The London School of Economics and Political Science

Identity and Reasons in Contemporary Liberal Theory

Andrej Keba

A thesis submitted to the Department of Government
of the London School of Economics for the degree of
Doctor of Philosophy, London, October 2009

UMI Number: U615319

UMI

Dissertation Publishing

UMI U615319

ProQuest®

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Abstract

Although it is commonly used in a variety of debates in normative political theory, the concept of practical identity is undertheorized and unclear. My dissertation aims to rectify this problem by providing a fully elaborated and philosophically robust account of identity. In developing this account I adopt the theory-indexed approach, which looks to define identity from the viewpoint of some normative conception of the person and society, rather than the viewpoint of shared but unsystematized intuitions about identity. Specifically, my enquiry is nested within the liberal perspective that affords individuals' reasons for action a central place in political justification. Starting from this theoretical background, I examine the prospects for defining identity as a determinant of individuals' reasons.

I first discuss some prominent current arguments that link identities to individuals' reasons. I maintain that the Rawls-inspired argument that identities generate reasons grounded in self-respect fails because Rawls's "social bases of self-respect" cannot be understood as the argument requires. Harry Frankfurt's view of identities as constraints on the individuals' power of willing rests, I suggest, on a flawed interpretation of volitional inability. Finally, I find that Christine Korsgaard's view of identities as the subjective grounds for maxim-adoption vacillates between an unacceptably naturalist understanding of identity and one that cannot account for the particularity of identities. Following these criticisms, I suggest that the link between identities and reasons needs to be investigated by reconsidering the notion of a reason for action. I argue that for a conception of reasons to be acceptable to liberals, it must present reasons as universal in scope. This condition is met by Kant's conception, according to which reasons consist in the conformity of maxims with certain objective principles. Accepting this conception, I argue that identity can be seen as a sensible, but inscrutable, condition for the possibility of universal reasons.

Table of Contents

Acknowledgements

I would like to thank my doctoral supervisor, Katrin Flikschuh, for her extraordinary patience and unswerving intellectual stimulation during a project that has taken many twists and turns.

My greatest personal debts are to my parents, Ivan and Nevenka Keba, whose unassuming support and encouragement have been boundless, and to Suman Sandhu, who has been the loveliest and most caring companion through turbulent times. This dissertation is dedicated to them.

1. From labels to reasons in the study of identity

1.1. Introduction

The objective of this thesis is to investigate the notion of personal identity as it is employed in debates in normative political theory. While the concept of identity is frequently and confidently used in a variety of such debates, there have been few attempts to define or explain it. This has led to uncertainty about the proper meaning of "identity". David Copp describes the situation as follows:

> Th[e] idea of "identity" is important both to moral and to political philosophy, but it has not been given an adequate philosophical explication. Indeed, it might be that there is not a single idea of "identity", but that instead there is a family of ideas that have not been well distinguished from one another.[1]

Another author states that "The concept of identity is... quite poorly circumscribed. Nothing in the nature of things dictates a particular usage and no convention has been elaborated to constrain its application."[2] Yet another writer observes:

> Although the English language is rich in synonyms, there are some words that are islands of desperate poverty in this respect.

---

[1] David Copp, "Social Unity and the Identity of Persons," *Journal of Political Philosophy* 10, no. 4 (2002): p. 365.

[2] Daniel Weinstock, "Is 'Identity' a Danger to Democracy?," in *Identity, Self-Determination and Secession*, ed. Igor Primoratz and Aleksandar Pavković (Aldershot: Ashgate, 2006), p. 15.

"Identity" is one of them. It stands in for so many different concepts that to use it at all is a recipe for confusion.[3]

Even within the sociological literature, which has been a major influence in the adoption of identity into the normative discourse, the term "tends to mean too much (when understood in a strong sense), too little (when understood in a weak sense), or nothing at all (because of its sheer ambiguity)".[4]

My aim in this thesis is to provide a clear and philosophically robust account of identity that is applicable in normative political theorizing. In this chapter I will present and defend one general approach to defining identity, as well as one particular orientation within that approach. The general perspective on identity that I will favour can be termed "theory-indexed". Its distinguishing mark is that it strives to define identity from the viewpoint of some normative conception of the person and society, rather than the viewpoint of shared but unsystematized intuitions about identity. I will introduce and argue for the theory-indexed perspective on identity in section 1.2. In section 1.3 I will discuss an especially popular account of identity that instantiates this general approach. This is Kwame Anthony Appiah's account of identities as descriptive "labels" that individuals attach to themselves. Its supposed basis, as will be seen, is Mill's theory of the development of individuality through choice. However I will argue, against Appiah, that the connection between Millian individuality and identities as self-applied descriptive labels is problematic. Therefore I will propose, in section 1.4, to develop an account of identity from the standpoint of another normative theory – the liberal conception of political justification through the use of reasons that hold for all citizens.

---

[3] Henry Harris, "Preface," in *Identity*, ed. Henry Harris (Oxford: Clarendon Press, 1995), p. v.

[4] Rogers Brubaker and Frederick Cooper, "Beyond 'Identity'," *Theory and Society* 29, no. 1 (2000): p. 1.

1.2. Two approaches to defining identity

Uncertainty about the proper meaning of the term "identity" can be resolved only by fixing the rules of its usage, or constraints on how the word "identity" is used. These rules will be given by a specification of some features of persons or the social world that the concept uniquely "picks out" and that set it apart from other concepts. In working out this unique set of features that belong to the concept of identity, two approaches are available. One of them is to define identity in relation to existing intuitions, or pre-theoretical notions about what falls under the concept, based on how it is ordinarily used and the situations in which using it seems appropriate.[5] The important aspect of intuitions, in this context, is that they are not a result of inference from some broader theory of the person or the society. Rather, they are beliefs and judgments that are independent of theoretical frameworks; they are natural or "gut" feelings about what seems right to say about identity. The intuitive approach to identity seeks to give a definition of identity that accommodates and, if possible, explains these natural feelings. The other perspective on conceptualizing identity stands diametrically opposite to the intuitive approach. It does not look for constraints on how the concept should be used in the untheorized beliefs of the users of the concept; rather, it starts from some normative theory, asking how "identity" ought to be understood in order to be consistent with that theory. This theory-indexed approach thus relates to the beliefs of users of the concept in a distinctive way: instead of accommodating accepted but unsystematized convictions about identity, it aims to produce an account of identity that those committed to some preferred theory *should* accept.

---

[5] The definition of intuitions I am working with here is very rough and loose, but serves to highlight the pre-theoretical nature of intuitions. A more stringent definition is given by Michael DePaul: "An intuition is just a belief in a proposition that (1) the person does not currently hold because of perception or introspection or memory or testimony or because the person has explicitly inferred the proposition, but (2) the person now holds simply because the proposition seems true to the person upon due consideration." (Michael R. DePaul, "Intuitions in Moral Inquiry," in *The Oxford Handbook of Ethical Theory*, ed. David Copp (New York: Oxford University Press, 2005), p. 595.)

The two approaches to identity have different aims. Minimally, the objective of the intuition-based approach is to deliver an account that systematizes diverse intuitions about identity. Typically, the hope is however also that intuitions can be brought under a concept that is practically useful: a concept that applies to and resolves existing disputes regarding identity, such as whether identities ought to be politically preserved or not, or whether a liberal state may legitimately promote a sense of liberal civic identity. On the intuitive approach, a concept of identity is constructed – pieced and assembled from shared convictions – in order to answer such questions. It is this anticipation of practical utility, more than anything else, that legitimates the activity of designing the concept. By contrast, the theory-indexed approach is not in the business of putting together a concept of identity from loose and potentially divergent material. Its starting point is an ordered system of thought, a theory of how to regard the man and society, from which a concept of identity must be drawn. The task here is to conceptually extend the theory to cover a new, and hitherto unaccounted for, aspect of individuals. What warrants the whole enquiry is not so much the expectation of settling normative disputes, but the goal of developing and enriching the theory. This activity can have practical import, but only indirectly: insofar as the overarching theory captures something about human practice, its extension will also have a bearing on real-world issues. The contestable points of both perspectives on identity are thus clear: in the case of the intuitive approach, it is the reliability and clarity of pre-theoretical intuitions; with the theory-indexed approach, it is the plausibility of the underlying theory.

I wish to present a case for theory-indexed conceptualization of identity by considering one version of the intuitive approach, and showing that it, too, relies on certain theoretical assumptions. While this does not amount to a positive defence of the theoretical approach, it at least shows that that approach cannot be easily dismissed in thinking about identity. This point can be illustrated by considering Copp's intuition-based analysis of identity. His notion of "self-esteem identity" fits the aforementioned characterization of the intuitive approach in its aim to capture the "central features of these intuitions [about

identity] better than the competing accounts".[6] It also promises to be "useful in a wide variety of contexts," from patriotism to the politics of multiculturalism.[7]

Copp begins his article by briefly reviewing some current understandings of identity: the philosophical accounts of Charles Taylor, David Miller, Christine Korsgaard, and Kwame A. Appiah, and the socio-psychological account of Michael Hogg and Dominic Abrams.[8] He sums up Taylor's approach to identity by saying that, according to Taylor, identity amounts to self-recognition – it is given by a person's answer to the question "Who am I?". Yet, counters Copp, that question may yield answers that are too trivial, or too bound up with the person's current situation, to be useful in thinking about identity. Copp charges Hogg and Abrams's similar account with the same error: it appears insufficiently selective because the person's 'self-image' is just down to her belief that she has such-and-such properties. But, intuitively, we do not count all of a person's characteristics among her identity, even she earnestly believes she has them, and in fact has them. We would not consider her statement that she is 6-foot tall, for example, a statement about her identity. Another account of identity in Copp's survey, that of Korsgaard and Miller, involves reference to a particularly entrenched system of personal values. However, suggests Copp, that reference implausibly rules out from the definition of identity those traits that are constitutive of the person's personality, but which the person *dis*values. Finally, Copp also distances himself from the "social identity" approach proposed by Appiah. While the latter captures an important dimension of *others'* response to a

---

[6] Copp, "Social Unity and the Identity of Persons," p. 369.

[7] Ibid.: pp. 365-66.

[8] The works Copp discusses are: Charles Taylor, "The Politics of Recognition," in *Philosophical Arguments* (Cambridge, MA: Harvard University Press, 1995).; David Miller, *On Nationality* (Oxford: Clarendon Press, 1995).; Christine M. Korsgaard, *The Sources of Normativity*, ed. Onora O'Neill (Cambridge: Cambridge University Press, 1996).; Kwame Anthony Appiah, "Identity, Authenticity, Survival: Multicultural Societies and Social Reproduction," in *Multiculturalism: Examining the Politics of Recognition*, ed. Amy Gutmann (Princeton: Princeton University Press, 1994).; and Michael A. Hogg and Dominic Abrams, *Social Identification: A Social Psychology of Intergroup Relations and Group Processes* (London: Routledge, 1988).

person's characteristics, it neglects the first-person perspective on identity that Copp is interested in.

However, Copp's criticisms of the rival accounts are themselves rooted in implicit theoretical pre-commitments, rather than merely reporting allegedly innocuous intuitions. The implicit theoretical background behind his supposedly intuitive points seems to be a hedonistic view of persons: namely, that persons are experiencers of pleasant or unpleasant emotions.[9] This view shapes his engagement with the rival philosophical accounts of identity, in that his objections to them presuppose the hedonistic view of persons. Taken at face value, without that background assumption, Copp's objections are overly quick and uncharitable. For instance, it is correct to say that for Taylor a person's identity consists in her answer to the question "Who am I?". However, Copp's criticism that answers to that question can be arbitrary or trivial is based on a misunderstanding of the context within which, for Taylor, the question arises. The question of identity is not triggered by the need for self-recognition, such as when one reviews an old school photograph; it is triggered by one's need to determine what one ought to regard as valuable. The proper context for the question is individuals' practical reasoning which, for Taylor, can only take place in view of fixed ideas of the good. These ideas are provided by the cultural structure of meaning and significance. To ask about one's identity is to ask about one's own place in relation to one's culture – the roles, norms and symbols of one's community. Individuals hence cannot arbitrarily invent their identities, but must work them out in dialogue with these determinate, historically given cultural components. To suggest otherwise, as Copp does, is possible only from

---

[9] Cf.: "[O]n the positive side, a person can feel satisfied, or feel sustained and enheartened, or feel boosted, bolstered or 'enhanced' by something. On the negative side, a person might feel worthless or despondent or have a sense of insecurity or lack of confidence. A person can feel shame, humiliation, or embarrassment, or feel disgraced, or feel discredited, or feel embarrassed or mortified. She can feel insulted. She can feel 'diminished' by something. All of these emotions can enter into a person's relevant feelings about herself." (Copp, "Social Unity and the Identity of Persons," p. 371.)

12

some preconceived standpoint that rejects Taylor's view of practical reasoning. But that just means that the criticism is theory-laden, not theory-free.[10]

Theoretical assumptions enter Copp's argument even more prominently when he elaborates his own notion of "self-esteem" identity. Based on the hedonistic view of persons, he defines identity as a set of putative facts about a person that ground that person's pleasant or unpleasant emotions about himself: "[A] person's 'identity' at a particular stage in her life is the set of propositions about her, each of which she believes, where her belief grounds an emotion of esteem."[11] Yet, it is unclear why putatively pure intuitions would mandate exactly *that* conceptualization of identity. To explain: Copp seeks support from psychological research that shows a link between self-conceptualization and self-esteem.[12] The link is clearly of interest to psychology, since that science is interested in detailing the various processes that govern human thought and behaviour. But the link, credible though it is, is not obviously relevant to political philosophy. For a psychological concept to be adopted in normative debates, such as those on nationalism and multiculturalism, it needs to be judged morally as well as descriptively appropriate. This is why, for instance, the theory of Pavlovian conditioning is not discussed by political philosophers, despite strong experimental evidence in its favour. The principal insight of that theory is that external stimuli can be used to induce desired physiological reflexes in conscious organisms, including humans. Although the theory is descriptively successful, it is normatively unacceptable since regarding persons merely through the prism of their behaviour, which can then be modified from the outside, is considered objectionable. Copp's decision to build his concept of identity around the idea of self-regarding emotions therefore presupposes that it is *valuable* that persons maintain a positive self-appraisal. Yet this claim is not self-justifying but must

---

[10] Copp similarly misrepresents Korsgaard's and Appiah's accounts of identity, but I cannot argue the point here. Appiah's and Taylor's views will be discussed in greater detail in the second half of this chapter, and Korsgaard's in chapter 5.

[11] Copp, "Social Unity and the Identity of Persons," p. 375.

