they are failing to be the ideal version of themselves. They are failing to become a person that can do something that falls within the scope of what the ideal version of themselves would do. Any disagreement beyond that is merely dislike. And the view accepts that dislike has to be regarded as strictly speaking a matter of taste and quite distinct from moral approval or disapproval.

This issue brings us to a debate within moral philosophy. There has been a tradition of eliding the difference between taste and morality, the sentimentalist tradition. Agent Particularism rejects that view. Since Hume is the great exemplar of that tradition I will look at how he sees the distinction (or indeed lack of it). Hume counts morality as falling within the province of taste but Agent Particularism insists they are separate. For Hume:

"... the distinct boundaries and offices of REASON and of TASTE are easily ascertained. The former conveys the knowledge of truth and falsehood: the latter gives the sentiment of beauty and deformity, vice and virtue. ... From circumstances and relations, known or supposed, the former leads us to the discovery of the concealed and unknown: after all circumstances and relations are laid before us, the latter makes us feel from the whole a new sentiment of blame or approbation" (Hume 1777, pp. 71-72).

We can see from the quote above that Hume groups together beauty and deformity, vice and virtue - it all counts as taste for him. Hume therefore doesn't accept a distinction between taste and morality. For him reason concerns itself with truths, and morality, which falls under taste doesn't involve rationality. Agent Particularism on the other hand sees rationality involved in morality and holds a robustly realist

approach to it. So it is natural for Agent Particularism to make a sharp distinction between taste and morality because there are no truths in what relates to taste. In fact it is essential to the view that there is a fundamental distinction between taste and moral disapproval – one of them has a rational basis, the other one doesn't.

Taste, that is a liking or disliking which has no rational basis, is the precise counterpart to accepting that there are just facts about people's character which are how they are and not some kind of version of the perfect human being. The distinction between moral approval or disapproval, on the one hand, and taste, on the other, is exactly parallel to the distinction between what each person ought to do, which involves limits in what is alright for them to do, and how we would like them to behave as an ideal human being.

The Agent Particularist claims that to respect a person means letting them be who they are. In the very same situation different people can face different obligations. This implies that even if someone can't do in a particular situation what we would ideally want someone to be doing in that situation that is not a moral criticism of them. At the same time Agent Particularism doesn't allow it to be the case that anything goes, it doesn't allow that the person be able to get out of their obligations. As I explained in Chapter Two it is precisely to rule out that possibility that Agent Particularism needs the idea of the demand for people to develop into the ideally developed version of themselves.

This raises a question which I haven't yet addressed. How does one improve oneself? In Chapter Two I have shown how negative freedom is necessary to and even a part of one's positive development. I expanded on that positive development by suggesting certain virtues of character which are required to make that development accord with each person's particular genius. In Chapter Three I complemented that view by explaining what other kind of virtue of character someone who

is developed into their ideal version must have, namely the virtues of character that allow one to develop a sensitivity to the requirements of circumstances on them. But I haven't yet addressed the question of how to recognize what one should do in order to develop themselves in the necessary way.

For Agent Particularism what is involved in respecting the individual involves at the same time letting other people be who they are and also letting yourself be who you are. It is the demand to respect who you are that stops the widespread drive towards eliminating any trace of individuality, which is the result of striving towards becoming some kind of ideal human being. Letting yourself be who you are requires a balance between striving towards developing yourself and respecting who you are.

In particular situations we can tell from the situation the kind of person that needs to be there. And we can see that in all sorts of situations there need to be different kinds of people. Because we can see the kind of person that we would want in each of the different situations what we do to improve ourselves is to try to become as much as we can the kind of person which anybody would want in any situation. This is what we do in order to work out how to improve ourselves.

But each person can only improve themselves in this way as much as they can. It shows us how to improve ourselves, but it is not the case that such improvement is required without limit. One can't be the person they are and be capable of doing all the different things that we would want people to do in each different situation. Each person will inevitably be a person who is better at handling some situations than others. Different people will be better at different situations. The demand is for people to try as much as they can but they also have to recognize who they are - and it is not a moral criticism of them that they aren't a different person.

The person tries to develop in the direction of someone who would be able to do all these things but at a certain point they have to recognize that there are limits to this striving. There are limits because there is a demand to respect oneself. This demand imposes limits to any kind of striving towards becoming an ideal person. There is no such thing as the person who is going to be equally good in every situation; there is no such person as the ideal person. The ideal person would be a person with no character. All there is is the ideal version of each person. Each person is who they are and as such they are better at handling some situations than others. This is what it is for people to be different.

This explains why the universalizability view seems attractive even though it is wrong. We need to start by trying to develop in the direction of someone who would be able to do everything we would want someone to do, but it is a mistake not to acknowledge that there has to be a stopping point in that pursuit. Universalizability seems to flow naturally from the way we have to think to improve ourselves. But it is not noticed that there is no such thing as the ideal person. It is just a failure to recognize that at a certain point one has to accept that this is what one is – good at certain things and not so good at others.

My view is that there are certain features of a person that are essential to them, and ought to be respected. There are a few clarifications related to that idea that I would like to make. By going through them we will be able to further clarify the sense in which there are objective features of a person that are essential to them. Precisely because people are particular and unique the same character features might be an ineliminable part of one person but a changeable part of another. I will also allow that it may be difficult to identify which are the features which are truly you and which are the ones you just happen to have. As I explained in Chapter Two your positive development into the ideal version of yourself is a process, and that also determines the nature of the virtues of character which have to do with being, or becoming

yourself, which I presented in the discussion of positive freedom. It is indeed a demanding process.

But my position does not require that it is easy to identify the features which can be changed. What my position does require is the idea that there are some features of your character which are themselves essential to you or such that developing them is essential to you and some features of your character which are incidental and you need to change, or you can change. That there are such features, and that they differ from person to person is what is needed to make a contrast between the richer-genius conception of the self and the blank conception of the self. It doesn't need to be the case that we can explain in any easy formula what is going to count as essential to a particular person. This is because the view that it has to be possible to specify which features are essential to someone and which are not goes one step towards rejecting Agent Particularism. If it is some specific feature that we have in mind we have a formula which we apply to everyone. In this case we wouldn't be Agent Particularists anymore.

Another point to clarify is the relation between a person's development and what is essential to them. I want to propose two models here, with the important thing being that in both cases we can make a clear distinction between what is true to the person and what is alien to them. One model holds that the ideal version of yourself is fixed from the moment you are born, and the challenge is to develop to become that person. The other model holds that that is implausible, because things happen to people in their lives which shape their characters. In this case at different stages of someone's trajectory of development there may be different fixed ideal versions. I want to organize the discussion of those models of development around the different senses in which the term 'fixed' could be applied here, and the contrast in terms of what is true to you and what is alien to you.

In one sense 'fixed' means something like objective. I take it to be an objective fact that some features are essential to a person and others not. For them to be objective is for them to be determined (fixed in that sense) independently of what anybody thinks. The other sense of 'fixed' is 'needing no development', i.e. unalterable. One's particular genius isn't fixed in this sense, as it is not yet developed. Let's see how this works in terms of the two development models that I have mentioned above.

According to the first model the ideal version of yourself is fixed (in the sense of 'determined') from the moment you are born, and the challenge is to develop to become that person. The ideal self, of course, isn't simply to be uncovered because you have to develop. It is there always from the moment you are born, already determined, what the ideal version of yourself would be. But the ideal version isn't yet present in you, you have to grow up and develop faculties in order to become that person. So the ideal version of yourself would be fixed in one sense, determined in advance, but not yet actually present – so not fixed in that sense. In this first model the contrast between what is truly you and what is alien is very simple. Some things are determined from the start as true to you and everything else is alien.

According to the second model things happen to people in their lives which shape their characters. Once you imagine decisions and choices being made, and things happening in people's lives you suppose that what is true to you doesn't fix (determine) a unique trajectory of development from the moment you are born. It doesn't determine every choice of significance in your life. So there are plenty of choices which are open to you which are true to yourself, but which of those true to yourself choices you make makes a difference to a proper development of yourself.

The thought that I am presenting at this point is that people can be changed by things that happen to them. They live in a particular environment, they get to know particular people, particular things happen to them. And the important thing is that this doesn't involve them at any point necessarily taking on features which are alien to them, but just require them confronting and attending as best as they can to the circumstances which they are faced with (I will explain this in more detail soon).

My central thesis is that what you are required to do depends on who you are. That means an important question is what determines who you are, what makes you who you are. What I am suggesting on the second model here is that there are some things that can intervene in your life that can make a difference to who you are and therefore make a difference to the ideal version of yourself. Again, different things will have a different effect on different people and it is not always going to be clear which events in life will actually change one's character.

We can think of obvious examples of transformative events either good or bad. So we can imagine that if you have a devastating injury, that might make a difference to the kind of person you could be. It might no longer be possible for you to be the person you might have been. From this we could see how it could be the case that at each point in your life there is in a sense fixed (determined independently of what you think) an ideal version of yourself, but of course things of various sorts that happen might push that around, i.e. change what it is. So possibilities of development which were open before are no longer open. And perhaps others become open that weren't. Some features of your character are just part of who you are. And some of those are going to be determined by physical characteristics, such as your size. Those are pretty much determined by birth. But different kinds of upbringing close down some possibilities and open up others. Now, the same process can go on in principle throughout your life. So you can imagine

that at each point there is relative to that point a fixed (determined independently of what anyone thinks) version of you. And your task at each point in your life would be to try to become the ideal version at each point.