[12] Ibid.: pp. 369-70.

rely on some further normative theory — for example, a theory of psychological well-being.

My claim that Copp's conceptualization of identity is held together by an implicit theoretical framework is not enough to discredit the intuitive approach to identity overall. This broader critique would require showing that a concept of identity can never be built upon the foundation of intuitions alone. It is plausible to assume that this type of argument would also raise deeper doubts about the status of intuitions in normative inquiry more generally. While I have sympathy for such claims, I cannot rehearse them here. What I can claim, though, is that the flaws I identified in a prominent version of the intuition-based approach provide enough of a rationale to consider the contrasting perspective. The principal payoff that Copp's perspective has to offer, a suitable systematization of our intuitions regarding identity, is of dubious value — since those intuitions seem to have been already ordered *prior* to the investigation into identity. Undoubtedly, other varieties of the intuitive approach can be formulated, and some of them perhaps will not be open to that objection. In the meantime, my suggestion is to try a different tack, with different rewards in store: to follow the contrasting theory-indexed route that starts from the constraints integral to some normative conception, with which the concept of identity must cohere. The hope is that the resulting concept of identity will enrich the underlying theoretical framework.

I will start by examining Appiah's influential theory-indexed account of identity. This account is remarkably attractive because it utilizes insights from other scientific disciplines — principally, sociology and social psychology — while remaining firmly rooted in the normative theory that sees individuality as a key component of persons' well-being.

1.3. Nominalism and individuality

In order to introduce Appiah's elegant and provocative conceptualization of identity it will be useful to first briefly sketch the broader context within which it is situated. Of particular importance is the relationship between the concern with

individuals' identity as a *practical* concept, as it figures in normative debates, and the one with individuals' identity as a *metaphysical* concept. In the general metaphysical context, the concept of identity denotes the constitutive properties of objects, with reference to which they can be identified as the objects they are, and re-identified through time. The question that the metaphysical notion of personal identity looks to answer is: which are the permanent and stable features of things that separate those things from other things, and guarantee their sameness despite changes to their other, non-essential characteristics? A typical test-case for this type of investigation is provided by the ancient tale of the "ship of Theseus", in which a ship is slowly and gradually repaired, plank by plank, until it ends up sharing no planks with the ship that first went in for repair. The problem is to determine whether it is meaningful to say that the original and the extensively modified ship are one and the same, and, if so, in virtue of what.

However, the identity of human individuals cannot be simply brought under the general heading of metaphysical identity of things and organisms. An enquiry into the identity of humans must take account of their unique ability to self-consciously make decisions and thus shape their own lives. This humans' ability, which sets them apart from not only inanimate objects but also other living beings, is to not only think or act but to also *know* that they are thinking or acting. Our existence is not merely a series of interactions with our environment; it contains also an "inner" dimension, a domain of our thoughts about ourselves in relation to the things that surround us, and to life in general. The issue of humans' identity therefore concerns not only the conditions of their being the beings that they are, but also the added factor of their own reflection on their constitutive features. The idea here is not that humans' self-reflection influences, or even fully determines, their metaphysical status – their numerical identity with themselves over time (as a special kind of entity among other entities, such as ships, trees, and ants). The point is merely that, although it must be presumed, the connection between humans' self-regard and their part in the order of nature remains unclear. This makes the topic of human identity especially difficult. Faced with this difficulty, the overwhelming majority of authors do not attempt to give an integrated account of identity, but instead side with one of two

approaches.[13] One of them abstracts from humans' self-consciousness, and regards their identity in parallel to the identity of inanimate objects and other living beings. The other approach divorces humans' self-regard from metaphysical issues from the opposite end, by confining the search for the essential properties of individuals to their first-person perspective. Its guiding question is: which of their features do persons regard as constitutive of themselves, as self-conscious beings? This latter perspective is summed up well by Bhikhu Parekh, when he writes that "To explore an individuals' identity is to ask what makes him who he is, how he views and relates to himself and the world, and why as a result he is this person and not anyone else."[14]

The overwhelming majority of authors that write about the ethical import of identity adopt the second approach to identity. It is concerned with what persons take to be the defining features of themselves as the persons that they are. Typically, the question asks about the particular ways in which persons *classify* themselves. An identity refers to a person's self-description in terms of categories such as witty, British, mother, risk-taker, trustworthy, and so on. In a passage representative of this approach Bernard Williams writes that identity

> has a sense which… relates to a type or a general thing. A gay or lesbian identity, a native American identity, or that of a Lombard as opposed to an Italian, are all type things, because such an identity is shared. Indeed, it is particularly important that it is shared, and an insistence on such an identity is an insistence on the ways in which it is shared."[15]

Furthermore, it is thought that to qualify as a constituent of a person's identity a self-description must be particularly important to the person – for example, one

---

[13] For a rare attempt at unifying the metaphysical and the practical perspective on identity see Kim Atkins, *Narrative Identity and Moral Identity* (New York: Routledge, 2008).

[14] Bhikhu Parekh, *A New Politics of Identity* (Basingstoke: Palgrave Macmillan, 2008), p. 9.

[15] Bernard Williams, "Identity and Identities," in *Identity*, ed. Henry Harris (Oxford: Clarendon Press, 1995), pp. 7-8.

that the person holds especially valuable, or that he is least prone to revise. It must "explain or underlie a lot of the individual's activities, emotions, reactions and, in general, life. It is from the point of view of those who endorse it a deep social classification."[16] This rules out from the scope of the concept of identity those self-descriptions that are trivial or ephemeral. For instance, looking out one's office window at children playing in the park, one might recognize one's exhaustion and envy at their careless existence. But this passing thought is not enough to conclude that the fatigue and the jealousy are parts of that individual's identity. These emotions capture the way he is feeling at the moment, but they do not describe his *person*, with a determinate character that stretches over time. Hence it can be said that an identity refers to any self-description that the person in question considers central to his being the person that he is. Despite his temporary jadedness, our office worker might continue with his work, intently and devotedly, because he recognizes its importance for his family's well-being. His identity, then, is that of a "provider". This approach to identity owes much to the sociological perspective of thinking about individuals' identity as their *social* identity, as "a way of locating ourselves in relation to other people".[17] As Hogg and Abrams explain, "just as we categorize objects, experiences and other people, we also categorize ourselves", which "causes one to perceive oneself as 'identical' to, to have the same social identity as, other members of the category – it places oneself in the relevant social category, or places the group in one's head".[18]

Nonetheless, although the view of human identities as essentially social, framed in terms of group categories, has wide currency in sociological circles, it raises a philosophical concern. The worry is whether defining identity as self-description

---

[16] Ibid., p. 9. Parekh similarly notes: "Not every distinguishing feature constitutes his [individual's] identity, only those that are an integral part of him, matter to him deeply, and in whose absence he would no longer be the same person." (Parekh, *A New Politics of Identity*, p. 9.)

[17] Michael A. Hogg, "Social Identity," in *Handbook of Self and Identity*, ed. Mark R. Leary and June Price Tangney (New York: Guilford Press, 2002), p. 462.

[18] Michael A. Hogg and Dominic Abrams, *Social Identifications* (London: Routledge, 1998), p. 19, original emphasis.

in terms of group categories can capture the sense in which identities are of, and refer to, discrete and unique *individuals*. One aspect of this thought is voiced well by Appiah: "If what matters about me is my individual and authentic self, why is so much contemporary talk of identity about large categories – gender, ethnicity, nationality, 'race', sexuality – t hat seem so far from individual?"[19] Appiah's complaint is normative, that regarding individuals through the prism of group memberships, even if they are self-attributed, reduces the multiplicity of their projects and attachments to that category alone. It diminishes their autonomy by corralling them into types of behaviour that are seen as "authentic" expressions of that category:

> There will be proper ways of being black and gay, there will be expectations to be met, demands will be made. It is at this point that someone who takes autonomy seriously will ask whether we have not replaced one kind of tyranny [that of neglect for identities] with another.[20]

The "identity model", Fraser similarly argues,

> puts moral pressure on individual members to conform to a given group culture. Cultural dissidence and experimentation are accordingly discouraged, when they are not simply equated with disloyalty. So, too, is cultural criticism, including efforts to explore intragroup divisions, such as those of gender, sexuality and class... The overall effect is to impose a single, drastically simplified group-identity which denies the complexity of people's lives, the multiplicity of their identifications and the cross-pulls of their various affiliations.[21]

---

[19] Appiah, "Identity, Authenticity, Survival: Multicultural Societies and Social Reproduction," p. 149.

[20] Ibid., pp. 162-63.

[21] Nancy Fraser, "Rethinking Recognition," *New Left Review* 3 (2000): p. 112. Following up on this worry, Anne Phillips argues with Fraser that the aim of authenticity is normatively

Appiah's work on the ethics of identity can be seen as a sustained attempt to chart a path between the two "tyrannies": the denial of equal dignity that comes with suppressing identities, and the reduction of persons' choices to group-specific ideals and types. In order to appreciate his way out of the dilemma it is necessary, first, to appreciate that he frames the difficulty with identity in distinctly Millian terms.[22] This is the sense in which Appiah's conceptualization is theory-indexed: it aims to resolve a problem that is conceivable *as a problem* only from the standpoint of Mill's conception of individuals' development. What Appiah means when he notes that rigid social categories may curtail persons' autonomy is that they undercut their possibilities for developing their *individuality*. By a person's individuality Appiah understands "the set of capacities she exercises and develops in managing her life well".[23] The ideal of the development of one's individuality (which is a distant, less metaphysically charged echo of Aristotle's views about the human function) is one of persons honing themselves to the best that they can be. As Mill memorably argues, an indispensable element of this pursuit is the free and extensive exercise of choice – not as an intrinsic end, but as a means of nurturing all of one's powers.[24] And it

---

undesirable, but also that it plays little part in justifying actual demands for recognition. More often than not, these demands amount to pleas for the acknowledgement of the group's distinctness, and equal validity of its political voice. According to Phillips, it is the value of equality, not authenticity, that drives identity politics. (Anne Phillips, "Recognition and the Struggle for Political Voice," in *Recognition Struggles and Social Movements*, ed. Barbara Hobson (Cambridge: Cambridge University Press, 2003).; Anne Phillips, *Multiculturalism without Culture* (Princeton: Princeton University Press, 2007).)

[22] See Kwame Anthony Appiah, *The Ethics of Identity* (Princeton: Princeton University Press, 2005), pp. 1-35.

[23] Kwame Anthony Appiah, "Reply to Gracia, Moody-Adams and Nussbaum," *Journal of Social Philosophy* 37, no. 2 (2006): p. 316.

[24] Cf.: "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 m aterials for decision, discrimination to decide, and when he has decided, firmness and self-control to hold to his deliberate decision. And these qualities he requires and exercises exactly in proportion as the part of his conduct which he determines according to his

is this irreplaceable instrument of self-improvement that is undercut by the constraints that accompany aligning oneself with rigid social categories. Appiah's solution to this problem is to claim that self-categorization is indispensable for, rather than inimical to, individuality and choice. In light of the just mentioned complaints about the "essentializing" tendencies of identity-discourse, this claim sounds surprising. As will be seen, it requires a sophisticated and original analysis of identity-claims, which I will examine now.

The first step in Appiah's programme is to show that self-categorization is compatible with the development of individuality. In order to make that claim he needs to show that self-categorization does not lock the person into fixed and immutable modes of behaviour. The idea that identities have a true, or authentic, essence to which identity-bearers must be true is hence clearly unsuited for Appiah's purposes. However, that idea cannot be dismissed out of hand as it seems to give the best account of the meaning of identity-statements, or self-categorizations. It explains, in other words, in what sense the categories to which individuals appeal have a real existence. It accounts for the sense that there must be *something* – a group, or a pattern of behaviour – to which individuals claim allegiance but which nonetheless has fixed boundaries. Without this fixity of categories, placing oneself under them would be meaningless: since there would be nothing real, nothing sharply delineated, about a label, using it would stand for nothing. If, for example, the application of the label "British" was completely unconstrained, there would be nothing that it *is* to be British. One simple and convenient way of avoiding this problem is to say that social categories are ontologically real, with a stable essence that logically precedes individuals' self-categorizations. On this interpretation of identity-statements, they mean that the person who makes them is a version of an ideal, or perfect, representative of the relevant category. To call oneself British is to claim some connection with prototypical and primordial "Britishness".

---

own judgment and feelings is a large one." (John Stuart Mill, *On Liberty*, ed. David Bromwich and George Kateb (New Haven: Yale University Press, 2003), p. 124.)

The view just presented can be termed the *realist* approach to identity-statements, after a position in the traditional philosophical debate about the status of universals. This debate, central to medieval philosophical discussions, focused on explaining how abstract signs – common terms that can apply to a multitude of objects, such as "horse", or "sweet" – can be predicated of particular objects. The realist answer is that particular objects that fall under a common term themselves have something in common, something that logically precedes the practice of calling them by the same name. Abstract terms point to the existence of abstract objects, of which particular objects are versions, or instantiations. In Plato's version of realism, these mind-independent, eternal, and primary objects are Forms. An object falls under a universal concept insofar as it imitates, or participates in, the universal Form that, for Plato, exists in a purely intelligible world, inaccessible to senses. We can call an animal a "horse" because it is, in some sense, a copy of the perfectly rational concept of the horse. The transcendental Form makes the talk of particular horses meaningful.

The realist approach to abstract terms seems unsatisfactory, in general as well as an explanation of identity-statements. One objection to it is that it does too little to justify its postulation of a contentious and, ultimately, mysterious world of pure categories – in other words, that the notion of logically basic categories is little more than a convenient, *ad hoc* account of how particular objects fall under universal terms. This was the criticism that the empirically-minded Aristotle levelled against Plato, and one that equally applies to the realist interpretation of identity-claims. The notion that our ontology should include basic and really existing social types, such as "the British", "the mother", and "the homosexual", seems too fanciful and wildly unsupported to be acceptable. One author who advanced claims of this sort, specifically with respect to national types, was Herder, with his Leibniz-inspired view of national characters as self-contained and self-directing monads. Herder thought that each culture has a "singular, wonderful, inexplicable, ineradicable" spirit.[25] However his mystical vision of

---

[25] Johann Gottfried von Herder, quoted in Samuel Fleischacker, *Integrity and Moral Relativism* (Leiden; New York: E.J. Brill, 1992), p. 167.

national spirit, as well as other attempts to reify social categories, is ontologically dubious and must be resisted.