Something similar can be seen in a different kind of case. Imagine someone who is 9 years old when their mother suddenly dies. There is no doubt that this kind of event makes a difference in an everyday sense to who someone is. The question is whether it makes a difference in the strict sense to who they are. It might be that it does. It means that the person is faced with kinds of choices of ways of behaving at a stage where you wouldn't perhaps face those choices and you take particular routes down them that are perfectly alright in terms of your development, but you could have done something else. Having gone there you've become a different person in some sense. This allows for life situations to influence who the person is without it being the case that you are doing something because someone else wants you to. It's just that you have to work out your own way of negotiating these things, and you are facing situations that you wouldn't have faced if that haven't happened.

The important thing is that there is a certain sense in which facing situations of a certain sort doesn't just require you to act in a particular way on one occasion, it requires you to take on a certain attitude. There is a way in which you have to learn to behave in a particular way to cope. And that makes a difference to who you are. So the acting in response to big things like that is a matter of partly training how you behave, and that is bound to make a difference to your character.

These are dramatic examples, but the same kind of thing can happen in all sorts of circumstances. If you find yourself living your life in a small community you're likely to become a different person from one who finds himself living their life in a big city. They could end up being



different in character from one another even if we imagine that they began in the same way and they don't just absorb what people say. One's way of living has to be different in different environments.

The important thing is that what this thesis requires is just a contrast between what is essential to you and what is not. It insists that what is essential varies from person to person, and it allows that it is not easily discernible. What is essential to a person is fixed just in the sense that it is determined independently of what anybody (whether that person or another) thinks. But depending on the model you choose, it may be subject to change according to the stage you are at in the process of development.

What is needed to contrast my view with the blank conception of the self is just the idea that there is something which is essential to each person, and what is essential varies from person to person. As I have already said, it doesn't need to be easy to identify what is essential to each person, and there won't be an easy formula which can be used to explain what is essential to each person. This is why personal education is a long and difficult process, on my view.

This discussion brings us back to an issue I discussed in Chapter Two. I asked, what is essential to the very idea of Agent Particularism? I said then that the basic Agent-Particularist position involves the view that what is required of each person depends upon who that person is, which depends upon features of their actual character. That in itself requires the idea of an ideal improvement of oneself. From there I introduced three possible positions for the basic Agent-Particularist view and indicated which one is mine, explaining why. I will now come back to that position.

The view is that a person can't be required to do things which they in some sense couldn't do while still being the same person. This means that who they are provides a constraint on what they ought to do. What

is important for the discussion we are seeing here is that this view doesn't imply that everything which an ideal version of themselves would do is required of them.

The notion of supererogation applies here. Supererogation refers to:

"... the class of actions that go "beyond the call of duty." ... supererogatory acts are morally good although not (strictly) required" (Heyd 2015).

There are, in my Agent-Particularist view, two senses of a supererogatory act. The first sense involves a more traditional understanding of supererogation and is less important to the point I want to make. Once a person has reached the point of their ideal development there is a contrast to be made between what they are required to do and acts that go beyond the call of duty. Some things that the ideal person would do might not be required of them but just optimal.

The second sense, the more relevant one, involves a contrast between what is required of one now and what would be required of the ideal version of oneself now. In what sense is the person now required to do what the ideal version of themselves would be required to do? There is some sense in which one ought to become the person who would be required to do that, and not everybody is required to become that person, but even so one might not now be required to do it. The reason is that it might be the case that one is not yet in a state of being capable of doing what the ideal version of themselves would do.

The point I have been making is about how a person knows what to do in order to develop themselves. As I said above what we do to improve ourselves is to try to become as much as we can the kind of person that anybody would want in any situation. But at a certain point the person has to recognize that there are limits to this striving because there is a demand to respect oneself.

We can now complement this view. A full story about what a person is required to do will be determined not just by who they are, but also by how far along towards becoming the ideal version of themselves they have got. They might face different obligations in different stages. They have the standard obligation to become the best version of themselves but then what other obligations they face might depend upon how far along that line they are. There is therefore this other kind of limit to what a person can be required to do.

### 3. Kant's Conception of Human Dignity:

I have presented above some key elements of an Agent-Particularist conception of human dignity. I will now further develop that conception by contrasting it with Kant's conception of dignity. I will use Kant for two reasons. The first is that Kant's conception is very carefully developed and structured, providing a rich and interesting point of contrast. The second is that anyone who thinks of dignity in philosophy will have Kant in mind. This means that orienting my position with respect to Kant's is the simplest way of making my position clear.

My aim in looking at Kant's conception of dignity is to provide an interesting alternative to my own view that I can explain my own view in contrast with. So it is not interesting to me to cover an interpretation of Kant which isn't going to produce a view that engages with the one proposed by this thesis. This has implications for the kind of literature that will be more helpful to our discussion. As will be clear shortly it turns out that the literature that discusses Kant's conception of dignity directly isn't going to be of much help. For this reason I will only illustrate what I mean in one case and explain why it wouldn't help us much, and then quickly mention a few of the points that the literature

tends to discuss. Then I will move on to the more helpful literature that in general focuses on Kant's idea of humanity as an end in itself.

Different authors will focus on distinct aspects of Kant's conception of dignity and in general they intend to propose an interpretation of the aspect they are looking at which they claim to be an alternative to how Kant is usually understood. In some cases other authors have picked up on the discussion and a debate has formed on the matter. Many of these authors will make an historical investigation of the concept of dignity and contextualize Kant in terms of that history.

I will illustrate this with the case of Oliver Sensen's interpretation of Kant. Sensen's main concern is with the view that for Kant human dignity is a non-relational value which is intrinsic to human beings. Sensen rejects this view and proposes that for Kant dignity is rather a relational property that entails in the quality of being elevated above something else. According to this view what has dignity is seen as superior to what hasn't got it in some way. Sensen distinguishes between distinct traditions (or paradigms as he puts it) of the concept of dignity and places Kant as a member of the tradition for which dignity is a relational property that is not intrinsic to human beings. Sensen claims that in Kant's view dignity is not what grounds a demand for respecting people. Sensen rejects the view according to which:

"... human dignity is a nonrelational value property human beings possess that generates normative requirements to respect them. Because of the prominence of this pattern of thought within and outside the Kantian literature, I call it the 'contemporary paradigm of dignity'" (Sensen 2009, p. 312).

This contemporary paradigm of dignity is according to Sensen the standard understanding of Kant:

"The standard view in the Kantian literature ... is that dignity is a certain type of value of human beings. This value is characterized with attributes such as 'absolute', 'inner', or 'unconditional'. ... The value is also described as

'incomparable', implying that human value cannot be traded against other value, for instance the value of things" (Sensen 2009, p. 311).

According to Sensen Kant is actually better seen as a part of what he calls the traditional paradigm of dignity according to which:

"... 'dignity' ... refers rather to elevation, a relational property. ... For instance, if one says that human beings are elevated over the rest of nature, it merely amounts to saying that human beings are distinguished from the rest of nature by having capacities (e. g. reason, freedom) that put human beings at a distance from immediate natural determination. This instance of elevation says something about humanity's place in nature, but it does not yet imply anything about how human beings should treat each other. For this one needs a further normative premise (Sensen 2009, p. 313)."

Sensen also rejects the view that for Kant humanity as a characteristic of human beings is the ultimate end of human action – a view which we will see in detail later. According to him:

"... it is a common view to think that every act must serve a good or value, and it is a common interpretation of Kant that every action needs an end; that moral actions need a special moral end; and that this end is humanity as an end in itself. ... However, I do not believe that this reading is correct (Sensen 2015, p. 121)."

Sensen's interpretation of Kant's conception of dignity is of little interest for our purposes. On the Agent-Particularist conception acknowledging the dignity of a human being is letting them be the person they are, and not regarding it as a fault in them that they are not another person even if it means that they are not the person one would most like to have in the situation. The idea is that there is something about a human being which is worth respecting and in the end it is their being the person they are.

This means that the central notion of the Agent-Particularist conception of dignity is the notion of being worthy of respect. Human dignity is that about a human being which is worthy of respect (in fact the

general notion of dignity is about what is worthy of respect). It follows from that that I am not going to be interested in interpretations of Kant which don't have something like this notion involved in it. In this sense a large portion of the literature which is specifically on Kant's use of the concept of dignity is concerned with issues that aren't the concern of this thesis.

This is what happens in the case of Sensen. Sensen basically wants to say that all that human dignity in Kant means is being in some sense higher than animals, rejecting the view that human dignity is what grounds the demand to respect a person. As interesting as it may be Sensen's interpretation doesn't engage with the conception of dignity proposed by Agent Particularism and so isn't helpful in terms of fleshing out the view. Whatever the truth of Sensen's claims about Kant none of what he discusses is going to be relevant to our concerns. The same applies to the debate which followed Sensen's proposal<sup>28</sup>.

Von der Pfordten is another author who attempts at a direct interpretation of Kant's conception of human dignity. He concludes in his interpretation that dignity is not really as central to Kant's moral account as it is taken to be. Finally I want to mention that some commentators call attention to Kant's different usages of the term dignity throughout his works. Michael Meyer for instance focuses on a distinction between dignity as a kind of distinguished or higher rank of a person in the socio-political hierarchies of society and dignity as a distinct quality of human beings. Meyer held that Kant used both senses of the term dignity. But none of these interpreters engage with the idea that dignity refers to that which is worthy of respect in human beings, and so they aren't for the most part helpful to our purposes.