The objection to the realist account of specific social categories is not the only one that can be raised against realism about identities. Although persuasive, it is also not the one that informs Appiah's own account of identity. While it is indeed difficult to show that there are really existing and pure types to which individuals correspond, and what they are, Appiah's opposition to realism runs deeper. For him, there is nothing outside the linguistic activity of categorization that underlies the application of social categories to individuals. Racial and other designations are entirely "socially constructed", which means that "there are no African Americans independent of social practices associated with the racial label".[26] Accordingly Appiah proposes the following definition of identity: a self-categorization X qualifies as an identity if (1) there is a social conception of Xs; (2) some people identify themselves as Xs; and (3) some people treat others as Xs.[27] By a social conception Appiah understands a set of shared views about what makes some social category distinctive – typically, these are loose and imprecise stereotypes about what it is to belong to the category. For an identity of, say, the British or the black, to exist these stereotypes also have to be affirmed and passed on by individuals. Furthermore, some persons' membership of the relevant group must be sufficient to provide reasons – for members as well as non-members – for dealing with them in certain ways.

Appiah's definition is rich and subtle, and warrants patient examination. However what primarily matters to me here are not its specific stipulations, but how it prepares the ground for Appiah's normative claims about identity. Specifically, I am interested in how Appiah's definition leads to an interpretation of identity-statements that is compatible with the development of individuality. In this regard the definition can be called markedly non-realist because it does not reify social types. The meaning of identity-statements therefore cannot be that they assert a connection between an individual and mysterious ontological

---

[26] Appiah, *The Ethics of Identity*, p. 23.
[27] Ibid., pp. 66-69.

entities called "culture", or "race", or "nation". Appiah calls his understanding *nominalist* because it "explains h ow the identities work by talking about the labels for them".[28] His claim is imprecise because it is not clear how the phrase "how identities work" should be understood. Nonetheless, it is possible to get a sense of what Appiah has in mind by returning to the earlier question of how abstract signs can be predicated of particular objects. The nominalist answer denies that particular objects that fall under a common term have anything deep and underlying in common: abstract terms are just names that we assign to particular objects, but that do not imply the existence of abstract objects. In the same vein Appiah argues that identities are just names, just labels that do not track any ontological entities. Consequently, the meaning of identity-claims does not consist in any connection between features of the person and a category prototype. Their meaning is just that the person adopts a certain mode of speaking about himself, or "script" as Appiah calls it.[29]

Claiming that identity-statements are just expressions of commitment to regard oneself in some particular, historically created and contingent, way fulfils the first task of Appiah's programme: it removes the tension between identities and individuality. To make sense of identities it is not necessary to postulate modes of being that represent their true or authentic cores. Identifying oneself with a group or a category therefore need not entail constraints on the person's freedom to develop and express his individuality. However, as I noted earlier, Appiah also wants to advance a further, stronger claim – that individuality *requires* identification. In making this point he appeals to the connection that Hacking,

---

[28] Ibid.: p. 365. He continues: "The main motivation for the nominalism is that it allows us to leave open the question of whether the empirical presuppositions of a labelling practice are correct. Since many social identities are like folk races in being shot through with false belief, this is a decided advantage." (Appiah, "How to Decide If Races Exist," p. 365.)

[29] When I place myself under a social category, writes Appiah, I "fit my life story into certain patterns—confirmation at puberty for a religious identity, tenure in your mid-thirties for a professorial one—and I also fit that story into larger stories; for example, of a people, a religious tradition, or a race. (Appiah, *The Ethics of Identity*, p. 68.)

borrowing from Elizabeth Anscombe, makes between actions and descriptions.[30] Hacking's discussion of "kinds of people" starts from the Foucauldian assumption that creating social labels is an active intervention into the social reality. Defining and naming social types such as "the child", "the depressed", "the perverse", does more than classify persons; it creates new ways of being a person. Devising a new social label "makes up" people by reconfiguring the horizon of social expectations and options that persons face. It opens up a whole new context for human action: it enables them to attend to new sorts of projects and pursue novel excellences, but also to experience distinctive setbacks and suffer fresh prejudices. In an instructive passage Hacking writes:

> People spontaneously come to fit their categories. When factory inspectors in England and Wales went to the mills, they found various kinds of people there, loosely sorted according to tasks and wages. But when they had finished their reports, millhands had precise ways in which to work, and the owner had a clear set of concepts about how to employ workers according to the ways in which he was obliged to classify them.[31]

People come to fit their categories because – and this is the principal thesis that Appiah takes from Hacking – all intentional action is done "under a description". That thesis consists in the apparently unremarkable claim that we cannot form an intention to perform an action which we cannot conceptualize. Cavemen, for example, could not intend to recite Shakespeare because that action was not among the options they could conceive. Equally, one could not intend to receive communion without having a notion of transubstantiation, which is in turn possible only if there exists the categorization "Catholic". Descriptions, not only of things around us but also of ourselves, ground our intentions. As Hacking

---

[30] See Ian Hacking, "Making up People," in *Reconstructing Individualism*, ed. Thomas C. Heller, Morton Sosna, and David E. Wellbery (Stanford: Stanford University Press, 1986).; Ian Hacking, *Rewriting the Soul* (Princeton: Princeton University Press, 1995), pp. 235-36.; G. E. M. Anscombe, *Intention* (Cambridge, MA: Harvard University Press, 1957).; G. E. M. Anscombe, "Under a Description," *Nous* 13, no. 2 (1979).

[31] Hacking, "Making up People," p. 223.

notes, "Who we are is not only what we did, do, and will do but also what we might have done and may do. Making up people changes the space of possibilities for personhood."[32]

The thesis of the dependence of actions on descriptions is nonetheless not trivial because it highlights the social embeddedness of choice. Identities do not, as the realist would have it, prescribe certain modes of being as uniquely possible. But they enable us to *conceive of* different modes of being, and thereby create options from which persons can choose in their efforts at self-improvement. When Appiah writes that we make our lives "*as* men and *as* women; *as* Americans and *as* Brits; *as* philosophers and novelists",[33] he is not merely reporting on people's opinions. He is saying something about the way in which persons must regard themselves in order to be able to reason practically. The only way in which we can develop their distinctive capacities is by making ourselves into a particular kind of person from a "tool kit of options made available by our culture and society."[34] In this regard, "To value individuality just *is* t o acknowledge the dependence of the good for each of us on relationships with others. Without these bonds, as I say, we could not come to be free selves, not least because we could not come to be selves at all."[35]

Appiah's analysis of identity is better elaborated and more sophisticated than most. It is also commendably sensitive to the difficulties with understanding the connection between individuals and social types, and to the different approaches to the meaning of identity-statements. Several aspects of Appiah's account of identity are open to questioning – from its underlying, and highly contentious, theory of the perfection of individuals' capacities, to its nowadays fairly

---

[32] (Ibid., p. 229.)

[33] Appiah, "How to Decide If Races Exist," p. 370, original emphases.

[34] Appiah, "Identity, Authenticity, Survival: Multicultural Societies and Social Reproduction," p. 155.

[35] A ppiah, *The Ethics of Identity*, p. 21, original emphasis. Ronald Dworkin comes from a similarly Millian position when he depicts his membership of the American political community as a "condition of a good life" for him. (Ronald Dworkin, *Sovereign Virtue* (Cambridge, MA: Harvard University Press, 2000), p. 261.)

uncommon nominalist stance.[36] However, I am here only interested in Appiah's account as a version of the theory-indexed approach to defining identity. My concern is whether it is a plausible version of that approach. In order to determine this, it is necessary to evaluate Appiah's conceptualization of identity on its own terms – in relation to the theory to which it is supposed to be indexed, rather than alternative normative frameworks or positions in the philosophy of language. My question is, therefore, whether Appiah's portrayal of identity really coheres with Mill's theory of the development of individuality through choice. If it does not, a different kind of theory-indexed construal of identity will need to be sought.

Unfortunately, it cannot be said that Appiah's account of identity satisfies the requirement that, by aiming to extend Mill's normative framework, it sets itself.

---

[36] For example, one might object that Appiah does not really explain what it is for a particular person to have an identity, as a general social type. This objection is an extension of the classic criticism of nominalism, which is that it "seems incapable of explaining the generality in thought and language". (Cynthia Macdonald, "Tropes and Other Things," in *Contemporary Readings in the Foundations of Metaphysics*, ed. Stephen Laurence and Cynthia Macdonald (Oxford: Blackwell, 1998), p. 331.) The general thought is that the nominalist view of language is, in an important sense, question-begging. It states that objects referred to by the same common name do not have anything in common ontologically, but does not explain in virtue of what they fall under the same common name. In the context of identities, even if the realist view of identity-bearers as versions of a common archetype is implausible, the nominalist position appears equally unsatisfactory. Since it treats identities as mere names, it fails to cast light on what warrants calling a multitude of persons by a same common name – in other words, on the criterion of correct categorization. It may be supposed that Appiah would wish to avert this criticism by insisting that identification is *self*-categorization. It is persons who attach identity-labels to themselves. Therefore it appears misplaced to ask about the criterion of correct categorization: it is appropriate for a person to call himself anything he *decides* to call himself. However this rejoinder is of dubious value because it destroys the generality of identities. If identification is self-justifying unilateral stipulation – of the type "I am French because I say so" – identities become a private rather than a common thing. Identification is then not a matter of placing oneself under an existing social type, but of tweaking the type so as to fit one's own, ultimately unconstrained, decision. Because it would undermine the generality of identities, the decisionist rejoinder would not be able to explain what it is to bear an identity. I cannot further discuss this objection or the way in which Appiah might respond to it here.

It looks to resolve a problem conceived in the terms of that framework by using resources that are available within that framework. Specifically, Appiah sets out to reconcile self-identification and choice by arguing that self-identification enables individuals to conceive of different options. If this argument is successful, it delivers a paradigm of identity that "matters for ethical and political life"[37] because it is rooted in Mill's defence of the value of individuals' choice. However it is unsuccessful. Therefore, Appiah's conceptualization of identity fails – not *tout court* but by its own theoretical lights, as it cannot be regarded as an extension of Mill's conception. It mistakenly claims connection with that conception because its claim about the ethical significance of identities cannot find support in Mill's normative claims.

The difficulty is that Appiah's nominalist outlook provides ultimately too slim a foundation for his Millian defence of the ethical import of identities. Specifically, it rules out the view that self-categorization makes intentional actions possible, which is Appiah's intended position. His argument for this view is quite terse. However it can be noted that it involves a remarkable transition: from the claim about the necessity of regarding *actions* under a description Appiah purports to show that one must regard *oneself* under a description. This transition cannot be made from the standpoint of nominalism about identity. Let me explain this point by elaborating on an earlier example, which is not Appiah's but can be used to illustrate his argument. It is a plausible enough thought that one can only intend to receive communion if one possesses the concept of transubstantiation. Now, the further claim that Appiah wants to add is the following: what makes it possible for anyone to have the concept of transubstantiation is the existence of the social category "the Catholic". By this he means that the Catholic identity creates distinctive ideals, excellences, and patterns of behaviour, one of which is attending the Eucharist. However this further claim cannot be borne out by Appiah's nominalism. Recall that identities are just *labels*. To be sure, there is a gamut of conventions, habits, and doctrines associated with social types. However, at bottom, identities are nothing more than names, manners of speaking about oneself and one's projects – and it is hard to see how, on its own,

---

[37] Appiah, *The Ethics of Identity*, p. 69.

attaching a certain label to oneself and others can make possible the communion. A great many things that are not just manners of speaking – a religious text, an ecclesiastic tradition, an appropriate social standing of the church, and so on – are also required for the communion to be conceivable, aside from the mere designation "Catholic". All these elements combined, it seems, co-operate in shaping a distinctively Catholic way of being a person.[38] Names, on their own, do not expand the horizon of human opportunities.

Appiah's attempt to connect identities to the Millian theory of individuality is unconvincing. The nominalist analysis of identity, which promises to fend off the essentializing tendencies of identity-discourse, turns out to be too slender to bear out the claim that identification is a precondition for choice. Therefore Appiah's particular version of the theory-indexed approach to defining identity cannot be accepted. This finding is important because Appiah's account represents one of the most prominent and sustained attempts at explaining identities. Given that it fails, it is legitimate to enquire into the possibilities for devising alternative theory-indexed understandings of identity. In the rest of this thesis I will look into the prospects for one such alternative account, which is rooted in a normative theory that is distinct from Appiah's Millianism but nonetheless very influential. This is the liberal theory of reasons-based justification, which I will briefly present in the next section. As I will show, the central place that conception affords to individuals' reasons for action provides an opportunity for an attractive conceptualization of identity.

---

[38] Appiah counters the argument I have just given by appealing to a famous sociological experiment. (Ibid., pp. 62-64.) In the "Robbers Cave" study two groups of eleven-year-old boys, previously unacquainted with each other, were placed in an isolated camp area and monitored over a period of four days. Left to their own devices, the two groups developed markedly different conventions, ideals of proper behaviour, and pastimes. Before long, a fierce competitiveness and sense of enmity arose between the groups, which however disappeared once the two groups were assigned to work on a common task. Appiah attributes the development of different social structures exclusively to the groups' adoption of different labels (the Rattlers and the Eagles). However, this suggestion is highly questionable in light of the variety of other factors influencing the groups' behaviour: absence of the boys' parents, the fact of group co-habitation, the camp setting, comparison with the other group, and so on.

1.4. Identities and reasons

The liberal theory of reasons-based justification states, in short, that all political principles must be justifiable by recourse to reasons that hold for all citizens. It can be introduced by, first, defining a reason as a consideration that, for the agent, counts in favour of some action.[39] By referring to reasons as considerations that favour some action, liberals adopt the *normative* perspective on reasons – a perspective that asks about the justification for acting. This approach can be usefully contrasted with the *explanatory* approach to reasons, which uses the notion of a reason in order to tell a coherent causal story about happenings in the world. Saying that a shipwreck survivor had a reason to drink saltwater is, from this point of view, perfectly acceptable. It serves to describe his behaviour as purposeful: even though we cannot comprehend what his reason was, his doing it shows that he must have had some reason for it. From the normative perspective, however, the notion of a reason is not descriptive but prescriptive: it tells us that an action is the sensible thing for the person in the circumstances. In this sense, reasons are always *good* reasons: they rationally recommend or support taking some action. Viewed from this angle, to say that an agent had a reason for doing something is not equivalent to saying that his purposeful behaviour produced certain outcomes in the world; it is to state that there was something in light of which his action was appropriate.

One common way of regarding the requirement for universal justification connects it to the issue of legitimacy.[40] On this view, which has roots in older liberal theories of the social contract, subjecting citizens to coercive measures that he could not consent to amounts to coercion. The condition of public justifiability is therefore a means of ensuring the legitimacy of the state's use of power. The liberal insistence on justification through universally acceptable

---

[39] This is a widely accepted definition of a reason. See for example Thomas M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Belknap Press, 1998), p. 17.; Jonathan Dancy, *Practical Reality* (Oxford: Oxford University Press, 2000), p. 1.