As I said earlier, if one is working with a conception of dignity as that which is worthy of respect in a person a more fruitful source can be

---

<sup>28</sup> For examples of this see Jochen Bojanowski and Stefano Bacin.

found in the literature that addresses Kant's idea of treating humanity as an end in itself. So I will now turn to that literature (besides, of course, Kant himself). My aim in looking at the conception of human dignity which we can find in Kant's idea of humanity as an end in itself is, as I said earlier, in using it as a helpful alternative to my view. This is not a thesis about Kant, so it will suffice to produce a version of Kant which is reasonably well rounded, reasonably plausible, and which provides the helpful alternative that I need. Therefore the interpretation of Kant which I will offer in what follows is simply one packed together from a range of sources and it provides a plausible and reasonably orthodox interpretation.

A way in which Kant formulates the idea that a person ought to be respected is the following:

"... all rational beings stand under the law that each of them is to treat itself and all others never merely as a means, but always at the same time as an end in itself" (Kant 2012, p. 45).

For Kant respecting a person is to act in such a way as to treat them always as an end, never merely as a means. In order to explain what this means I will start by coming back to the discussion on Kant which I introduced in Chapter Two. In Chapter Two I explained Kant's distinction between actions done out of inclination and actions that are independent from the influence of inclination (rational action). Actions which are done independently of inclination - which for Kant are the free and moral actions - have an absolute worth, i.e. their worth is not conditioned by the attainment of some end that they aim at. Whereas actions done out of inclination always aim at a certain end, and the worth of those actions is conditioned by the attainment of that end.

Kant also makes another distinction, not of the actions but rather about the objects or ends of those actions:

"All objects of inclination have a conditional worth only; for if the inclinations, and the needs founded on them, did not exist, their object would be without worth. ... Beings whose existence rests not indeed on our will but on nature, if they are non-rational beings, still have only a relative worth, as means, and are therefore called things, whereas rational beings are called persons, because their nature already marks them out as ends in themselves, i.e. as something that may not be used merely as a means (and is an object of respect)... " (Kant 2012, p. 40).

According to Kant something is treated as merely a means if one acts in relation to it from inclination. It means that one uses it in order to achieve, for instance, something that one desires. It is thus a means to achieving it. Things and even non-rational creatures have only a conditional worth, i.e. they have a worth for us when they are the objects of our actions. The worth of a rational creature on the other hand doesn't come from our achieving some particular end that we might have, but they are rather worthy independently of all the ends that come from inclination. They are an end in itself and in that sense have absolute worth.

In order to further clarify what Kant means by his idea that humanity is an end in itself let's look more carefully at it. Richard Dean puts Kant's view like this:

"An end is a ground of the will's self-determination, or, in more common terms, a reason that a person adopts for acting" (Dean 2006, p. 111).

Kant's view is that a human being always acts for a reason, i.e. there is always something which grounds their action. According to Allen Wood, Kant:

"... distinguishes two types of grounds on which a will may act: subjective grounds, based on empirical desire for an object, and an objective ground, which is also an end, but one given by reason alone and valid for all rational beings" (Wood 1998, p. 167).



The idea here is that a person acts on subjective grounds, i.e. their end is subjective, when that end is based on inclination. It depends on the particular inclination of a human being. According to Dean's view, Kant contrasts the notion of subjective end with the notion of objective end:

"An objective end does not depend on a person's inclinations, but rather is 'given by reason alone' (G 427). Since this end is required by reason alone, it must hold equally for all rational beings', meaning it provides every rational being with a reason to act in certain ways, regardless of her inclinations" (Dean 2006, p. 111).

So Kant's view is not that an action that has absolute worth, which is independent of inclination, has no end while action based on inclination has an end. If we follow Korsgaard's interpretation the view is rather that:

"... every action "contains" an end; there is no action done without some end in view. The difference between morally worthy action and morally indifferent action is that in the first case the end is adopted because it is dictated by reason and in the second case the end is adopted in response to an inclination for it" (Korsgaard 1996, p. 107).

So Kant's basic distinction is between ends that come from reason and ends that come from inclination. Subjective ends are based on the particular inclinations of individual human beings and in this sense are not shared by all persons. Objective ends, which are dictated by reason and reason alone, are according to Kant shared by all human beings. In Wood's view Kant also refers to objective ends as ends in themselves:

"... there is the conception of an end in itself or objective end, whose worth, as the ground of categorical imperatives, is unconditional, independent of desire and valid for all rational beings, a "motive" constraining the will through its own faculty of reason. The opposite of an end in itself is therefore a relative end, whose worth is relative to, conditioned by and dependent on the subjective constitution of each particular rational being, and hence varying contingently from one such being to another. This sort of end would be only an "incentive" and could ground only hypothetical imperatives" (Wood 1998, p. 168).

So Kant's distinction between subjective (or conditional) ends and objective ends (end in itself with unconditional worth) is at the same time a distinction between action which is morally indifferent and morally good action. In order to give a bit more context into Kant's idea here I want to go back at a point which I explained in Chapter Two.

Kant thinks that the rational being is free. This means that as rational they are bound to act according to some freely adopted maxim, which can be of two kinds. It can be a maxim grounded in some inclination - a hypothetical imperative with no intrinsic worth. Or it can be a maxim that has an intrinsic value independent of anyone recognizing that value, which according to Kant is the moral maxim. This maxim, Kant thinks, is grounded on an objective moral law – which is his categorical imperative.

Kant identifies what he calls hypothetical imperatives with actions to which he attributes a conditional worth, while he identifies what he calls categorical imperatives actions to which he attributes an absolute worth. Since, according to Kant, the actions of the free will of the rational being are supposed to be independent from inclinations they are supposed to be motivated by categorical imperatives as opposed to hypothetical ones. This means that moral action is the action that is motivated by some end which is an end in itself.

As I hinted at in the beginning of this discussion on Kant he identifies humanity with his notion of end in itself:

"This principle of humanity and of every rational nature as such, as an end in itself ... in it humanity is represented ... as an objective end that, whatever ends we may have, as a law is to constitute the supreme limiting condition of all subjective ends and hence must arise from pure reason" (Kant 2012, p. 43).

In fact according to Wood's interpretation:

"Kant's position is that there is only one thing, namely humanity or rational nature, which satisfies the concept of end in itself" (Wood 1998, p. 171).

The morally good action is the action whose motivation must be an objective end or end in itself. And the thing that qualifies as an end in itself is humanity. So it is from this idea of humanity as an end in itself that Kant derives the idea that it should always be treated as an end and never merely as a means. In Dean's view of Kant:

"Humanity is an end in itself, or an objective end, because each agent is rationally required not to sacrifice her own or others' humanity ... for the sake of her inclinations" (Dean 2006, p. 129).

Kant's view is that a human being always acts on a maxim. Morally indifferent action is done out of one's inclinations. But in the idea of treating humanity as an end in itself Kant has already ruled out action done out of inclination. So the only candidate left for the kind of maxim that morally good action follows is categorical imperatives. And one of Kant's formulations of the categorical imperative is precisely:

"So act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means" (Kant 2012, p. 41).

So the requirement is that we respect humanity in our own person and in that of others. This means that what grounds the requirement to respect people in Kant is what he refers to as humanity. But what exactly does Kant mean by humanity? According to Hill, there is a common understanding of the phrase humanity in a person:

"On the usual reading this is treated as a quaint way of saying "a human person". That is, treating humanity in persons as an end is just to treat human beings as ends. "Humanity," on this view, refers to the class of human beings,

and what is meant is simply that each member of the class is to be treated as an end" (Hill 1992, p. 39).

But Hill argues convincingly that this is not what Kant means<sup>29</sup>. Kant's view is not that we must treat humanity as the group of individual human beings as ends in themselves. What grounds the demand for respect is not a person, but rather humanity in a person as Hill's view of Kant suggests:

"Kant thought of humanity as a characteristic, or set of characteristics, of persons. ... humanity is contrasted with our animality ..." (Hill 1992, p. 39).

So by humanity Kant means something like that which is uniquely human. But as I hinted at in some of the quotes above Kant doesn't think that it is only human beings that should be treated as ends in themselves. Kant refers to the: "principle of humanity and of every rational nature as such, as an end in itself" (Kant 2012, p. 43). Since for Kant what distinguishes human beings is something that they actually share with rational beings in general it is not difficult to imagine that it is something like our capacity to reason. This conception of what Kant means by humanity is shared also by Robert and Adam:

"Our "humanity" is that collection of features that make us distinctively human, and these include capacities to engage in self-directed rational behavior and

---

<sup>29</sup> In Hill's reading of Kant: "There is no temptation to think of "man" as referring to something in a person, or a characteristic of a person, though "humanity" can be so understood, for example, when we contrast a person's animality with his humanity .... A review of Kant's repeated use of "humanity in a person" in *The Metaphysics of Morals* and elsewhere strongly suggests that, contrary to the usual reading, Kant thought of humanity as a characteristic, or set of characteristics, of persons. Kant says, for example, that we can even contemplate a rogue with pleasure when we distinguish between his humanity and the man himself (*Lectures on Ethics* 196-97; *The Doctrine of Virtue* 107 [441]; *Metaphysical Principles of Virtue* 104 [441]). Again, humanity is contrasted with our animality; and it is said to be something entrusted to us for preservation (*The Doctrine of Virtue* 51 [392]; 85 [423]; *Metaphysical Principles of Virtue* 50, 84). Its distinguishing feature is said to be "the power to set ends," and we are supposed to respect it even in those who make themselves unworthy of it (*The Doctrine of Virtue* 51 [392]; 133 [463]; *Metaphysical Principles of Virtue* 50, 128)" (Hill 1992, p. 39).

to adopt and pursue our own ends, and any other rational capacities necessarily connected with these" (Robert and Adam 2017).