[40] See for example John Rawls, *Political Liberalism*, Expanded ed. (New York: Columbia University Press, 2005), p. 217.

reasons can be illustrated by briefly looking at the way in which reasons figure in the design of John Rawls's "original position". The appeal to the original position represents the contractarian aspect of Rawls's theory, in which he seeks to justify certain principles of justice by recourse to the consent of individuals under certain hypothetical circumstances.[41] This type of contractarianism is morally constrained, which means that among the stipulations that govern the construction of the choice-scenario are moral stipulations – principally, that the ensuing principles of justice must be fair.[42] The resulting conception of justice has an inevitable moral component because it is built into it from the very start.

At the heart of morally constrained contractarianism is the idea that justice results from individuals' choices grounded in proper reasons – namely, those reasons that are relevant to justice. In Rawls, reasons relevant to justice are those that are exclusively based on two interests of the choosing parties: their interest in developing and exercising their conception of the good, and their interest in developing and exercising their sense of justice.[43] For political principles to be just, according to Rawls, it must be possible to regard them as the preferred option of individuals guided only by these interests. However, he also recognizes that individuals are ordinarily not guided by such considerations. Making social organization dependent on the actual choice of persons in their real circumstances would therefore result in unjust principles. For that reason, Rawls places deliberators in a deliberately counterfactual context of choosing, the

---

[41] However, contractarianism is not the only aspect of Rawls's theory: the principal burden of justification in his conception is carried by the procedure of "reflective equilibrium". This issue is discussed in greater detail in chapter 7.

[42] For the distinction between morally constrained and morally unconstrained contractarianism see for example Christopher W. Morris, "Justice, Reasons, and Moral Standing," in *Rational Commitment and Social Justice: Essays for Gregory Kavka*, ed. Jules L. Coleman and Christopher W. Morris (Cambridge: Cambridge University Press, 1998), p. 189.; and Cynthia A. Stark, "Hypothetical Consent and Justification," *Journal of Philosophy* 97, no. 6 (2000): pp. 314-15.

[43] Although implicit in *A Theory of Justice*, the idea of "higher-order" interests as the basis for deliberation on principles of justice is explicitly discussed only in later Rawls's writings. See John Rawls, "Kantian Constructivism in Moral Theory," *Journal of Philosophy* 77, no. 9 (1980): p. 525.; and Rawls, *Political Liberalism*, pp. 73-75, 105-07.

"original position". That imagined standpoint isolates the parties' relevant reasons by radically restricting the set of facts that they can rely upon in deciding.[44] In particular, they are imagined as shorn of any of their particular circumstances, such as their natural talents, race, wealth, and so on. Deprived of these facts about themselves, Rawls argues, the choosing parties could only opt for political conditions that allow them to utilize and preserve their conception of the good and sense of justice. The design of the choice situation thus embodies procedural fairness, which gives us, ordinary citizens with full knowledge of their circumstances, reason to accept the contractors' principles.

In the rest of this dissertation I will examine the prospects for framing the concept of identity in terms of individuals' reasons. More specifically, I will ask whether or not identity can be seen as a determinant of individuals' reasons for action. One of the facts than might be thought to constitute a person's identity is that he endorses some evaluative self-description, as Christine Korsgaard thinks of identity; or, in Harry Frankfurt's conceptualization, that he has some commitment that he cannot bring himself to violate. Facts of this sort, it may be claimed, stand in an intimate connection with the person's reasons: they either generate certain reasons, or they constrain the range of reasons that it is possible for the person to have.

My question is therefore whether such facts function as what Peri Roberts calls the theoretical limits on reasons – as "assumptions that necessarily underpin our reasoning".[45] If culturally mediated and subjective identities do set limits to reasons in this way, Roberts argues, we may worry about reasons being "different for different people at different times and places".[46] For him, it is their potential threat to the universality of liberal justification that makes identities normatively significant. However since there is considerable uncertainty

---

[44] Cf.: "[T]he veil and other conditions of the original position are designed to focus our attention upon the reasons that are m orally relevant, and to exclude those that are not, to justifying principles of justice." (Samuel Freeman, *Rawls* (London: Routledge, 2007), p. 144.)

[45] Peri Roberts, *Political Constructivism* (London: Routledge, 2007), p. 120.

[46] Ibid.

regarding how to understand identity, as I have shown, it cannot be simply assumed that they pose this threat. My enquiry is therefore not primarily guided by the objective of defending the possibility of universally valid reasons. I will probe the thesis that identities constrain reasons in order to determine what can be plausibly said about identities, since there is currently no rigorous and widely accepted account of identity. This investigation has a normative dimension insofar as it looks to nest the concept of identity within the liberal conception of reasons-based justification. If identities are the roots of all reasons, they are an issue of some normative importance. They then determine what kind of reasons persons can have. And this, it seems, has implications for the kind of justifications that can be given to others, which is relevant to those who think that all political principles must be supported by shareable reasons.

The dissertation is organized as follows. Chapter 2 addresses the methodological question of whether it is possible to give an individualist analysis of identity, in the face of Charles Taylor's thesis of "irreducibly social goods". Chapter 3 considers the Rawls-inspired argument that identity constrains individuals' reasons insofar as they care to maintain their self-respect. Chapter 4 evaluates Harry Frankfurt's claim that identities determine persons' reasons by constraining their power of willing. Chapter 5 discusses Christine Korsgaard's view that identities are the subjective grounds for maxim-adoption, and as such lie at the basis of all reasons. Chapters 6 and 7 develop a Kantian account of identity as a transcendental condition of reasons, and chapter 8 concludes.

2. The possibility of individualism about identity

2.1. Taylor and the dynamic concept of the self

In this chapter I will discuss a comprehensive objection against the kind of account of identity that I wish to develop. In the previous chapter I announced that I am interested in the prospects for devising a concept of identity from within the theory of reason-based justification. One way of conceiving of identity from that standpoint, and the one I care to examine, is to think of it as a determinant of individuals' reasons for action. However, before developing such a concept of identity it is necessary to put to bed a well-elaborated line of argument that contests its very possibility. That argument is Charles Taylor's rejection of the "atomist" method of studying the society through an examination of the properties of individuals that compose it.

If successful, Taylor's critique discredits the "atomist", or analytical, approach to a variety of phenomena that involve humans in their social setting. I am here only concerned with how Taylor's arguments relate to the investigation of identity, and specifically to the project of devising a reasons-based concept of identity. That project is liable to Taylor's critique of the analytical method because it assumes that, since individuals are the only units of moral standing, reasons can only be of individuals. A concept of identity that is based in this perspective can therefore also admit only of individuals' identities. However, Taylor has argued that this narrowly individualist approach to identity impoverishes the reality of social and ethical life. There is, on his view, something intrinsically "undecomposable" about lived communal practice, which exists above and beyond any attempt to compartmentalize it. Whittling the worth of identity to a set of propositions about individual attachments is flawed from the start. A practical implication of this Taylor's view is that there can be no talk

of partial cultural accommodation of identities, only full-house autonomy for the community that shapes and nurtures identities. In this chapter I will present and evaluate Taylor's critique of individualism about identity, focusing on his thesis of "irreducibly social goods".

Taylor defines identity as follows:

> To know who I am is a species of knowing where I stand. My identity is defined by the commitments and identifications which provide the frame or horizon within which I can try to determine from case to case what is good, or valuable, or what is or ought to be done, or what I endorse or oppose. In other words, it is the horizon within which I am capable of taking a stand.[1]

This instructive passage is extremely compressed, weaving together several distinct and complex argumentative strands. Nonetheless, it can be summarized as follows. Its central theme is opposition to the analytical separation of the subject's identity from his actions and commitments as a member of a particular social world. Concepts, including ethical concepts, are based on prior engagement with the world, where this engagement is not random or arbitrary but guided by the horizon of shared understandings. The distinctive feature of this constitutive horizon, in turn, is its groundedness in the community's conception of what constitutes a good or fully-realized life. Since the conception is itself the product of a dynamic process, an expression of the community's unique way of life and deliberative practices, the subject's selfhood is also best conceived not as fixed but a matter of dialogical self-interpretation. Underpinning the self-interpretive project is an essentially communal understanding of the good.

As it may perhaps be gathered from this brief sketch, Taylor's philosophical treatment of identity incorporates two mutually reinforcing insights: one is the dynamic Hegelian notion of the emerging self; the other is the more Aristotelian

---

[1] Charles Taylor, *Sources of the Self* (Cambridge: Cambridge University Press, 1989), p. 27.

idea that self-interpretation can only be conducted in view of the good. In conjunction, the two notions work against two constituents of what Taylor terms the malaise of "atomism". The dynamic notion of the self is directed against that part of the atomistic outlook which Taylor calls "ontological individualism", the view according to which one "can and ought to account for social actions, structures, and conditions in terms of properties of the constituent individuals".[2] The contested view here is that societies must be understood and arranged from the perspective of individual human persons. However this claim is easily misunderstood. In particular, it may be thought that Taylor wishes to advocate an alternative picture of personhood – for example, one in which collectivities can also be moral persons. But Taylor is not a collectivist.[3] In fact, his objection to ontological individualism is not at all directed at the sort of entities that it picks out as subjects. Rather, his criticism targets its analytical method, which draws its roots from Descartes – and specifically its freezing of the social categories. The analytical approach in philosophy breaks down the object of its scrutiny into progressively smaller constitutive parts with the objective of observing their individual contributions to the working of the whole. In the study of society, this method manifests itself in the reduction of the lived practice of the society to the series of relations between distinct individual agents. An inevitable part of this reduction, argues Taylor, is the conceptual separation of individuals from each other and from the practices in which they are involved. Without this separation, the project of observing the way in which these disparate agents interact would be impossible.

However, Taylor claims, this strict division of the societal whole into neatly individuated agents is contrived and incoherent. In looking for neat and

---

[2] Charles Taylor, "Cross-Purposes: The Liberal-Communitarian Debate," in *Philosophical Arguments* (Cambridge, MA: Harvard University Press, 1995), p. 181.

[3] Collectivism about personhood is not a commonly held view nowadays. A rare recent proponent of it is Vernon Van Dyke, for example in Vernon Van Dyke, "Collective Entities and Moral Rights: Problems in Liberal-Democratic Thought," *Journal of Politics* 44, no. 12 (1982).; Vernon Van Dyke, "The Individual, the State, and Ethnic Communities in Political Theory," *World Politics* 29, no. 3 (1977).; and Vernon Van Dyke, "The Cultural Rights of Peoples," *Universal Human Rights* 2, no. 2 (1980).

manageable categories it compromises fidelity to reality. This is evident in two explanatory failures of the analytic concept of the individuated self: its failure in accounting for the nature of the self, and in expressing why selfhood is valuable. On the first note, that approach supposedly cannot explain how we relate to ourselves. This argument starts from the idea, which echoes Hegel's criticism of Kant's theoretical philosophy, that "our representations of things – the kinds of objects we pick out as whole, enduring entities – are grounded in the way we deal with those things."[4] The analytical approach runs counter to this idea by assuming that individuals can be regarded in abstraction from their engagement with the social world. What drops out of consideration within this perspective is that persons necessarily regard themselves under certain descriptions. In other words, they always think of themselves as such-and-such persons: as honest, witty, sporty, conniving, intellectual, Danish and so on. In identifying themselves in this way they cannot but use concepts borrowed from their culture, which is why the self is always coloured by communal ideas and ideals. Secondly, reduction also fails to acknowledge that one values and takes pride in oneself precisely because of the traits and dispositions captured in this comprehensive self-conception – not because one is an abstract moral unit, but because one is a human being of a certain sort: for example, a caring mother, or a good socialist. For Taylor, these deficiencies of the analytical account of selfhood warrant embracing an alternative, dynamic concept of the self. It is a concept in which, rather than a finished article, the self is a continuous project: the project of becoming a person of a determinate sort through ongoing interaction with others. Essential to this project is the process of receiving and granting recognition – that is, assurance of selfhood. Recognition is a dialectical notion, which means that affirming the selfhood of another person (as "the mother", "the socialist", or whatever) at the same time establishes one's own subjectivity.[5]

---

[4] Charles Taylor, "Overcoming Epistemology," in *Philosophical Arguments* (Cambridge, MA: Harvard University Press, 1995), p. 11.

[5] This view of the self is developed extensively in Georg Wilhelm Friedrich Hegel, *Phenomenology of Spirit*, trans. Arnold V. Miller (Oxford: Clarendon Press, 1977).

Although the dynamic conception of selfhood that Taylor adopts is developed most systematically by Hegel, the connection between recognition and unimpaired selfhood has earlier roots. Aspects of it can, for example, be found in Rousseau's sentimental romanticism about self-expression.[6] There is a way of being that is peculiar and appropriate to my own unique self, Rousseau writes, and happiness and self-realization can be only be found in a life that reaches this genuine mode of existence. But his writings also reveal a special kind of worry about the fragility of identity. Civilization has supplanted man's natural instincts with self-indulging inclinations: all traces of compassion have been eradicated, innate and innocent self-love exchanged for selfishness. The modern man is shut off from all that was once true and untainted about him, the great corruptors being the over-production of goods and, resulting from it, private property. They have imposed on humanity the stultifying dictate of uniformity answering to the demands of economic expediency. In this precarious position the duty of the alienated individual is to cry out in search of himself – to embark on the most penetrating project of self-discovery in which the only assurance of his own selfhood is the cry itself.[7]

But, so the argument continues, not only must the quest for authenticity be given expression, it must also be publicly recognised as worthwhile. Without recognition the quest is futile, for the individual's objective of asserting that he is somebody, somebody worthy of respect, is baldly and easily defeated when the enormous strength of common opinion is harnessed into convincing him that he is an insignificant speck. Or, for that matter, when his plea is plainly ignored. This worry is adopted and taken extremely seriously by contemporary identity-

---

[6] On a radical interpretation, Rousseau can even be seen as a direct predecessor of Hegel's dynamic concept of the self. Guignon thus notes: "For the more authentic form of self-revelation Rousseau envisions, what the self-portrait presents is not a faithful copy of the subject but a representation of the subject's ongoing search for the truth of the self. The image is authentic because *the self just is this search*." (Charles Guignon, *On Being Authentic* (New York, London: Routledge, 2004), p. 69, original emphasis.)