Hill helps to make the understanding of humanity in Kant even more precise:

"... it is most reasonable to construe "humanity" as including only those powers necessarily associated with rationality and "the power to set ends"" (Hill 1992, p. 40).

At this point I would like to call attention to something that I won't go into now, but will come back to when I get to my objections to Kant's conception of human dignity. We have seen that Kant refers to what is an end in itself and as such worthy of respect as humanity in human beings. And humanity, regardless of variations among different interpreters, is always understood as a person's rational capacity. So in the end when it comes to Kant's account of what is worthy of respect in a person it is rationality. There are two issues about this. One is that according to Kant what is worthy of respect in a human being is just a limited capacity in them, which is the same for everybody. There is no sense of respecting an individual. The second issue is that this reinforces the point which I made in Chapter Two that Kant, with his blank conception of the self, treats a person as a kind of personification of reason. Dean is even clearer in his interpretation:

"... each agent is taking the same aspect of herself, namely her rational nature, as valuable or as an end in itself, so what she is really valuing or taking as an end in itself is rational nature as such ..." (Dean 2006, p. 127).

When we talk about Kant's idea of humanity as an end in itself there is a clarification to be made. The term 'end' suggests the idea of something like a goal to be achieved, but that is not the sense in which Kant sees humanity as an end. As we have seen in a quote above Kant

says that as an end in itself humanity: "... constitute the supreme limiting condition of all subjective ends ..." (Kant 2012, p. 43). Humanity limits what can be done towards a person. According to Robert and Adam's interpretation, because Kant sees it as a limiting condition, he believes humanity is an end in a negative sense:

"Humanity is ... an end in this negative sense: It is something that *limits* what I may do in pursuit of my other ends. ... Humanity is an *objective* end, because it is an end that every rational being must have. Hence, my own humanity as well as the humanity of others limit what I am *morally permitted* to do when I pursue my other, non-mandatory, ends" (Robert and Adam 2017).

This negative sense is the essence of Kant's understanding of treating humanity as an end in itself. But according to some interpreters Kant tries to also show a positive sense of treating humanity as an end, both in terms of one's own humanity and of humanity in others. On Korsgaard's view of Kant, treating one's own humanity as an end in itself in a positive sense involves:

"... developing and realizing the capacities which enable you to exercise your power of rational choice .... This, indeed, is as close as Kant comes to assigning a positive function to humanity as an end. What makes this possible is the fact that rational nature is a sort of capacity. ... we can realize our rational capacities more or less fully ..." (Korsgaard 1996, p. 127).

So according to Korsgaard the closest that Kant gets to assigning a positive function to humanity as an end in itself is in the view that we have a duty to promote our own humanity in terms of developing or realizing our rational capacity. There is also the idea of a positive sense in treating others as ends in themselves. According to Robert and Adam Kant's view is that:

"... insofar as humanity is a positive end in others, I must attempt to further their ends as well. In so doing, I further the humanity in others, by helping further the projects and ends that they have willingly adopted for themselves" (Robert and Adam 2017).

This anticipates a point to which I will come back later, when I get to my objections to Kant. Kant's view is clear in terms of treating humanity as an end in itself in this negative sense of limiting one's actions. But it is much less clear when he attempts at a positive sense to it.

We have seen above that Kant identifies our humanity with the power to set ends as Hill suggested. This power to set ends is according to Kant an aspect of humanity, something unique to rational creatures. This rules out any ends that rational creatures have in an instinctual way. According to Hill's view of Kant, the power to set ends is therefore a kind of capacity that allows rational creatures to determine themselves:

"... to acknowledge that something, such as humanity, is an end is to grant that one has a "ground" for choosing, or "determining oneself," to do or refrain from doing various things" (Hill 1992, p. 43).

So humanity's being an end means that it gives us reason to act or refrain from acting in certain ways. The key thing here is that a human being has the power to set its own ends and in this way determine itself to action. And this capacity, according to Kant, has to be treated as an end in itself.

This capacity for self-determination is key to understanding why in Kant's view humanity is to be considered an end in itself. But before exploring this point in more detail I would like to explicitly bring out the connection of Kant's notion of end in itself with Kant's notion of dignity. If we follow for instance Patrick Riley's interpretation we can talk about:

"... the idea of a rational nature as an end-in-itself, or the dignity of a person ... as an end-in-himself..." (Riley 1979, p. 53).

In Riley's view Kant's suggestion is a straightforward one according to which what is an end in itself has dignity. But this thought requires explanation. According to Dean's interpretation of Kant:

"There is reason to think Kant took this idea of dignity to be included in the concept of something being an end in itself. When he introduces the term 'dignity', he uses it interchangeably with 'inner worth', and contrasts it with 'relative worth'. ... the claim that something is an end in itself directly implies that it also has a dignity" (Dean 2006, p. 37).

Dean's suggestion is that for Kant what has absolute worth has dignity, and in this sense what is an end in itself has dignity. As we have seen earlier for Kant there are basically two kinds of ends. Some ends have only a conditional (or relative) worth, in the sense that their worth for us is conditional to our fulfilling the desire of achieving that end. Other ends have unconditional or absolute worth, i.e. they are ends in themselves. Hill distinguishes between the two different kinds of ends in terms of dignity:

"Kant distinguishes (relative) personal ends from ends in themselves by saying that the latter have dignity whereas the former have only price" (Hill 1992, p. 47).

The distinction between dignity and price can be found on Kant in the following way:

"In the kingdom of ends everything has either a price, or a dignity. What has a price can be replaced with something else, as its equivalent; whereas what is elevated above any price, and hence allows of no equivalent has a dignity" (Kant 2012, p. 46).

The view is that for Kant what has dignity has no equivalent. So what has absolute worth is seen by Kant as having no equivalent. What has an equivalent has only a relative worth and is associated with the notion of price. Wood's interpretation of Kant helps us to understand what it means to say that what has absolute worth has dignity:



"... an end with absolute worth or (as Kant also says) dignity, something whose value cannot be compared to, traded off against, or compensated for or replaced by any other value" (Wood 1998, p. 170).

Hill explains the point in more detail:

"Dignity is characterized as "an unconditional and incomparable worth". ... it is a value not dependent upon contingent facts. Thus ... whatever has dignity has value independently of any effects, profit, or advantage which it might produce. ... No amount of price, or value dependent on contingent needs and tastes, can justify or compensate for sacrifice of dignity. We may express this by saying that what has dignity is priceless" (Hill 1992, pp. 47-8).

I want to call attention to a fact that I won't explore in detail now, but will come back to later. A central idea in Kant's conception of human dignity is thus that what has dignity has no equivalent. The idea is in itself interesting and I will develop it later in an Agent-Particularist fashion. But in the context of Kant's conception of dignity that comes out as an odd suggestion. We have seen that what has dignity for Kant is humanity in a person, and that humanity basically means one's rational capacity. And the rational capacity that has dignity is exactly the same for everyone. So there seems to be a contradiction involved in this, in the sense that what has dignity is at the same time exactly the same for everyone and is supposed to have no equivalent. As I said the idea that what has dignity has no equivalent is a nice one, but it seems to have no real substance in Kant's conception of dignity.

I now want to go back to the point I made earlier about the meaning of having something like humanity as an end. The view is that humanity refers to a person's rational capacity. This capacity is seen by Kant as an end in itself in as much as it gives the human being the power to set its own ends and determine itself to action. This means that according to Kant the dignity of a person comes from their capacity for self-determination. Kant sometimes refers to this capacity as legislation:

"... the will of a rational being must always at the same time be considered as legislating since it could not otherwise be thought as an end in itself. ... the dignity of a rational being that obeys no law other than that which at the same time it itself gives" (Kant 2012, p. 46).

As I have showed in Chapter Two Kant thinks that autonomy (positive freedom) follows from negative freedom. His view is that as negatively free we are free from the determination of causal laws of nature (free from the influence of inclination). But Kant thinks that rational beings can't avoid acting for reasons. And acting for reasons is understood in Kant in terms of acting according to laws. This means that even if the free person doesn't act according to causal laws of nature they still act according to some law. As we by now know Kant's view is that if we are not acting out of inclination we are acting out of reason, i.e. according to causal laws of reason. And as Hill points out:

"If we assume (with Kant) that one's nature as a rational will is in some sense one's "true" self ... then we could conclude further that the rational principles in question are 'one's own' or 'given to oneself by oneself' ..." (Hill 1985, pp. 111-2).

That is precisely what autonomy means: one follows one's own law, a law that one gives to oneself. Once we have ruled out the possibility that the person acts out of inclination, i.e. motivated by a hypothetical imperative, the only possible candidate is Kant's moral (universal) law – the categorical imperative. Kant sees it as the only kind of law an autonomous rational creature (in contrast with heteronomous – creatures that follow causal laws of nature or laws given by alien forces) would give to itself.

In Kant's view autonomy is worthy of respect. But it is not merely autonomy in my own person which is worthy of respect but rather that capacity for autonomy is worthy of respect wherever it can be found,

including other rational beings. As Paul Guyer's interpretation of Kant suggests:

"The thought that there is an incomparable value or dignity in self-mastery over nature, which can be attained only through the willing of universal law, is what is ultimately necessary in order to explain why rational being is an end both in one's own person and in that of everyone else. The dignity of autonomy is what makes the preservation and promotion of one's own rational nature an absolute end; but it makes humanity an end in everyone else as well because one's own autonomy is not desired merely as a matter of personal preference, but as something of incomparable dignity wherever it occurs" (Guyer 1998, p. 237).

So Kant's dignity is grounded on a person's capacity to follow only laws that it gives itself (i.e. laws that are given by reason alone):

"... the dignity of humanity consists in just this capability, to be universally legislating, if with the proviso of also being itself subject to precisely this legislation" (Kant 2012, p. 51).

Kant's view is that an autonomous rational being is one that follows only laws that it gives to itself. But we need a careful understanding of what that means. First we need to understand what it means in Kant to follow a law given by oneself. It doesn't mean just any law that one gives to oneself. Laws that recognize subjective ends or laws that come from one's inclinations (hypothetical imperatives) aren't laws one gives to oneself. Rather they are laws given by nature and so following them isn't really acting autonomously for Kant. This is what Riley's quote below suggests:

"The reformulated categorical imperative is a "law of one's own giving", a law legislated by one's own will, but not in terms of a mere "interest" such as happiness. The moral laws to which a man is "subject," he urges, are given by his own will—a will which, however, is "designed by nature to give universal laws." A will which determines itself by laws which recognize objective ends is autonomous, Kant says, while one which makes merely contingent ends the maxims of its action is heteronymous" (Riley 1979, p. 50).

This brings us to the issue of the form of the law which an autonomous creature gives itself. Since these laws are categorical imperatives rather than hypothetical ones Kant's view is that the laws one gives to oneself are universal laws as we saw in Chapter Two. This means that:

"... morality is the condition under which alone a rational being can be an end in itself... morality and humanity, in so far as it is capable of morality, is that which alone has dignity" (Kant 2012, p. 47).

So what we can see from Kant's conception of human dignity is that for him what has dignity is what is an end in itself. The only thing which is an end in itself is humanity. Humanity doesn't refer to a human being or the class of human beings, but rather to a specific capacity which Kant attributes to rational beings: a rational being's capacity for rationality. Specifically what is supposed to be worthy of respect is a rational creature's capacity for self-directed (autonomous) action. Autonomous action is understood as action done out of reason alone. And action done out of reason in its turn is understood as action which accords with universal laws of reason, i.e. laws which are universalizable. Action that follows universal laws of reason is the action of a moral being. So a rational being's capacity for morality has dignity for Kant.

### 3.1. Problems with Kant's Conception of Dignity and the Contrast with the Agent-Particularist Conception:

The point of looking at Kant's conception of human dignity is to use it as a way of fleshing out the Agent-Particularist conception of dignity. In this sense Kant's conception of dignity is rich and helpful, but it also has some fundamental problems. I will now discuss those problems in detail. I will also look again at the most useful notions involved in

Kant's account of dignity and develop them in an Agent-Particularist way.

The Agent-Particularist conception of human dignity works with a richer-genius conception of the self. Coming from that perspective a worry about Kant's understanding that humanity is what is worth respecting becomes clearer. There is a worry about that because as we have seen Kant's view is not that humanity refers to the class of human beings. Rather by humanity Kant refers to a kind of feature of human beings which is worth respecting. This means that it is not people that Kant thinks need respecting but some peculiar feature of them. And there is something odd about the idea of respecting someone's humanity without respecting the person. Hill points to an important implication of this:

"The set of traits Kant calls "rational nature" or "humanity" is not something which distinguishes one man from another but is something which men have in common and which marks them off from animals" (Hill 1985, p. 61).

When we note that by humanity Kant means merely their rationality the point becomes particularly pressing. In this case what is valuable about a person is just some rational capacity that they have. As the quote above suggests, the implication of that is that what is valuable about a person is the same in every human being. Since what is valuable in everyone is the same there is no sense of respecting the individual. We can see that from Hill's quote below, in which he explains Kant's idea of how an autonomous human being (a member of the kingdom of ends<sup>30</sup>) chooses which law to adopt:

"... in trying to decide which rules to make, the members will "abstract from personal differences." That is, they will disregard the various factors which distinguish individuals: for example, differences in appearance, height, weight, sex, race, family, heritage, special talents, social roles, and so on. It follows

---

<sup>30</sup> I won't go into the details of Kant's notion of the kingdom of ends. What is relevant for our purposes is that every autonomous human being is a member of the kingdom of ends, and that such kingdom includes all the objective laws that its members make.

that their rules will be universal in intent as well as in form. ... When one legislates as a member of the kingdom, it is as if he were ignorant of his own special traits and circumstances. ... Ignoring the factors which distinguish one man from another ..." (Hill 1985, p. 60).

So in Kant's conception there is no sense of an individual person having dignity. Even if that is already problematic it becomes even more so when we bring to the context of this discussion a point which I made in Chapter Two. I explained that from a neutral point of view (if one isn't working with a richer-genius conception of the self) we don't have an idea of what it is to be a particular person at all. As Hill points out Kant's understanding of one's rational will as one's true self:

"... raises deep questions about personal identity" (Hill 1985, p. 112).

In the end it is simply unclear what account can be given of what it is to be a person in Kant. Once we exclude everything that distinguishes a person from another and see the person's true self as their rational capacity, there is nothing substantive to say about personal identity at all. The only thing that is the person is common to every person. So not only in Kant there isn't a sense of an individual person having dignity but also there is nothing substantive which constitutes personal identity anymore.

This makes a sharp contrast with the Agent-Particularist conception of dignity. According to the Agent-Particularist conception each person is worthy of respect. The whole of the Agent-Particularist position is shaped to allow that possibility. The Agent-Particularist characterization of what it is to respect someone is to let them be who they really are rather than impose anything on them or subjugate them to anything other than themselves.

The structural feature of Agent-Particularism is that there is no further moral criticism that we can make of the person once they have

developed themselves into their ideal version. Once that is the case there can't be a moral criticism of someone because they don't do in a given situation what we would ideally want someone to do. This means that for Agent Particularism there isn't a way in which someone could be compared to another person and found less good because they are a different person. And that simply is respecting the individuality of people.

So in contrast with Kant's conception of dignity which fails to have a sense of a person having dignity and so requiring respect, the Agent-Particularist conception of human dignity has at its very core a demand to respect an individual person as the person they are. Whereas for Kant what is worthy of respect in people is the same for everyone, for Agent-Particularism it is the particular person who is worthy of respect. There is something about the uniqueness of each person that Agent-Particularism value as being worthy of respect.

There is also a worry in Kant's sense in which humanity is an end. As we have seen earlier Kant's view is that humanity limits what can be done towards a person. As Robert and Adam point out:

"... my own humanity as well as the humanity of others limit what I am *morally permitted* to do when I pursue my other, non-mandatory, ends" (Robert and Adam 2017).

Kant's view is that what it means for something to be an end in itself is for it to be a reason for acting or refraining from acting in certain ways. There is a clear sense in the case of giving it reason for refraining from action, but when we try to give a reason for acting it is more difficult. This means that the sense in which humanity is an end in Kant's view is basically a negative one. It provides a prohibition rather than an encouragement. The key thought is 'don't use them merely as a means'. Within Kant's conception of dignity a positive sense of treating humanity as an end is actually difficult to develop.

According to Korsgaard the closest that Kant comes to giving a positive sense to treating a person as an end is in what he calls imperfect duties. His view as we have seen is that we have a duty to keep our rational capacity in good shape. But the main way in which there may be something positive about treating a person as an end is in something like the view that we need to advance the ends of other people. In discussing the notion of the kingdom of ends Hill points out that an:

“... implication of the idea that members regard each other as ends in themselves is that they have a prima face concern to see each person's ends realized, or at least to ensure each person freedom to pursue his ends. Persons necessarily have ends, and one way of showing our special respect for persons is to favor rules which enhance their opportunity to satisfy their ends” (Hill 1985, pp. 61-2).

Hill talks of a concern to see other people's ends realized and adds, “or at least to ensure each person freedom to pursue his ends”. So even in this case the positive sense is not clear. But even if it were there would still be a worry about the idea that we need to foster other people's ends in order to treat them as an end. The key question is, what happens in cases in which their end is a bad one?

This brings us back to a point I made in Chapter Two about Kant's positive freedom. Kant understands negative freedom as freedom from being determined by empirical causes, while he understands positive freedom in terms of having autonomy. As we have seen earlier to have autonomy just means that one follows one's own law as opposed to externally given ones. The difficulty is that following one's own law as opposed to an externally given one is precisely what negative freedom is. So in the end because Kant is working with a blank conception of the self there is no real substance to his positive freedom.

We end up with a similar result in trying to distinguish between a negative and a positive sense of treating a person as an end. There is a clear negative sense of treating someone as an end in terms of it



limiting the pursuit of our other ends. But there is little substance to a positive sense of treating someone as an end. And fostering the ends of others (i.e. giving other people what they want), which is the substance that there is, is problematic. In order to try to do more in the case of his notion of the ends in themselves than he did in terms of a distinctive positive freedom Kant has got into a position which is just morally ambiguous. By contrast Agent Particularism has a much more morally straightforward positive side.

From the perspective of Agent Particularism there is also a clear negative side which limits what can be done to a person. We need to respect the person they are and not try to impose on them something which isn't in accord with their particular genius. There is also a clear positive sense of treating people as ends, which ties into the Agent-Particularist positive freedom and the idea of the development of each person into the best version of themselves. For Agent Particularism there is a dignity that comes when the human being flourishes, developing in accord with their particular genius to the fullest of their potential.