[7] The brief sketch of Rousseau's position is extracted from *The Social Contract* and the two *Discourses*, both in Jean-Jacques Rousseau, *The Social Contract, and Discourses*, trans. G. D. H. Cole (London: Dent, 1993).

theorists, to the degree that Nancy Fraser is justified in saying that "the usual approach to the politics of recognition – what I shall call the 'identity model' – starts from the Hegelian idea that identity is constructed dialogically, through a process of mutual recognition."[8] This demand for recognition is thus the demand for an appropriate societal attitude towards individuals' striving for authenticity, where it encompasses both a negative aspect – that one's quest be tolerated, meaning that he is not held in contempt or abused because of his pronouncements of identity; and usually also a positive aspect – that the value of the quest is given proper public acknowledgement (or that diversity is cherished, as the slogan goes).

2.2. Self-interpretation and irreducibly social goods

A further and distinctive part of Taylor's position is the special importance he gives to self-interpretation in the context of recognition.[9] It is at this point that the Hegelian strand of Taylor's argument is joined by the Aristotelian insistence on the communal nature of evaluative standards. This insistence can be seen as Taylor's rebuttal of the other aspect of the atomistic doctrine: what may be termed "justificatory (or value-) individualism". Whereas ontological individualism is an approach to the study of persons, justificatory individualism is an approach to the study of goods. Its central claim is that the property of goodness can be examined in relation to individuals alone. Against this claim, Taylor argues that the goodness of some goods – for instance language, and civic virtue – cannot be explained with reference to lone individual agents. This is the position of justificatory (or value-) *holism*. Taylor's idea can be approached by returning to the thought that in addition to recognizing ourselves as such-and-such persons, we also take an evaluative stance towards our being the way that

---

[8] Nancy Fraser, "Rethinking Recognition," *New Left Review*, no. 3 (2000): p. 109. For arguments that draw, to a lesser or greater extent, on this Hegelian idea see Axel Honneth, *The Struggle for Recognition* (Oxford: Polity Press, 1995).; Axel Honneth and Avishai Margalit, "Recognition," *Aristotelian Society, Supplementary Volumes*, no. 75 (2001).; and Avishai Margalit, *The Decent Society* (Cambridge, MA: Harvard University Press, 1996).

[9] Cf. Taylor, *Sources of the Self*, 46-47.

we are. One does not only care to know that the road she is following in her endeavours is one of a mother, or socialist; one also wants to know where she is standing in this course and how well she is doing in her strivings. In fact, one *must* ask oneself these strong evaluative questions, as Taylor calls them, because there is no way of attaching these labels to oneself without appealing to public standards of goodness. Asking about the good comes before saying anything about oneself, as its transcendental condition.

In explaining the inevitability of strong evaluative frameworks Taylor often invokes the spatial metaphor: when a lost traveller asks where is Mont Tremblant, it will not do for the native of the region to blindfold him and drop him off at the foot of the mountain the next day. The traveller will, to be sure, have found the spot and in one sense his curiosity has been satisfied; but, in another and more important respect, he is as lost as he was before. For, he has no concept of the geographical area and, despite his being brought to the destination, no understanding of where he is in relation to where he was before, or to other places in the known world.[10] But this is not the position one finds oneself in when trying to work out his identity. Self-interpretation is an own effort, consisting in placing oneself in some role or under some description *for oneself* – which means that one cannot rely on the benevolent native for guidance but is compelled to go by his own well-informed understanding of the map. The framework that fixates the ethical terrain and make finding one's way possible, continues Taylor, is the shared understanding of what is a good, or worthwhile, way of being.[11]

---

[10] Ibid., 41-42.

[11] Cf.: "In the light of our understanding of identity, the portrait of an agent free from all frameworks rather spells for us a person in the grip of an appalling identity crisis. Such a person wouldn't know where he stood on issues of fundamental importance, would have no orientation in these issues whatever, wouldn't be able to answer for himself on them. If one wants to add to the portrait by saying that the person doesn't suffer this absence of frameworks as a lack, isn't in other words in a crisis at all, then one rather has a picture of frightening dissociation. In practice, we should see such a person as deeply disturbed." (Ibid., p. 31.)

However, Taylor's communitarian understanding of goods goes further than the claim that any self-interpretation must make use of cultural categories and ideals. This claim is not very controversial or uncommon. In fact one of the major liberal currents, the Millian version of liberalism, explicitly acknowledges the deep entanglement of individuals' choices and the cultural context. One of central points of Appiah's approach to identity, as we saw in the previous chapter, is that the development of individuality requires "the complex interdependence of self-creation and sociability".[12] And Joseph Raz maintains that aside from independence and appropriate mental abilities, there is also a social condition of individuals' autonomy: an adequate range of options.[13] Autonomous choosing is choosing between valuable options. What differentiates Taylor from these socially sensitive liberals is that he regards some goods as unanalyzable to the interests, or well-being of individual agents. For Appiah and Raz, the social context provides indispensable resources for the development of individuals' character and their exercise of autonomy, respectively. Taylor goes beyond these claims by postulating a class of goods that are only meaningful in relation to a community that is composed of, but cannot be reduced to, individuals.

Taylor's irreducibility thesis should be clearly distinguished from the benign claim about *public goods* which present-day liberalism can easily wrestle with. One cannot enjoy fresh air without the same good being had by others. On the modern understanding of the state, where it is not seen as a decentralized cluster of corporations with private armies, there can be no system of national defence that benefits only some citizens and not all. Provision of the good of national security requires the pooling of considerable financial resources, which can in itself stimulate public debate on current norms of political inclusion and state's duties; it may be associated with intense feelings of pride and obligation to the national community; and in some extreme cases it engages individuals in undertaking sacrifices that would seem to outweigh the expected benefits, especially when their lives are endangered. But none of this affects the fact that

---

[12] Appiah, *The Ethics of Identity*, p. 17.

[13] Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon, 1986), pp. 372-73.

national defence, like fresh air, is open to a reduction. In the final analysis, they are both goods for *lone individuals*, so that "the good is a good only because it benefits individuals".[14] Each is intrinsically "decomposable", meaning that in order to gain insight into how and why they carry value it is sufficient to look into the separate satisfactions that compose them.

Public goods thus point to intersubjectivity only in the weak, non-constitutive sense. However, in addition, Taylor argues for another category of goods: those by their nature unsusceptible to any reduction. He points out that thinking of culture and language, in particular, as decomposable takes something away from their goodness: how individuals relate to their culture slips out of sight on an analytical account. So, one part of this objection to the atomist theory of the good is that the individuals' relationship towards their culture is best conceived as holistic. The actions we find ourselves performing and the social roles we find ourselves occupying are the actions and roles that are prescribed and delineated by the totality of social interchange. Describing certain behaviours as morally praiseworthy and others as reprehensible, working out and conceptualizing the ingredients of the good life, coming to the awareness of our unique traits and capacities – all these activities we inevitably conduct in the language of our cultural community. We are thoroughly implicated in the practices and shared understandings of our culture.

But this is not all that Taylor is saying. His claim is not only that culture is an irreducible feature of the society; it is that culture is an irreducible feature of the society *and* that it is a good.[15] Moreover, culture is a good *because* it is an irreducible feature of the society. He states that culture must be valued because it provides the indispensable background and structure to human action, where his favourite way of making the point is by likening culture to language. Adopting Ferdinand de Saussure's distinction, he speaks of the difference between *parole*, or speech act, and *langue*, or the totality of the linguistic system. It cannot be

---

[14] Charles Taylor, "Irreducibly Social Goods," in *Philosophical Arguments* (Cambridge, MA: Harvard University Press, 1995), p. 129.

[15] Ibid., p. 138.

denied that language perpetuates and recreates itself through individual speech acts, as external manifestations of that language, but the important thing to note is that *langue* cannot be conceived as the lump sum of all acts of *parole*. The reason for this is that linguistic utterances cannot be created *ex nihilo*: they must themselves be modelled after the pre-existing structure of meaning provided by *langue*. Now, it might be thought that this claim's reference to individuals ultimately subverts the whole irreducibility thesis. However, the fact that a linguistic code is analyzable into disparate words, propositions and speech acts, and that each speech act is attachable to an individual speaker does not in itself speak against Taylor. It does nothing to undermine the existence of an irreducible backdrop which only gives *parole* meaning. Individual utterances are made possible by this abstract structure, which continues to hold even if language must necessarily be affirmed in concrete situations, by concrete people. The same dialogical relationship, argues Taylor, exists between culture and the individual. Hence, since culture is as fundamental to action as *langue* is to speech, we must value culture if we value agency.

With these claims in place, the depth of Taylor's disagreement with the analytical approach to identity can be more readily appreciated. Since he endorses the Hegelian idea of the emerging self, he regards recognition in the public domain as essential to the development of selfhood. This is a remarkably strong defence of the value of unhindered participation in the public life of a community. It is focused on a value that precedes and transcends the ideals and principles that liberal theorists usually invoke, such as liberty, or welfare, or human rights – the value of selfhood itself. However, Taylor does not regard selfhood in static terms, as something that can be had or lost once and for all. Rather, to be a person is to take part in the process of becoming a determinate kind of human being in dialogue with others. Furthermore, for Taylor this dialogical self-interpretation must be framed in terms, and draw on ideals, inherited from one's culture. What is distinctive about Taylor's position, however, is that self-interpretation must be conducted in view of a special class of goods, ones that cannot be reduced to a sum of goods for disparate individuals. Points of reference for individuals' self-identification are objects and virtues that are good for the community.

At the practical level, Taylor's claims about identity have the potential of representing one of the strongest cases conceivable for the autonomy of cultural groups. The core of such a strongly communitarian defence would be the startlingly forceful claim that nothing less than moral subjectivity is at stake in debates about identity. This is because the emergence of integral selfhood heavily depends on full and unobstructed participation in the community's cultural interchange. Furthermore, the strong argument for cultural autonomy would deny that this can be achieved through partial accommodation of cultural claims, for the essential feature of a cultural community is that it expresses itself in its political constitution. It collectively probes and answers the question about the human good and then builds institutions around it, where the important thing is that it does the searching and the regulating *on its own*. In this regard, Taylor may be seen as recommending a return to some insights of the civic-humanistic tradition of political theorizing.[16] This tradition, which among others includes ancient writers, Machiavelli, Montesquieu, and Tocqueville, places great emphasis on the civic virtue of patriotism. It is a virtue which cannot be explained by appealing to the separate virtues of individuals because it is not a property of any of the parts of the polity but of the whole. It embodies a common spirit of devotion to the public good, stemming from the shared sense that participation in the political body forms part of the good citizen's dignity. Politics is, on this conception, not an instrument in the service of individual interests, but a fate that we are sharing, an enterprise that is truly *ours* and not mine, yours and the third person's. As with Taylor's irreducibly social goods, the meaning of patriotism remains elusive if one stubbornly persists with the idea of patriotic citizens where talking about common patriotic virtue is more appropriate.[17]

---

[16] The connection between Taylor's normative views and the republican tradition is drawn especially in Taylor, "Cross-Purposes: The Liberal-Communitarian Debate."

[17] Michael Sandel offers an account of the self that has similarly republican inclinations. (Michael J. Sandel, *Liberalism and the Limits of Justice*, 2nd ed. (Cambridge: Cambridge University Press, 1996), pp. 79-103.) After considering Nozick's objection to Rawls' difference principle, Sandel argues that Rawls is only able to treat natural talents as a common societal asset at the expense of violating his own principle of distinctness of persons. More specifically, what precludes Rawls

In sum, Taylor's arguments undermine individualist analyses of identity by making a case for the inclusion of reference to essentially communal goods in the concept of identity. Within this perspective, the central determinant of an individual's identity is not any decision or property of that individual, but the dialogical community of which he is part. As Taylor puts it, "To understand our predicament in terms of finding or losing orientation in moral space is to take the *space* which our frameworks seek to define as ontologically basic."[18]

2.3. Human agency and culture

I have already noted that it is possible to understand Taylor's position on identity, and the critique of individualism that it involves, as an amalgam of two theses. One is the Hegelian thesis about the dynamic nature of selfhood; the other is the Aristotelian thesis about the orientation of practical reasoning towards the communal good. The notion that serves to connect these two disparate claims is that of self-interpretation. For Taylor, individuals constitute themselves through the process of self-interpretation, which stands in need of validation by others. Self-interpretation is in turn always evaluative: we do not merely describe ourselves as a certain kind of person, but ask whether we are doing well in being that kind of person. One wonders whether she is a good mother, a good communist, and so on. And it answering such questions one must appeal to an essentially communal notion of what constitutes a good life.

---

from taking that route is that individuation of persons is taken as *a priori* given, and includes the assumption of mutual disinterest. Sandel's solution (in which he of course differs from Nozick) is to include such attributes as attachment to others in the definition of selfhood, as constitutive rather than merely accidental properties of the self. This Sandel's positive conception of intersubjectivity (for which he also finds support in Rawls' idea of the social union) is sometimes overlooked in favour of his more widely cited negative critique of the "unencumbered self". However, the republican disposition also carries over into his subsequent writings, with a more prominent role to play.

[18] Taylor, *Sources of the Self*, 29, emphasis added.

A full assessment of Taylor's critique of reductionism about identity would require an appraisal of both theses that compose it. I will however focus on only one of them, the Aristotelian claim about irreducibly social goods. This would be impossible if the two strands of his argument were inseparable, but this is not the case. One can be a Hegelian, of a more orthodox type than Taylor, about recognition without thinking that recognition is primarily a matter of self-interpretation. Equally, it is possible to espouse a more conventional Aristotelian position that affirms the orientation of practical reasoning towards communal good, but takes selfhood as fixed. The merger of these two lines of argument is a highly original and idiosyncratic feature of Taylor's philosophy. That said, it is still necessary to justify the decision to omit further consideration of the Hegelian element of Taylor's position. Given that the Hegelian line of argument advocates the ontology of fluid subjectivity, it is difficult to both present and assess in an entirely perspicuous fashion. Since it breaks down the conceptual separation of persons from their engagement with the world – the subject-object relation, as it is sometimes called – it cannot be done justice without abandoning the analytical style of philosophizing. Since liberal political theory is mostly, if not wholly, done in the analytical mode, it would need to be stretched considerably to properly confront the idea of the emerging self.[19] Of course, this is no argument against doing so. However this meta-theoretical enquiry would not only require much more space than I can afford here, it would also distract from my main topic of practical identity. Therefore, I cannot further discuss the first part of Taylor's critique.[20]

---

[19] The need of the analytical method for fixed units of analysis is expressed well in Kukathas's critique of Iris Marion Young's Hegelian account of the self: "If we are to theorize about the good for human beings (or about anything, for that matter), something must be kept constant. The suggestion in this work is that that constant is the individual, as the entity with whose good we must ultimately be concerned." (Chandran Kukathas, The Liberal Archipelago (Oxford: Oxford University Press, 2003), p. 90.)

[20] This restriction of my discussion should not be understood as an implicit concession to Taylor: it does not rule out the possibility that the idea of the emerging self is implausible after all. Hegel's idea rests on a critique of the formality of Kant's philosophy – specifically, of Kant's ambition in the Critique of Pure Reason to uncover the necessary conditions of cognition itself, abstracted from any content of cognition. Hegel's charge is that this ambition, and the notion of a

Instead, I wish to focus on the other part of Taylor's rebuttal of atomism about identity, his view of irreducibly social goods as the basis for evaluative self-interpretation. My suggestion is that this view should be rejected because the concept of irreducibly social goods is unacceptable. It is unacceptable because its treatment of the conditions of human agency is arbitrary and hence illegitimate. My objection is thus different from the criticism that irreducibly social goods are not really *irreducibly social*. An exponent of this latter charge is Chandran Kukathas. He argues that "the condition of human beings is, ultimately, one of solitariness; and the goods each knows he knows not in common with others but alone. The fact that human beings are social beings does nothing to alter this."[21] What leads him to this conclusion is the difficulty in explaining the supposed intrinsicness of communal goods. The only way in which the goodness of culture and language can be understood, writes Kukathas, is if they are ultimately good for individuals. Since all goods are *consumed* individually rather than collectively, they can only be goods for individuals.[22]

It is doubtful whether Kukathas's critique succeeds. It is rooted in the uncontentious claim that "For something to be valuable it must, at some point, be valuable to *someone* for the value it gives *someone*."[23] From that claim Kukathas

---

critique of *pure* reason, is incoherent. One cannot, he claims, investigate the necessary structure of cognition without a basis in some prior cognition. To claim otherwise is to make the mistake of "refusing to enter the water until you have learnt to swim". (Georg Wilhelm Friedrich Hegel, *Hegel's Logic: Being Part One of the Encyclopaedia of the Philosophical Sciences*, trans. William Wallace, 3rd ed. (Oxford: Oxford University Press, 1975), p. 66.) Yet, this objection is not obviously well-placed. For instance, Karl Ameriks has claimed that Hegel wrongly assumes that Kant aims to deliver a general criterion for knowledge. Instead, Ameriks contends, Kant starts by assuming that there are some warranted knowledge-claims, and then looks to answer which a priori principles would be required by them. (Karl Ameriks, "Hegel's Critique of Kant's Theoretical Philosophy," *Philosophy and Phenomenological Research* 46, no. 1 (1985).) On this construal, the criticism that forms the basis of Hegel's dynamic concept of the self is misdirected. Although there is much to be said for this claim, I lack the space to do so here.

[21] Kukathas, *The Liberal Archipelago*, p. 67.
[22] Ibid., p. 69.
[23] Ibid., p. 68, original emphases.

moves on to a further claim: there cannot be intrinsically group goods because it is unclear how anything can be good for a group, as something that exists above and its members. Ultimately, only individuals have interests and concerns in virtue of which something can be good for them. Since groups are composites of individuals, with no interests and concerns that are independent of the interests and concerns of their constituent members, all group goods are in fact reducible to goods for group members. However this objection misses its target because, as I noted earlier, Taylor is not a collectivist. In his opposition to justificatory individualism he does not contest the claim that goods can only be for individuals. His claim, which he calls value holist, is that there are some goods whose value for individuals cannot be understood if these individuals are studied in isolation from others. Aside from things that are valuable for me *and* you, suggests Taylor, there are also things that are valuable for *us*.[24] We are ontologically separate but can experience and enjoy certain goods only together, as a product of our cultural interchange.

Taylor's claims can be further explained by noting that he thinks of culture and language as background conditions of human agency. Their value consists in the fact that they are indispensable for making any kind of choices. His reasoning is that if agency is taken to be a good, then that which is its necessary prerequisite must also be considered valuable. It should be kept in mind, however, that Taylor's argument does not pick out culture and language as the *causes* of agency, but their *conditions*.[25] So, Taylor is not committed to the claim that shared norms and understandings are to be valued because they bring human action to existence. If this straightforward relation of determination existed, culture would destroy rather than promote agency. Instead, Taylor maintains that the level to which culture meshes with individual choices allows one to think that

---

[24] Taylor, "Cross-Purposes: The Liberal-Communitarian Debate," p. 189.

[25] He zeroes in on "the way in which thoughts presuppose and require a background of meanings to be the particular kind of thoughts they are. But the term "presuppose" and "require" in the previous sentence point to a peculiarly strong relation. It is not a contingently causal one, which we could imagine a way around − the kind we invoke when we say that neolithic villages couldn't have built pyramids because this requires and presupposes a larger labour force." (Taylor, "Irreducibly Social Goods," pp. 131-32.)

action could not take place in the absence of these norms and understandings. Despite the lack of causal determination, it can be said that they carve out a niche in which being an agent presents itself as a possibility. Without them, thinking and acting would be inconceivable. The rationale then is clear and compelling for prizing culture as the enabling condition for engaging in any sort of activity.

However Taylor's defence of the value of culture as a background condition of human agency is, in the end, unpersuasive. It overlooks the fact that even if culture is a condition of human agency, it cannot be regarded as the *ultimate* condition of agency that is not itself conditioned. Taylor's decision to arrest the enquiry into the conditions of agency at the level of culture is arbitrary and philosophically unacceptable. To explain: throughout his discussion, Taylor assumes that the flip-side of rejecting justificatory individualism is accepting the status of culture as a good. The negative claim of value holism is its rejection of the study of all goods in relation to lone individuals. Its positive claim, Taylor suggests, is that culture is a good and as such matters in political life. It is this assumption that underlies his practical normative prescriptions – for example, his claim that a full appreciation of holism entails granting some form of autonomy to territorially demarcated and institutionalized cultural corpuses, such as Quebec. Only if cultural autonomy is granted to them, suggests Taylor, can the Quebecois be reunited with the cultural core that informs, and is in turn informed by, their moral reasoning. He writes: "Where the nature of the good requires that it be sought in common, this is the reason for its being a matter of public policy", the policy in question being that of a distinct cultural community.[26]

But it is implausible to suppose that the rejection of justificatory individualism entails accepting the status of culture as a good. What discredits this supposition is the possibility that even if culture is the background condition of agency, it is itself conditioned. In other words, the possibility is that there are *pre*-cultural, or "brute", facts and conditions that make the cultural interchange possible. If this is the case, rejecting justificatory individualism does not entail accepting the normative significance of culture. Instead, it warrants accepting the normative

---

[26] Taylor, "The Politics of Recognition," p. 59.

significance of brute facts, which are the more fundamental condition of human agency.

The objection I am advancing would be unavailable if the following argument succeeds: the fact that all our concepts are mediated by culture means that there can be no meaningful talk of anything outside cultural understandings. Culture accounts for all we know and can represent, which is why there are no *other things* beside culture that constrain our behaviour. However this argument is deeply problematic. In a passage representative of the view I wish to counter Bhikhu Parekh writes:

> Even something as basic and inevitable as death is viewed and experienced differently in different cultures. In some it is a brute fact of life, like the falling of leaves or the diurnal setting of the sun, and arouses no strong emotions; in some others it is a release from the world of sorrow and embraced with joy; in yet others it is a symbol of human weakness, a constant reminder of inadequate human mastery over nature, and accepted with such varied emotions as regret, puzzle, incomprehension and bitterness.[27]

Parekh's suggestion is that facts or events in the physical world, even those as drastic as death, do not affect us *in themselves*; what brings them to our attention and renders them an occasion for feelings of stoic acceptance, or joy, or puzzlement and anguish is the cultural interpretation. Yet, this claim contains an exaggeration: it is certainly true that our reception of physical facts is not immediate, but holding that they therefore do not impinge upon our concepts is simply far-fetched. They do because they constrain the scope of cultural understandings. In other words, they are what the interpretation is *about*. Thus, at the most general level, there could be no concept of death as release from worldly suffering if there were not for the underscoring brute fact of biological termination of life. But moreover, it could also be argued that this specific way of

---

[27] Bhikhu Parekh, *Rethinking Multiculturalism* (Basingstoke: Macmillan, 2000), p. 121.

49

dealing with death finds its grounds in certain objective features — of the event itself, which often comes after a period of painful physiological deterioration; or of human life, which presents most people with innumerable adversities and frustrations; perhaps also of the attestations of the dying, who sometimes express profound placidity in the face of imminent perishment. This is why we can, to some degree, understand how one can think of death as merciful release.[28]

Now, there is no reason to deny that there may be great scope for interpretation in these matters. The underlying brute fact may be multifaceted (as most would seem) — that is, such that we might regard plausible construing it in a number of ways: as salvation, ultimate defeat, or an ordinary part of what it means to be human. But, again, there is something about death that accounts for these diverse imaginable attitudes. The crucial point then is that culture is not the terminus of conceptual explanation, and as such cannot be presented as the ultimate arena for holist interchange. Factors standing beyond it ensure that, as far as interpretations are concerned, not anything goes. Death cannot be thought of as "blue", or "salty", or "modest", whether one is Amish, Burmese or Finnish. Seen in this light, the fact that the physical world only affects our concepts indirectly ceases to be as pivotal as Taylor and Parekh maintain. The important thing is that it does nonetheless.

The purpose of discussing beliefs and attitudes surrounding death has been to point out that they cannot be merely the product of what Taylor understands by culture — of the institutionalized corpus of norms and understandings that is distinct from other such corpuses and attaches to some community of people. These norms and understandings draw on an order of things that is external to them, and that constrains them. One can then say, to adopt Taylor's terminology, that this external and pre-cultural order is a background condition of the cultural fabric, just as the latter is the background condition of human agency. If this is

---

[28] Thomas Nagel's *The Last Word* represents a careful and persuasive defence of the same point, considered on a wider scale — that at the bottom of every justification, be it in the area of religion, ethics, or epistemology, there must be objective principles or facts which do not depend on our point of view. (Thomas Nagel, *The Last Word* (Oxford: Oxford University Press, 2003).)

so, it makes sense to criticize Taylor on the account that it is not included in the Saussurian circle that his multicultural recommendations are meant to sustain. So, when he argues that cultural communities are part of the interactive relationship that makes agency possible, it is legitimate to remark that other things do too. Simply arresting the enquiry into the conditions of agency at the level of culture, as Taylor does, is arbitrary and illegitimate.

Moreover, Taylor's arbitrary treatment of the conditions of agency makes him vulnerable to the charge of parochialism. The charge is that Taylor's advocacy of the normative significance of culture is just an expression of his partiality to the political programme of securing the political autonomy of particular cultural groups. It may just be down to the fact that Taylor thinks that some specific cultures (including, presumably, the Quebecois) need political protection. This criticism can be explained by recalling that Taylor presents culture as indispensable for any kind of agency. However, closer inspection reveals that he does not actually make this point. His argument is not that shorn of the cultural backdrop we cannot have the concept of action in general – but that without the backdrop there can be no concept of an action that has some *determinate* character and falls under some *specific* linguistic description. His modest claim is that thoughts rely for being "the particular kind of thoughts they are" on the cultural framework. He writes:

> Nothing could count as making the claim "she's sophisticated" among neolithic farmers in upper Syria (if our surmises are right about their culture), in somewhat the same way as nothing could count as making the queen's gambit in a checkers game. The move presupposes a background of rules or, in the case of language, conditions of possible validity; and in both these cases the background is missing.[29]

This line of reasoning is bound to cause some worry about the scope of its conclusions. The proposition that the queen's gambit is only possible within a

---

[29] Taylor, "Irreducibly Social Goods," p. 132.

game the rules of which stipulate such movement of the piece is not informative. It is perfectly expected and constitutive of how we understand things that one cannot make that same move in checkers, or in water polo, cooking, upholstery, climbing, or tooth brushing. Chess moves only take place in chess, just as French can be spoken in Quebec only if there exists a francophone culture to spawn it. In other words, it does not seem a very deep point to say, as Taylor does, that *culture-specific* agency presupposes the cultural background. Furthermore, it is difficult to see how this claim can yield any universal conclusions. Since it does not concern agency in general, it can only throw up the following statement: if one values some particular type of human expression, one must also value the particular cultural corpus that fosters it. This is a conceptual truth not unlike saying that if one takes pleasure in a certain type of chess opening then one must also care for the system of rules that enables it. To put it slightly differently, prizing Quebecois cultural products – such as Quebecois nationalism, francophony and the tradition of Roman Catholicism – inevitably entails prizing the distinctness of Quebecois culture. However, these words are plainly addressed to those already converted. It takes a prior conviction that the Quebecois way of life has worth in all its specificity to agree that cultural autonomy is merited. Because it never reaches so deep as to ground *this* conviction and consequently has only conditional force, Taylor's transcendental argument can be called parochial.

It is worth noting that Taylor can produce an elegant response to the parochiality objection that keeps in line with his Hegelian leanings. He might say that all concepts are necessarily "parochial", in the sense that they are contaminated with lived practice and the beliefs stemming from it. As was remarked in relation to the Hegelian critique of Kant's formalism, any self-knowledge is for him self-description, and any moral ideal is the ideal of some community. Therefore, it is no criticism to point out that a precept derives its plausibility from appealing to my particular situation – one construed not in connection with some void category of humanity, but the totality of my presence in the world. There is simply no other way, Taylor will reply, a precept could exert force on me. I do not wish to weigh the merits of the rejoinder, but merely to indicate that the challenge compels Taylor to employ the already discussed notion of "background

conditions of human agency". His claim is that I, as an active agent, am at one with the complex ongoing interchange that provides me with the means and the capacities for impressing myself on the world. Isolating the pure centre of consciousness from the full set of factors contributing to the interchange is illegitimate, as it fails to explain the nature of self-referencing and the value we attach to being ourselves. However if Taylor does wish to resort to this line of defence against the parochiality objection, it is vital that he provides an appropriate depiction of the Saussurian circle. This, as I suggested, is where his argument for *culture-enclosed* exchange misses the target.

It will be useful to quickly retrace my steps in this chapter. My aim has been to defend the viability of enquiring into the identities of individuals against Taylor. He claims, first, that the concept of identity must be understood in relation to the totality of exchanges within a community. Second, that totality cannot be reduced to the sum of individuals' contributions and properties, in the same way that social goods cannot be regarded as aggregates of what is good for lone individuals. Therefore, Taylor maintains, a strictly individualist concept of identity is fundamentally ill-conceived. I have not tried to contest this Taylor's argument directly. Instead, I have merely maintained that the scope of that argument is arbitrarily arrested. Taylor assumes that culture is the ultimate "undecomposable" factor that influences the choices of individuals, but it is difficult to see why this assumption should be accepted. If one thinks that individuals' agency is not self-sufficient but depends on wider background conditions, there is no reason to simply equate those conditions with culture. Culture, I have suggested, is also conditioned – by an underlying order of brute, or natural, facts. Therefore culture cannot be thought of as the highest, or the only, precondition of agency, or valued as such.