So on the Agent-Particularist conception of dignity when it comes to respecting a person there is no need for an extra requirement to make it positive, like the idea of giving people what they want. The position already has a positive side built into it that doesn't get us into the kind of moral ambiguity involved in just giving people what they want.

There is another problem with Kant's conception of dignity, and quite a serious one. For Kant what has dignity is humanity in a person and humanity is just understood as one's rational capacity. So in the end it is a person's rational capacity that has dignity. And what grounds that dignity for Kant is a person's capacity for autonomous action:

"Autonomy is thus the ground of the dignity of a human and of every rational nature" (Kant 2012, p. 48).

That is problematic because there is a worry about Kant's conception of autonomy, as we saw in Chapter Two. And if there is a problem with Kant's account of autonomy there is inevitably a problem about his conception of dignity. Autonomy means that the agent only follows their own law. It comes in in Kant as a consequence of the fact of our negative freedom. We are negatively free from the determination of alien forces coming from the empirical world, and as an inevitable consequence we are autonomous.

Another way to put it is by saying that we always act according to some law. And we basically follow either causal laws of nature or causal laws of reason. So when we are free from determination by causal laws of nature (i.e. we have negative freedom) the only alternative left is for us to follow causal laws of reason. Kant points out that in this case:

"Reason must regard itself as the author of its principles independently of alien influences" (Kant 2012, p. 54).

So basically in Kant's view the agent's true self as rational and negatively free gives to itself its own law. But we have seen in Chapter Two that the situation in which Kant sees autonomy is one in which the agent is following laws given to them by rationality itself independently of everything that is particular to the person, independently of everything about the person except for their rational capacity. So the autonomous act is one in which reason follows only laws that it gives to itself. This means in the end that it is not a person which is autonomous for Kant, but rather rational agency itself.

We already know that it is not a person that has dignity in Kant's account but humanity in a person. And humanity is understood as one's rational capacity. So we know that for Kant it is rationality that has dignity. The reason why rationality has dignity is because it has

autonomy. And autonomy has dignity because it means that the person themselves make the law. So one might think that the fact that dignity is grounded on the rational being's autonomy gave one a route to a more substantial and different account of what is worth respecting. But when one sees what autonomy amounts to for Kant - it just means that reason rather than nature determines action - then one sees that it doesn't bring in any more than rationality at all. Therefore in Kant's account of human dignity there is no sense of a person having dignity at all. The only thing that has dignity in Kant's account is rational agency itself.

On the Agent-Particularist conception of human dignity there is a much clearer sense of a person having dignity. It is the person itself that has dignity. Just their being the person they are calls for respect. And what grounds dignity is not just autonomy, though that is part of it. What grounds a person's dignity is negative freedom (i.e. autonomy) and positive freedom. It is autonomy and their developing into an ideal version of themselves.

This brings us to another point of contrast between Kant's conception of dignity and the Agent-Particularist conception. In the Agent-Particularist view there is something immediate and unconditional about the demand to respect a person. In Kant's account we must respect a person because of their humanity, i.e. rational capacity. And rational agency is capable of autonomous behaviour. So in the end Kant's view is that we must respect a person because they are a personification of reason.

For Agent-Particularism you must respect a person not because they meet some further condition. It is because they are who they are. There is something about not requiring them to be other than they are that is a really absolute and unconditional respect for a human being, and that is just not to be found in Kant. In fact this makes a point of

contrast even with Rousseau. Even though Rousseau is also working with a richer-genius conception of the self there is also something conditional in the need to respect a person.

His view is that we are good because we are natural, i.e. we are creatures of nature. And for Rousseau “[i]t is an incontestable maxim that the first movements of nature are always right” (Rousseau 2010, p. 161). His view is that nature is good, and so if the person is natural they are good. So it is natural to understand Rousseau as holding the view that we must respect a person because they are natural.

For Agent-Particularism it is not because they are natural that we must respect a person, it is because it is them. There is no further because. The reason we should respect them is that is who they are. There is a point at which we just accept that they are different people. For Agent Particularism once they are developing into the ideal version of themselves we just let people be. And there is something absolute and unconditional about the respect for individuals in terms of that. It is finding a dignity in humanity in each person just for the person they are.

Kant’s conception of human dignity has the worries which I pointed out above. But it is at the same time a rich view. In particular it has notions that when developed in an Agent-Particularist way can be helpful. In Chapter Two I have used some of Kant’s notions to present a natural way of developing the Agent-Particularist position. Something similar can be done here. So I will now use some of Kant’s notions as a natural way of developing the Agent-Particularist conception of dignity.

For Kant we must not treat a person as merely a means but always at the same time as an end. For Agent-Particularism as we have seen a person is worthy of respect just for the person they are. To use Kant’s language, they must not be treated as a means to the pursuit of one’s

ends. But how does it follow from the Agent-Particularist view that we shouldn't treat other people merely as means?

Agent-Particularism accepts each person as the person they are. And it is hard to see how we could have a consistent attitude which on the one hand accepts each person as the person they are and on the other is prepared to use people merely as an instrument for attaining one's ends. These attitudes seem impossible to combine. The problem with it is that it could only be an accident that someone who was treating a person as a means was not changing them to suit the aim that they were pursuing.

When we think in terms of using someone merely as a means, what we want is someone who would do a certain job in a certain situation. And we want the person to be the person that does that job, without caring about what is right for them to do. What we want is the job done. This means failing to respect them in order to get what we want.

According to Agent Particularism it is not a criticism of an individual that they aren't the person that we would ideally want in that position. There is a respect for the person as the person they are. There is no trying to change them in order to get what we want. It doesn't treat them as a means. So it is integral to Agent Particularism that it respects people and treats them as ends.

Any version of the universalizability view on the other hand inevitably does something quite like treating human beings as means. This connects with an idea I have presented earlier about universalizability and its demand for people to behave like some kind of ideal human being in situations. When someone is treating a person as a means they always want the person to be the person who they would ideally want in the situation and that is what the universalizability view does. It always wants each person to be the person we would want for some purpose and the ideal human being is the person we would want for

every situation. This is taking them as role performers rather than people. Because Kant's view is fundamentally a universalizability view he is ironically unable to treat people as anything other than means. A person is always the person we would want to do a certain job. Kant's thought about not treating people merely as means is a great one, but his own theory doesn't really accommodate it properly.

Another of Kant's ideas that actually work better if understood in an Agent-Particularist way is the idea that what has dignity has no equivalent. As we have seen earlier for Kant there are basically two kinds of ends. Some ends have only a conditional worth. Their value for us is dependent on the effects it might produce. Other ends have unconditional or absolute worth, i.e. their worth is independent of any effects it might produce – they are an end in itself.

As we saw earlier Kant also distinguishes between them in terms of the notions of price and dignity. What has absolute worth have dignity and what has only a relative worth has a price. What has only a relative worth has a price in the sense that it can be replaced by something of equal worth or as Kant puts it by something equivalent. What has dignity on the other hand can't be compared to some other thing, or exchanged by something else because it has no equivalent.

So a central idea in Kant's conception of human dignity is that what has dignity has no equivalent. And that is a brilliant thought. But in the context of Kant's conception of dignity it comes out as an odd suggestion (I anticipated this point earlier). We have seen that what has dignity for Kant is in the end one's rational capacity. And the rational capacity that has dignity is exactly the same for everyone. So there seems to be a contradiction involved in this, in the sense that what has dignity is at the same time exactly the same for everyone and is supposed to have no equivalent.

In this way it turns out that there isn't much substance to the idea that what has dignity has no equivalent in Kant's conception of dignity. Yet it is a useful idea, and works nicely in an Agent-Particularist conception of dignity. For Agent-Particularism what has dignity is simply the person as the person they are. The view is that each particular person is unique. And it is from that uniqueness that we can say that according to Agent Particularism what has dignity has no equivalent, because quite literally the view is that each person has no equivalent.

Finally I would like to look at Kant's notion of absolute worth (which I already addressed in Chapter Two) as another notion that is helpful and actually works better if understood in an Agent-Particularist way. Kant's view is that things and non-rational creatures have only a conditional worth, i.e. they have a worth for us when they are the objects of our actions. They can be replaced by something equivalent. The worth of a rational creature on the other hand doesn't come from our achieving some particular end that we might have. They have dignity and can't be compared with or exchanged for something else because they have no equivalent. They are an end in itself and in that sense have absolute worth.

But we have seen that in Kant's view it is not the person who has absolute worth or dignity. In the end for Kant it is the rational capacity – Kant's version of a person's true self - which is an end in itself and is taken by him to have absolute worth. And such rational capacity is exactly the same for everyone.

For Agent-Particularism on the other hand the true self that has absolute worth is not just the rational capacity of the person as Kant would have it. It is the person as the person they are. When we think of the true self in terms of one's particular and rich genius the idea that human beings have absolute worth acquires real substance. Instead of wanting to ground people's absolute worth in rational agency, itself

detached from everything about them as a particular person, Agent Particularism grounds people's absolute worth in their particular genius, in what makes them the person they are. This is what gives peculiar content to the Agent-Particularist conception of the notion of human dignity.

Behind the Agent-Particularist understanding of the notion of absolute worth is the idea that in order for someone to become moral at least part of what is required is that they develop into the ideal version of themselves. As I said in Chapter Two, in accepting the idea that people have an absolute worth the Agent Particularist accepts that we can't make sense of an agent not having a self which has an ideal development, and that we can't make sense of an ideal development of a person not being a good thing. This issues in the view that respecting people's dignity is a matter of letting them be themselves even if one dislikes the way they are. That is what it is to respect the dignity of a person just as a person.

#### 4. The Essence of the Agent-Particularist Conception of Human Dignity:

Human dignity refers to what deserves respect in people. On the Agent-Particularist conception of human dignity it is the person itself that has dignity and calls for respect. The Agent-Particularist characterization of what it is to respect someone is to let them be who they really are rather than impose anything on them or subjugate them to anything other than themselves. So it is integral to the Agent-Particularist position that it respects people's individuality.

For Agent Particularism you must respect a person not because they meet some further condition. It is because they are who they are.



There is something about not requiring them to be other than they are that is a really absolute and unconditional respect for a human being. The structural feature of the position is that there is no further moral criticism that we can make of the person once they have developed themselves into their ideal version. This means that Agent Particularism acknowledges that there is a limit to how much someone can be criticized.

Agent Particularism itself means that there can't be a moral criticism of someone because they don't do in a given situation what we would ideally want someone to do. For Agent Particularism that is an illegitimate demand and an illegitimate criticism, because it is wrong to criticize someone for not doing what we ideally want if it is not demanded of them under the ideal development of their particular genius. The position consists in accepting that this person is not worse than the person that behaves as we would ideally want. This fact about the Agent-Particularist position is itself a respect for the individual. For Agent-Particularism there isn't a way in which someone could be compared to another person and found less good because they are a different person. It requires them to reform themselves to become the ideal version of themselves but it does not require them to be other than they are.

Agent Particularism makes a distinction between moral disapproval and dislike. It accepts that we can disapprove of somebody morally and dislike them because they are morally mistaken. But it is a fundamental feature of the view that those will be cases in which they are failing to be the ideal version of themselves. Anything we dislike about somebody else when they have improved themselves to their ideal limit is merely dislike and can't be properly understood as moral disapproval. It is merely a matter of taste.

For Agent Particularism each particular person is unique, and there is something about that uniqueness that it values as being worthy of respect. It is this sense of each person being unique that brings the Agent Particularist to claim that what has dignity has no equivalent, and can't be traded by or compared with anything else. Agent Particularism accepts the individual for who they are, not for how well they live after some general standard. Agent Particularism is about the dignity of each human being.

From the perspective of Agent Particularism there is a clear negative sense of treating people as ends rather than means which limits what can be done to a person. There is also a clear positive sense of treating people as ends, which ties into the Agent-Particularist conception of positive freedom and the idea of the development of each person into the best version of themselves. For Agent Particularism there is a dignity that comes when the human being flourishes, developing in accord with their particular genius to the fullest of its potential.

It is integral to Agent Particularism that it respects people and treats them as ends. For this view the true self of a person has absolute worth. When we think of the true self in terms of one's particular and rich genius the idea that human beings have absolute worth acquires real substance. In accepting the idea that people have an absolute worth the agent particularist accepts that we can't make sense of an agent not having a self which has an ideal development, and that we can't make sense of an ideal development of a person not being a good thing.

Agent Particularism grounds people's absolute worth in what makes them the person they are. This issues in the view that respecting people's dignity is a matter of letting them be themselves even if one disagrees with the way they are. That is what it is to respect the dignity of a person just as a person. And this is what gives peculiar content to the Agent-Particularist conception of the notion of human dignity.

## Bibliography

Allison, H.E. (1998) 'Morality and freedom: Kant's reciprocity thesis' in Guyer, P. *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. Boston: Rowman & Littlefield. pp. 273-302.

Allison, H.E. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Anon (2017) 'Genius' in *Oxford English Dictionary* [online] Available at:  
<http://www.oed.com.ezproxy.sussex.ac.uk/view/Entry/77607?redirectedFrom=genius> (Accessed: 10 April 2015).

Anon (2017) 'Dignity' in *Oxford English Dictionary* [online] Available at:  
<http://www.oed.com.ezproxy.sussex.ac.uk/view/Entry/52653?redirectedFrom=dignity#eid> (Accessed: 22 June 2017).

Ashford, E., Mulgan, T. (2012) 'Contractualism', *The Stanford Encyclopaedia of Philosophy*. Available at:  
<http://plato.stanford.edu/entries/contractualism/> (Accessed August 2014).

Bacin, S. (2015) 'Kant's Idea of Human Dignity: Between Tradition and Originality', *Kant-Studien*, 106(1), pp. 97-106.

Backhurs, D. (2000) 'Ethical Particularism in Context', in Hooker, B. and Little, M.O. *Moral Particularism*. Oxford: Oxford University Press. pp. 157-177.

Berlin, I. (1958) 'Two Concepts of Liberty', in Berlin, I. *Liberty*. Oxford: Oxford University Press. pp. 15-35.

Bittner, R. (1989) *What Reason Demands*. Translated by Theodore Talbot. Cambridge: Cambridge University Press.

Bojanowski, J. (2015) 'Kant on Human Dignity: A Response to Oliver Sensen', *Kant-Studien*, 106(1), pp. 78-87.

Carter, I. (2012) 'Positive and Negative Liberty', *The Stanford Encyclopaedia of Philosophy*. Available at:  
<http://plato.stanford.edu/entries/liberty-positive-negative/>  
 (Accessed: 21 May 2014).

CNRS (2005) 'Genie', *Centre Nationale de Ressources Textuelles e Lexicales*. Available at: <http://www.cnrtl.fr/definition/genie> (Accessed: 10 April 2015).

Crisp, R. (2000) 'Particularizing Particularism', in Hooker, B. & Little, M.O. *Moral Particularism*. Oxford: Oxford University Press, pp. 23-47.

Crisp, R. and Slote, M. (1997) *Oxford Readings in Philosophy: Virtue Ethics*. Oxford: Oxford University Press.

Crowell, S. (2016) 'Existentialism', *The Stanford Encyclopaedia of Philosophy*. Available at: <https://plato.stanford.edu/archives/spr2016/entries/existentialism/> (Accessed: 30 May 2017).

Cullity, G. and Holton, R. (2002) 'Particularism and moral theory', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 76, pp. 169-209.

Dummett, M. (1978) *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press.

Dancy, J. (1981) 'On Moral Properties', *Mind*, 90(359), pp. 367-385.

Dancy, J. (1993) *Moral Reasons*. Oxford: Blackwell Publishers.

Dancy, J. (2006) *Ethics Without Principles*. Oxford: Clarendon Press.

Dean, R. (2006) *The Value of Humanity in Kant's Moral Theory*. Oxford: Clarendon Press.

Emerson, R. W. (1841) *Self-Reliance*. Available at: <http://www.emersoncentral.com/selfreliance.htm>. (Accessed: 01 August 2016).

Ferrara, A. (1993) *Modernity and Authenticity: a study of the social and ethical thought of Jean-Jacques Rousseau*. Albany: State University of New York Press.

Formosa, P. (2014) 'Dignity and Respect: How to Apply Kant's Formula of Humanity', *Philosophical Forum*, 45(1), pp. 49-68.

Fromm, E. (1997) *The Fear of Freedom*. London: Routledge Classics.

Geuss, R. (2001) *History and Illusion in Politics*. Cambridge: Cambridge University Press.

- Gewirth, A. (1969) 'The non-trivializability of universalizability', *Australasian Journal of Philosophy*, 47(2), pp. 123-131.
- Guyer, P. (1998) *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. Boston: Rowman & Littlefield Publishers.
- Guyer, P. (1998) 'The Possibility of the Categorical Imperative', in Guyer, P. *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. Boston: Rowman & Littlefield Publishers. pp. 215-246.
- Halwani, R. (2003) *Virtuous Liaisons: Care, Love, Sex, and Virtue Ethics*. Chicago: Open Court.
- Hare, R.M. (1961) *Language of Morals*. Oxford: Oxford University Press.
- Hare, R.M. (1963) *Freedom and Reason*. Oxford: Oxford University Press.
- Hare, R.M. (1981) *Moral Thinking*. Oxford: Oxford University Press.
- Hare, R.M. (1989) *Essays in Ethical Theory*. Oxford: Oxford University Press.
- Heyd, D. (2015) 'Supererogation', *The Stanford Encyclopaedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/supererogation/> (Accessed: 18 August 2017).
- Hill, T.E. Jr. (1992) 'Humanity as an End in Itself', in Hill, T.E. Jr. *Dignity and Practical Reason in Kant's Moral Theory*. New York: Cornell University Press. pp. 38-57.
- Hill, T.E. Jr. (1992) 'Kant's argument for the rationality of moral conduct', in Hill, T.E. Jr. *Dignity and Practical Reason in Kant's Moral Theory*. New York: Cornell University Press. pp. 97-122.
- Hill, T.E. Jr. (1992) 'The Kantian Conception of Autonomy', in Hill, T.E. Jr. *Dignity and Practical Reason in Kant's Moral Theory*. New York: Cornell University Press. pp. 76-96.
- Hill, T.E. Jr. (1992) 'The Kingdom of Ends', in Hill, T.E. Jr. *Dignity and Practical Reason in Kant's Moral Theory*. New York: Cornell University Press. pp. 58-66.
- Holmes, R. L. (1966) 'Descriptivism, Supervenience, and Universalizability', *The Journal of Philosophy*, 63(5), pp. 113-119.

Hooker, B. (2000) 'Moral Particularism: Wrong and Bad', in Hooker, B. and Little, M. O. *Moral Particularism*. Oxford: Oxford University Press, pp. 1-22.

Hooker, B. and Little, M. O. (2000) *Moral Particularism*. Oxford: Oxford University Press.