However my conclusions so far are insufficient to defend the possibility of individualism about identity. My discussion has only highlighted the inadequacy of one, albeit very influential, version of justificatory holism – Taylor's strongly "cultural" version. But it has not challenged the overall holist outlook. This concession, it seems, commits me to accepting that consistent and thoroughgoing holism, if it can be developed, is the correct approach to identity. This would be

the approach that takes full account of the embeddedness of individuals' agency in a variety of natural and non-natural conditions. Here I want to distance myself more decisively from the holist perspective. My claim is that thorough holism is, in fact, not a viable option in the study of identity, which indirectly vindicates the analytical approach. There are two considerations to support this view, both of which can be explained by speculating on why Taylor himself steers clear from full-blown holism. It is plausible to assume that at least a part of his reason for doing so is practical: namely, that a wider concept of the background conditions of human agency precludes any cultural politics. Culture can be fitted into a system of normative claims, as is shown by Taylor's activist advocacy of autonomous Quebec; it remains unclear however whether such arguments can find room for the foundational order of brute facts. Consistent holism is just too impractical to stand as a political credo. Putting the point more generally: it is difficult to see how any kind of normative claims can be derived from a perspective that regards as valuable *all* of the various preconditions for choice.

Taylor's more philosophical rationale could be turning on a problem with justification in pluralistic societies. Aristotle's moral philosophy can be called, uncontroversially, the archetypal example of deep holism, insofar as it incorporates the category of brute, culture-independent facts that determine cultural understandings. The treatment of the gender distinction in *Politics* is a case in point. Aristotle maintains that "between male and female the former is *by nature* superior and ruler, the latter inferior and subject".[30] What deserves attention here is not Aristotle's chauvinism, but that he neither aims for, nor attempts, a *justification* of the natural superiority thesis. The remark, as well as the more comprehensive discussion in *De Generatione Animalium*, merely explains or elucidates the thesis to an audience that already lives the truth of it in everyday moral practice.[31] Aristotle can start from a shared understanding which

---

[30] Aristotle, *Politics*, trans. Stephen Everson (Cambridge: Cambridge University Press, 1996), 1254b2, p. 68, emphasis added.

[31] This clarificatory aspect of Aristotle's approach comes out best in the famous passage from the *Nicomachean Ethics*: "Presumably we have to begin from things known by us. This is why anyone who is going to be a competent student in the spheres of what is noble and what is just –

he does not need to vindicate – the assessment that men are better placed to fulfil their natural rational function than women. But Taylor can enjoy no such privilege. The basis of shared understandings is precisely what is lacking in culturally divided societies, especially relating to such deep issues as the meaning of brute inequalities, natural predilections of human beings, matters of religion, and so on. Under radical pluralism, the thoroughgoing holist would have to *persuade* those coming from different systems of meaning, while holding onto some fixed yet uncontroversial idea of how nature determines moral practice. This task seems altogether vexed.

In sum, not only is comprehensive holism unable to deliver normative political claims, it also requires a wide agreement on what qualifies as a natural precondition for choice, which is absent in contemporary societies. Although this finding does not amount to a direct defence of the analytical approach to identity, it shows that a prominent objection to it is inconclusive. Individualism about identity, which is the approach I will favour, is hence still a possibility. I first turn to a line of argument that links identities with reasons via the notion of self-respect.

---

in a word, politics – must be brought up well in his habits. For the first principle is the belief *that* something is the case, and if this is sufficiently clear, he will not need the reason *why* as well." Aristotle, *Nicomachean Ethics*, trans. Roger Crisp (Cambridge: Cambridge University Press, 2000), 1095b4-8, p. 6, original emphasis.)

3. Identity and Rawlsian self-respect

3.1. Identities and impartiality

The typical contemporary state comprises diverse cultural groups that are, nonetheless, compelled to co-operate and share the burdens and benefits of common political life. At least some of these groups uphold unique group-specific ideals and practices. Take, for instance, the case of religious rituals. Observing Shabbat by praying and abstaining from the thirty-nine activities that were required for the building of the Tabernacle is specific to the conservative and orthodox Jewish communities. This custom is not shared by other religious denominations or non-believers simply because they are not devout Jews. Another example is provided by minority linguistic groups, whose linguistic practices set them apart from the society at large.

Brian Barry has forcefully argued that the cultural diversity of contemporary states invalidates demands for special cultural rights. The correct normative response to diversity is to withdraw cultural matters from politics, rather than politicize them. Anything else, Barry argues, would run counter to fairness and justice. His remarks about the fair treatment of religion are illustrative of his broader position:

> What can be said about the liberal proposal for privatizing
> religion, then, is that it is the only way in which religions can be
> given equal treatment, and equal treatment is what in this context
> is fair. This contention is, of course, open to dispute. But it cannot
> be proved wrong merely by observing that the kind of settlement
> it recommends will be inimical to the beliefs of some people. A
> fair distribution of property will be inconvenient to those who

56

have an unfairly large amount. Similarly, a fair way of dealing
with religions will incommode those who wish to make claims on
behalf of their own religion that cannot be accommodated within
the constraints prescribed by fairness. There is nothing surprising
in this.[1]

Some discomfort and inconvenience will certainly be inflicted upon orthodox
Sikhs by a policy that disallows wearing a small dagger (*kirpan*) in public.
However, argues Barry, the loss to their psychological well-being will be
outweighed by the public gain in justice. This claim needs further clarification as
it can easily be misunderstood. It may be thought that Barry's idea is just that,
given the cultural diversity of contemporary societies, cultural rights lead to
unequal treatment of citizens. Under conditions of diversity different individuals
have different identities standing to be expressed, and require different liberties
and amounts of resources to attend to them. For instance, Millian individuals that
are highly mobile and experimental with regard to their identities will have
different demands of the state than those persons that strongly identify with their
cultural background and local community. Those unconcerned about their
identities, preferring to focus on advancing their material comfort or on fulfilling
their family duties, will demand to be treated in yet another way. Barry might be
taken to argue that the described requirement for differential treatment of citizens
is incompatible with justice. The granting of special liberties or resources to
some individuals with regard to some matter disadvantages other individuals in
that same matter. The right of publicly wearing daggers stipulates an exception to
the system of uniform rights, as well as placing those enjoying it in a position of
possible physical dominance over other citizens. Therefore, inconvenience
through non-recognition is not only a perfectly expected outcome of neutral
arrangements; it is also insufficient to warrant unjustly privileging some citizens
over others.

However, while it is correct to say that equality lies at the root of Barry's
objection to cultural rights, his argument is not best understood as directed

---

[1] Brian Barry, *Culture and Equality* (Cambridge, MA: Harvard University Press, 2001), p. 28.

against inequality of treatment. What is wrong with special identity-rights is not that they give some citizens certain privileges that others do not have. Barry is thus prepared to accept that citizens' equality is not synonymous with their possession of the same rights and obligations, and that it may sometimes even demand treating them *un*equally. Ronald Dworkin expresses this point forcefully by saying that the principle of treating persons equally is subsidiary to the principle of treating persons *as equals*.[2] In an example he gives, providing the same amount of aid to flood-stricken areas that are equally populous but have suffered different levels of devastation would be unfair. The appropriate response would be to differentiate the relief in accordance with the gravity of damage.

Barry's position can accommodate Dworkin's point. His objection is not that identity-politics would result in inequality of rights; instead, it concerns the kind of justification that multiculturalists can provide for unequal cultural rights. His argument states that sectional privileges cannot be impartially justified – that is, supported by arguments that would be acceptable to all citizens. The demand for impartial justifiability is, according to Barry, rooted in the classical liberal striving to protect the individual from the abuse of political power.[3] What makes this concern relevant in the context of identity politics is that claims to state support of essentially sectional activities cannot be expected to be welcomed by other communities. That some practice is "a part of my culture", writes Barry, cannot be enough to convince those from *other* cultures that the practice should be enforced by the state. If the state does enforce such a practice, it will do so without the consent of some citizens – those who have no interest in preserving the practice, or who might be inconvenienced by it. For Barry, this means that these citizens will be coerced. With regard to the right of publicly wearing daggers, non-Sikhs may object that it places Sikhs in a position of possible physical dominance over other citizens. Granting and enforcing this right would

---

[2] Ronald Dworkin, "Liberalism," in *A Matter of Principle* (Oxford: Clarendon Press, 1986), pp. 190-91.

[3] As he notes, "liberalism is, both historically and logically, the result of generalizing the proposition that it is no business of the state to enforce the observance of the true religion – however and by whomever that is defined." (Barry, *Culture and Equality*, p. 65.)

then limit the lives of non-Sikhs without their consent. In effect, it would coerce them into accepting a constraint on their choices and actions. And this, argues Barry, is incompatible with the liberal principle of ensuring that individuals have maximum freedom to design their own lives.[4]

Faced with Barry's forceful objections, multiculturalists have to provide an impartial defence of special cultural rights. They need an argument for recognition of particular identities that persons of other identities, as well as those unconcerned about their identity, cannot reasonably contest. Parekh offers one such argument that draws on Dworkin's idea that treating persons as equals sometimes requires treating them unequally. He argues that only a special right to wear weapons in public equalizes the position of observant Sikhs with that of the rest of society.[5] This argument is parallel to the one Dworkin uses in his flood example, with culturally fostered disadvantages taking the place of naturally caused ones: since Sikhs have to bear larger costs for preserving their culture in Western societies than non-Sikhs, equality requires that they be compensated in the form of special cultural rights. However, Parekh's equality-based argument is contentious and at odds with the liberal commitment to individualism. His account insists more on equality between groups than on interpersonal equality. Parekh's reasoning rests on a strongly contextualized understanding of opportunity, according to which a person's opportunities are relative to his cultural background.[6] It is for this reason that cultures, as authors of meaning and value, deserve equal standing in intercultural dialogue that is based on shared operative values.[7]

---

[4] Barry thus writes that "the complaint made by liberals is not against the objective of remaining true to some ancestral culture but against the coercion of those who do not share that objective." (Ibid., p. 66.) And further: "Liberals must stand up for the rights of those who wish to pursue individual goals of self-development." Barry, *Culture and Equality*, p. 66.)

[5] Bhikhu Parekh, "Equality in a Multiracial Society," in *Equality*, ed. Jane Franklin (London: Institute for Public Policy Research, 1997), p. 135.

[6] Ibid., pp. 150-51.

[7] Parekh, *Rethinking Multiculturalism*, pp. 264-94.

A more plausible impartial argument for special cultural rights must avoid Parekh's suspect relativization of agency. Barry maintains that multiculturalists typically do advance an argument of this sort, and discusses a form of it. The reasoning he considers is that non-recognizing identities, especially cultural identities, is unfair because persons cannot dissociate themselves from them. Insofar as persons are culturally embedded, the claim goes, there are certain beliefs and practices that they cannot do without. The problem with this thesis, as Barry readily points out, is that it is plainly implausible. There are simply no practices and beliefs to which individuals would be so organically attached that they could not step back from them. Culture-related beliefs and practices are no exception to this, which is to say that their being "a part of my culture" does not make them special in any way. A part of what it is to belong to a culture is that one is disinclined to pursue activities that the culture forbids – for instance, when a devout Muslim rejects meat from inappropriately slaughtered animals. Sometimes this disinclination is so strong that it blocks the person's ability to see any value in culturally shunned practices: the Muslim might not even be tempted to eat meat from inappropriately slaughtered animals. But none of this means that it is, strictly speaking, *impossible* for him to have the meat. In order to prove this the multiculturalist would have to liken the said constraints to the incapacitation of physically disabled persons, which the latter would rightly find offensive.[8]

Barry's rebuttal seems correct on the face of it: if the claim about the impossibility of dissociation carries the entire weight of justification, the multiculturalist argument must collapse. But it is also true that another, and more promising, impartial argument for recognition escapes Barry's attention.[9] The argument he considers states that persons necessarily have certain *ends*, principally to preserve their cultural identity. A more attractive argument, which I will examine in the rest of this chapter, is framed in terms of individuals' *reasons*. Some prominent multiculturalists endorse Rawls' justification of the worth of self-respect, adding the further thesis that secure identification is

---

[8] Barry, *Culture and Equality*, 37.

[9] Although I cannot say more on this point here, it is also uncertain whether anyone really subscribes to the rather simple-minded 'necessary tie' argument that Barry discusses.

indispensable for a stable sense of self-respect. If this thesis is correct, all persons have a reason, related to the idea of self-respect, to want to have their identities publicly affirmed and promoted. In what follows, I will first present some formulations of the Rawls-inspired argument that links identities and reasons. I will then discuss the psychologistic conception of self-respect that is required for recognition to be presented as one of Rawls' "social bases of self-respect". My final assessment will focus on the compatibility between Rawls' own and multiculturalist understandings of recognition.

3.2. Self-respect and its bases

Rawls' claim that attracts the interest of multiculturalists is that the highest function of a conception of justice is to protect citizens' self-respect. Since parties behind the "veil of ignorance" do not know the particulars of their situation – their natural characteristics, or the details of their life-plan – Rawls maintains they will choose conservatively. Uncertain whether they will turn out to be black or white, they will wish to safeguard fair equality of opportunity; not knowing whether they will be politically active or complacent, they will opt for maximal liberty of expression, compatible with equal liberty for others; unsure of their gender, they will shelter gender equality, and so on. These *primary goods* would be rationally desired by individuals, argues Rawls, whatever their life-plan, because they are prerequisite to the success of any life-plan.