Humboldt, W.V. (2014) *The Sphere and Duties of Government*. Mansfield Centre: Martino Publishing.

Hume, D. (1777) *An Enquiry Concerning the Principles of Morals*. Available at: [http://econfaculty.gmu.edu/klein/Assets/Hume\\_Morals.pdf](http://econfaculty.gmu.edu/klein/Assets/Hume_Morals.pdf). (Accessed on 26 July 2017).

Hume, D. (2011) 'A Treatise of Human Nature'. In: Hume, D. *The Essential Philosophical Works*. London: Wordsworth Editions Limited.

Hursthouse, R. (2010) *On Virtue Ethics*. Oxford: Oxford University Press.

Hursthouse, R. (2013) 'Virtue Ethics', *The Stanford Encyclopaedia of Philosophy*. Available at: <http://plato.stanford.edu/archives/fall2013/entries/ethics-virtue> (Accessed: 09 October 2016).

Hursthouse, R. and Pettigrove, G. (2016) *Virtue Ethics*. Available at: <https://plato.stanford.edu/entries/ethics-virtue/>. (Accessed: 31 July 2017).

Johnson, R. and Cureton, A. (2017) 'Kant's Moral Philosophy'. *The Stanford Encyclopedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/kant-moral/> (Accessed: 17 August 2017).

Jollimore, T. (2011) 'Impartiality', *The Stanford Encyclopaedia of Philosophy*. Available at: <http://plato.stanford.edu/entries/impartiality/#ImpUni>. (Accessed: July 2014).

Kant, I. (2001) *Religion and rational theology*. Translated by Allen Wood and George di Giovanni. Cambridge: Cambridge University Press.

Kant, I. (2012) *The Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor and Jens Timmermann. Cambridge: Cambridge University Press.

Korsgaard, C. (1989) 'Morality as freedom', in Yovel, Y. *Kant's Practical Philosophy Reconsidered*. Jerusalem: The Hebrew University of Jerusalem. pp. 23-48.

Korsgaard, C. (1996) 'Kant's Formula of Humanity', in Korsgaard, C. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press. pp. 106-132.

Korsgaard, C. (2005) *The Sources of Normativity*. Cambridge: Cambridge University Press.

Kramer, M.H. (2005) 'Supervenience as an Ethical Phenomenon', *The American Journal of Jurisprudence*, 50, pp. 173-224.

Levin, M.E. (1979) 'The Universalizability of Moral Judgments Revisited', *Mind*, New Series, 88(349), pp. 115-119.

Locke, D. (1968) 'The Trivializability of Universalizability', *The Philosophical Review*, 77(1), pp. 25-44.

MacCallum, G.C. Jr. (1967) 'Negative and Positive Freedom', *The Philosophical Review*, 76(3), pp. 312-334.

Macintyre, A. (1957) 'What Morality Is Not', *Philosophy*, 32, pp. 325-335.

Mackie, J. L. (1977) *Ethics: Inventing Right and Wrong*. London: Penguin Books.

Maclean, Anne (1984) 'What Morality is', *Philosophy*, 59(227), pp. 21-37.

McDowell, J. (1979) 'Virtue and Reason', *The Monist*, 62(3), pp. 331-350.

McDowell, J. (2002) *Mind, Value and Reality*. London and Massachusetts: Harvard University Press.

McDowell, J. (2002) 'Non-Cognitivism and Rule-Following', in McDowell, J. *Mind, Value and Reality*. London and Massachusetts: Harvard University Press. pp. 198-218.

McDowell, J. (2002) 'Values and Secondary Qualities', in McDowell, J. *Mind, Value and Reality*. London and Massachusetts: Harvard University Press. pp. 131-150.

Mendus, S. (2006) 'Innocent Before God: Politics, Morality and the

Case of Billy Budd', *Royal Institute of Philosophy Supplement*, 58, pp. 23-38.

Meyer, M.J. (1987) 'Kant's Concept of Dignity and Modern Political Thought', *History of European Ideas*, 8(3), pp. 319-332.

Mill, J.S. (1991) *On Liberty and Other Essays*, ed. John Gray. Oxford: Oxford University Press.

Moore, G. E. (1948) *Principia Ethica*. Cambridge: Cambridge University Press.

Morris, M. (2015) 'Realism and Representation: The Case of Rembrandt's Hat', *European Journal of Philosophy*, 23(4), pp. 909-932.

Nafsika, A. (2013) *Virtue Ethics*. London: Bloomsbury Publishing.

Neiman, S. (2016) *Why Grow Up? Subversive thoughts for an infantile age*. London: Penguin Random House.

Nelson, E. (2005) 'Liberty: One Concept Too Many?', *Political Theory*, 33(1), pp. 58-78.

Nelson, W. (2008) 'Kant's Formula of Humanity'. *Mind*, New Series, 117(465), pp. 85-106.

O'Neill, O. (1989) *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press.

Plato (1952) *Gorgias*. Translated by W. C. Helmbold. Michigan: The Liberal Arts Press. Available at: <http://babel.hathitrust.org/cgi/pt?id=mdp.39015005370005;view=1up;seq=8> (Accessed: 01 December 2015).

Plato (2014) *Phaedrus*. Translated by Benjamin Jowett. The University of Adelaide Library: eBooks@Adelaide. Available at: <https://ebooks.adelaide.edu.au/p/plato/p71phs/complete.html> (Accessed: 27 August 2016).

Plato (2014) *The apology*. Translated by Benjamin Jowett. The University of Adelaide Library: eBooks@Adelaide. Available at: <https://ebooks.adelaide.edu.au/p/plato/p71ap/complete.html> (Accessed: 01 March 2017).

Pollock, L. (1973) 'Freedom and Universalizability', *Mind*, New Series, 82(326), pp. 234-248.



Riley, P. (1979) 'Kant on persons as "ends in themselves"', *Modern Schoolman*, 57, pp. 45-55.

Ross, D. (2007) *The Right and the Good*. Oxford: Oxford University Press.

Rousseau, J. J. (1754) *Discourse on the Origin of Inequality*. Translated by Victor Gourevitch. Cambridge: Cambridge University Press.

Rousseau, J. J. (2010) 'Emile or On Education', in *The Collected Writings of Rousseau*, 13. Translated and edited by C. Kelly and A. Bloom. Hanover: Dartmouth College Press.

Rousseau, J.J. (1968) *Julie, or The New Eloise*. Translated by J. H. McDowell. Pennsylvania: The Pennsylvania State University Press.

Scanlon, T. M. (1998) *What We Owe To Each Other*. Cambridge, Massachusetts: Harvard University Press.

Schneewind, J.B. (1992) 'Autonomy, obligation and virtue: an overview of Kant's moral philosophy', in Schneewind, J.B. *Essays on the History of Moral Philosophy*. Cambridge: Cambridge University Press.

Sensen, O. (2009). 'Kant's Conception of Human Dignity', *Kant-Studien*, 100(3), pp. 309-331.

Sensen, O. (2011) *Kant on Human Dignity*. Berlin: Walter de Gruyter GmbH.

Sensen, O. (2015) 'Kant on Human Dignity reconsidered', *Kant-Studien*, 106(1), pp. 107-129.

Sidgwick, H. (1962) *The Methods of Ethics*. London: Macmillan and Company.

Singer, P. (1989) 'All Animals are Equal', in Singer, P. and Regan, T. *Animal Rights and Human Obligations*. New Jersey: Prentice-Hall.

Skinner, Q. (2001) *A Third Concept of Liberty*. (The Isaiah Berlin Lecture), *Proceedings of the British Academy*, 117, pp. 237-68.

Skorupski, J. (2006) *Why Read Mill Today?* New York: Routledge.

Spence, J. (1966) *Observations, anecdotes, and characters of books and men*. Ed. James M. Osborn, vol. 2. Oxford: Clarendon Press.

- Sullivan, R.J. (1989) *Immanuel Kant's Moral Theory*. Cambridge: Cambridge University Press.
- Taylor, C. (2003) *The Ethics of Authenticity*. Cambridge, Massachusetts: Harvard University Press.
- Taylor, C. (2006) 'Winch on Moral Dilemmas and Moral Modality', *Inquiry: An Interdisciplinary Journal of Philosophy*, 49(2), pp. 148-157.
- Von der Pfordten, D. (2009) 'On the Dignity of Man in Kant', *Philosophy*, 84(3), pp. 371-391.
- Ward, A. (1973) 'Morality and the thesis of Universalizability', *Mind*, 82 (326), pp. 289-291.
- Wiggins, D. (1987) 'Truth as Predicated of Moral Judgements', *Needs, Values, Truth*. Oxford: Blackwell.
- Wiggins, D. (1987) 'Universalizability, Impartiality, Truth'. *Needs, Values, Truth*. Oxford: Blackwell.
- Wiggins, D. (1991) 'Categorical requirements: Kant and Hume on the idea of duty' *The Monist*, 74(1), pp.83-106.
- Wiggins, D. (2002) *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Oxford University Press.
- Williams, B. (1981) 'Internal and External Reasons', *Moral luck: Philosophical papers, 1973-1980*. Cambridge and New York: Cambridge University Press.
- Williams, B. (1985) 'Foundations: practical reason', *Ethics and the Limits of Philosophy*. London and New York: Fontana Press. pp. 54-70.
- Winch, P. (1965) 'Universalizability of Moral Judgements', *The Monist*, 49(2), pp. 196-214.
- Wood, A. (1998) 'Humanity as an End in Itself ' in Guyer, P. *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. Boston: Rowman & Littlefield Publishers.