What makes Rawls so appealing is his contention that one of the pronouncements of justice as fairness is the uncompromising worth of self-respect:

> When we feel that our plans are of little value, we cannot pursue
> them with pleasure or take delight in their execution. Nor plagued
> by failure and self-doubt can we continue in our endeavours. It is
> clear then why self-respect is a primary good. Without it nothing
> may seem worth doing, or if some things have value for us, we
> lack the will to strive for them. All desire and activity becomes
> empty and vain, and we sink into apathy and cynicism. Therefore

the parties in the original position would wish to avoid at almost any cost the social conditions that undermine self-respect.[10]

Nir Eyal advances a potentially dangerous objection to Rawls' treatment of self-respect, claiming that it confuses self-respect with "confidence in one's determinate plans and capacities".[11] According to Eyal, the conclusions of Rawls' argument are different to those he intended. Rawls believes, on Eyal's interpretation, that parties in the original position would choose self-respect "in the Kantian sense" – belief in their dignity as ends in themselves – as a primary good. However Rawls shows at most that they would wish to enhance the chances that, once the veil of ignorance is lifted, their projects will bear fruit.[12] In Eyal's example, a scientist will not undertake research on HIV unless he trusts the value of his research plan and his own capabilities. This is a good incentive for him to place premium value on the state securing optimal conditions for the development of his research potential and resources. All rational persons, it might be said, would similarly wish for guarantees that they will be well-placed to materialize their life-plans. But this rational desire to preserve confidence in one's determinate plans and capabilities is not, argues Eyal, the kind of self-respect that Rawls has in mind. To understand this point, it is enough to consider that preserving one's confidence in one's determinate plans and capabilities may require a far greater threshold of guaranteed income and wealth than Rawls envisages. The self-confidence of the HIV researcher may, for instance, require substantial financial investment into expensive equipment, supporting medical and administrative staff, and so on. Other individuals will have similar, or even higher, demands related to their life-plans, in which situation the state may insist,

---

[10] John Rawls, *A Theory of Justice*, Original ed. (Cambridge, MA: Belknap Press, 2005), p. 440.

[11] Nir Eyal, "'Perhaps the Most Important Primary Good': Self-Respect and Rawls's Principles of Justice," *Politics, Philosophy & Economics* 4, no. 2 (2005): p. 202.

[12] A similar critique is also advanced in David Johnston, *The Idea of a Liberal Theory: A Critique and Reconstruction* (Princeton: Princeton University Press, 1994), pp. 137-85. In broad agreement with multiculturalist reasoning, Johnston maintains that among the social bases of self-respect are status and recognition, which he illustrates with the example of the Black community in the USA.

for practical purposes, on perfect equality in income and wealth.[13] This is hardly the planned outcome of Rawls' argument about self-respect.

One way of combating Eyal's objection would be to highlight the conditions of radical ignorance that obtain in the original position. The defence might be that the self-respect of Rawls' deliberators is independent of their success in practical pursuits because they do not now what these pursuits happen to be. They wish to preserve their belief in the value of their own agency absent any knowledge of what goods they actually strive for. Under this radical ignorance, Rawls might be taken to argue, parties can at most ensure that the pursuit of goods in general, rather than the enjoyment of unknown particular goods, is institutionally ensured. The social bases of self-respect would then just be the conditions without which persons could not develop and further their moral powers, taken abstractly. However, this line of defence is not persuasive. From the fact that parties do not know the specifics of their life-plans, it does not follow that they will not wish for their chances of success in their plans to be maximized. To recall, caution advises parties ignorant about their natural characteristics and conception of the good to conservatively opt for Rawls' primary goods. Now, the parties equally cannot be certain that their life-plan is not too costly or too exotic to materialize. Nothing guarantees that in actuality they will not aspire to such largely inaccessible careers as that of an astronaut or a high-fashion model. But sheer ignorance of the content of one's plans does not rule out one's concern about the *success* of one's plans. The deliberators' inability to ascertain whether or not they are in fact aspiring astronauts is not enough to produce their disinterest in the prospects for aspiring astronauts. They might still, once the veil is dropped, find themselves falling in that class, which might adversely affect their self-respect.[14] Consistent with the precautionary principle that Rawls' deliberators

---

[13] Eyal, "'Perhaps the Most Important Primary Good': Self-Respect and Rawls's Principles of Justice," pp. 208-09. For a similar argument see also Norman Daniels, "Equal Liberty and Unequal Worth of Liberty," in *Reading Rawls*, ed. Norman Daniels (New York: Basic Books, 1975).

[14] Thomas Pogge makes a similar claim about natural talents, in relation to self-respect. He contends that since Rawls' parties do not know whether or not they possess any outstanding natural talents, they will be inclined to ensure themselves against the possibility that they are

63

otherwise follow, they might for that choose maximal guarantees that even the most exotic conception of the good will be satisfied.

The just rehearsed defence of Rawlsian self-respect is unsuccessful because it is empty: it says nothing about what self-respect *is*. It just puts forward the claim that self-respect does not rely on one's success in practical pursuits – a claim that is implausible without support from some positive account of the proper bases of self-respect. Yet, Rawls is not committed to this vacant defence because he does have a substantive conception of self-respect, one that holds significant appeal to multiculturalists. Rawls believes that the self-respect of citizens has its basis in their mutual recognition of equal worth. Rawls thus notes that "our self-respect normally depends upon the respect of others. Unless we feel that our endeavours are honoured by them, it is difficult if not impossible for us to maintain the conviction that our ends are worth advancing."[15]

The appeal of Rawls' notion of recognition to multiculturalists can be readily identified. According to some prominent multiculturalist authors, there is but a short step between affirming the centrality of equal recognition and demanding the state protection of identities. On their view, secure identification can be understood as one of Rawls's social bases of self-respect. Anna Elisabetta Galeotti thus presents her idea of toleration as recognition as an extension of "the line of justification pointed out by John Rawls."[16] Shame, self-hatred and other

---

talentless: "Would not one's sense of self-worth be gravely damaged if one realized all along that one's limited natural talents give one no chance at being admitted to higher education (because admitting the less talented would reduce the lowest index position)? In fact, is not one's self-respect damaged more when one is excluded from higher education on account of one's lack of intelligence than when one is excluded on account of one's race or gender or the poverty of one's family? Unless exclusion based on *social* factors can be shown to be substantially more damaging to individuals than exclusion based on natural factors, then the special injustice we see in restrictions of opportunity based on social factors cannot be reaffirmed within a contractualist framework." (Thomas Pogge, *John Rawls: His Life and Theory of Justice*, trans. Michelle Kosch (Oxford: Oxford University Press, 2007), pp. 131-32, original emphasis.)

[15] Rawls, *A Theory of Justice*, p. 178.

[16] Anna E. Galeotti, *Toleration as Recognition* (Cambridge: Cambridge University Press, 2002), p. 113.

forms of negative self-regard that are associated with non-recognition are, according to her, condemned by Rawls' inclusion of self-esteem on the list of primary goods.[17] For Will Kymlicka, "The relationship between cultural membership and self-respect gives the parties to the original position a strong incentive to give cultural membership status as a primary good."[18] The loss of cultural membership is one condition that undermines self-respect, and that contractors behind the veil of ignorance would wish to avoid at any cost. Daniel Weinstock similarly writes:

> [I]f we accept that self-respect depends in part upon our ends being affirmed, or at least not demeaned, by our fellow citizens, then it will not be sufficient for the self-respect of those members of society for whom community membership is fully constitutive of their ends that their relation with their fellow citizens be mediated only through the two principles of justice as Rawls articulates them... For such people, being respected by their fellow citizens will require not only that they be respected as isolated individuals, but also *qua* members of the community which fully or partly constitutes their ends.[19]

---

[17] Ibid., p. 9.

[18] Will Kymlicka, *Liberalism, Community and Culture* (Oxford: Clarendon, 1989), p. 166.

[19] Daniel M. Weinstock, "How Can Collective Rights and Liberalism Be Reconciled?," in *Blurred Boundaries*, ed. Rainer Bauböck and John F. Rundell (Aldershot: Ashgate, 1998), p. 299. Laegaard argues in a similar fashion: "Provided a plausible causal connection between the location of public holidays and the self-respect of minority members can be established empirically, it seems that a liberal theory of recognition based on a concern with the social bases of self-respect within the scope set by the publicity constraint is at least as good a justification for such proposals as theories demanding affirmation of the value of minority cultures or identities." (Sune Laegaard, "On the Prospects for a Liberal Theory of Recognition," *Res Publica* 11, no. 4 (2005): p. 344.) For a similar argument see also Joel Anderson and Axel Honneth, "Autonomy, Vulnerability, Recognition, and Justice," in *Autonomy and the Challenges of Liberalism: New Essays*, ed. John Christman and Joel Anderson (Cambridge: Cambridge University Press, 2005).

3.3. The psychologistic conception of self-respect

In order to assess the presented impartial argument for recognition, it is important to spell out its central assumptions in more detail. To begin with, it relies on a claim that is a mainstay of multiculturalist and national writing, the claim that individuals' self-respect crucially depends on whether their identities are publicly affirmed and protected. Avishai Margalit and Joseph Raz thus write:

> It may be no more than a brute fact that people's sense of their own identity is bound up with their sense of belonging to encompassing groups, and that their self-respect is affected by the esteem in which these groups are held. But these facts, too, have important consequences. They mean that individual dignity and self respect require that the groups membership of which contributes to one's sense of identity be generally respected and not be made a subject of ridicule, hatred, discrimination, or persecution.[20]

The reader is assured that this last requirement is grounded in considerations of self-respect when Margalit and Raz declare it an "unexceptionable" premiss that "people's membership of encompassing groups is an important aspect of their personality, and their well-being depends on giving it full expression."[21] Yael Tamir is another author who strongly connects secure access to public self-identification with leading a good life:

> Membership in a nation is a constitutive factor of personal identity. The self-image of individuals is highly affected by the status of their national community. The ability of individuals to lead a satisfying life and to attain the respect of others is contingent on, although not assured by, the ability to view

---

[20] Avishai Margalit and Joseph Raz, "National Self-Determination," *The Journal of Philosophy* 87, no. 9 (1990): p. 449.

[21] Ibid.: p. 451.

themselves as active members of a worthy community. A safe, dignified and flourishing national existence thus significantly contributes to their well-being.[22]

However, the claim that individuals' self-respect depends on whether their identities are publicly recognized also needs clarification. In particular, it is important to note that it assumes a special understanding of self-respect, which can be explained by returning to Margalit and Raz. They argue that that the public protection of persons' identities is warranted by the fact that "their self-respect is affected by the esteem in which these groups are held". Their thought is that whether, and how far, one is able to nurture one's deepest attachments has a profound bearing on one's s elf-regard, and consequently one's self-respect. When a person lacks opportunities for publicly expressing his deepest

---

[22] Yael Tamir, *Liberal Nationalism* (Princeton, NJ: Princeton University Press, 1993), p. 73. According to Michelle Moody-Adams, three factors ensure that "the ability to have and affirm a robust sense of self-respect is greatly influenced by social circumstances": "First, the vocabulary in which one learns to give expression to one's self-conception, and even the concepts that initially shape that self-conception, are products of the linguistic conventions of a given community... Second, a society's normative expectations about emotion, thought, and action have an especially powerful influence on the development of self-respect. Every society gradually develops intricate patterns of normative expectations about what talents and abilities one ought to use in the service of self-preservation – even about what really constitutes survival or self-preservation... Further, self-contained communities within complex societies sometimes produce their own self-contained expectations about selves and self-respect. The self-conceptions of those in such communities will overlap very little with the self-conceptions of those outside such groups." (Michele M. Moody-Adams, "Race, Class, and the Social Construction of Self-Respect," in *Dignity, Character, and Self-Respect*, ed. Robin S. Dillon (London: Routledge, 1995), pp. 276-77.) Charles Taylor is even more ardent in condemning the negative effects of non-recognition: "The demand for recognition in these latter cases is given urgency by the supposed links between recognition and identity, where this latter term designates something like a person's understanding of who they are, of their fundamental defining characteristics as a human being. The thesis is that our identity is partly shaped by recognition or its absence, often by the *mis*recognition of others, and so a person or group of people can suffer real damage, real distortion, if the people or society around them mirror back to them a confining or demeaning or contemptible picture of themselves. Nonrecognition or misrecognition can inflict harm, can be a form of oppression, imprisoning someone in a false, distorted, and reduced mode of being." (Taylor, "The Politics of Recognition," p. 25, original emphasis.)

attachments, he may be deprived of an important means of affirming his own worth because it may bar him from various activities which he regards as central to his self-understanding. Suppose, for instance, that the individual considers it constitutive of his very personality that he is able to appeal to institutions of communal Islamic law to settle his disputes. It is plausible to expect that a fully secular political arrangement that disallows the appeal would foster in the person a feeling of estrangement from the society, and dissatisfaction with himself. The rejoinder that he is nonetheless free to attend to his projects privately gives rise to the objection that upholding a system of law cannot be a private matter altogether.[23] The psychological harm is even stronger in extreme cases where individuals are systematically and forcefully made to behave as if things that matter to them are in fact worthless. The real danger is that, in these instances, persons will begin to think of *themselves* as worthless; such duress radically thwarts their self-respect.[24] Although the diagnosis is weaker with regard to most identity-blind political arrangements, it may still warrant regarding them as morally objectionable.

Margalit and Raz's objection to strictly neutral arrangements is thus that persons under them are left with no room for self-respect, being deprived of the possibility of engaging in projects they hold valuable.[25] The understanding of self-respect that underpins their reasoning may be termed psychologistic, insofar

---

[23] Margalit and Raz similarly note: "To the extent that a person's well-being is bound up with his identity as a member of an encompassing group it has an important public dimension." (Margalit and Raz, "National Self-Determination," p. 452.)

[24] David Middleton expresses the point by introducing the term "reflexive self-respect", which is "the way in which others treatment of us affects the way we feel about ourselves. This can be very powerful and its effect on our well-being should not be underestimated. To be treated as worthless, as a means and not an end, as an object not a subject conveys powerful symbolic and material messages." (David Middleton, "Three Types of Self-Respect," *Res Publica* 12, no. 1 (2006): pp. 65-66.)

[25] Another interpretation of the objection, following David Sachs's conceptualization, might be that liberal neutrality affects persons' sense of their pride by suppressing some important activities, such as wearing one's religious symbols in school, that make them proud. (David Sachs, "How to Distinguish Self-Respect from Self-Esteem," *Philosophy and Public Affairs* 10, no. 4 (1981).)

as it relates the moral worth of persons to their subjective appraisals of their moral worth. The psychologistic approach equates self-respect with one's *sense* of self-respect. As Stephen Massey notes, on this view "it is both necessary and sufficient for respecting oneself that one have a certain kind of favourable self-attitude, which can be adequately defined in psychological terms."[26] According to the psychologistic interpretation, one's evaluation of one's qualities, achievements, character traits and other personal attributes is fully constitutive of one's self-respect. When the evaluation is positive, the person can be said to have self-respect; when it is negative, the person is lacking in self-respect. A more precise formulation of Margalit and Raz's complaint is, therefore, that positive self-appraisal is impossible in a system that demotes the pursuit of all commitments to the private sphere.

Another explanation of the psychologistic approach to self-respect can be given with reference to Stephen Darwall's well-known distinction between appraisal and recognition respect for persons.[27] For Darwall, the two kinds of respect stem from two kinds of attitudes: while appraisal respect for persons is based on a favourable assessment of their traits or achievements, recognition respect consists in an acknowledgement of some status that they have.[28] Darwall further maintains that acknowledgment of persons' moral status, associated with attaching certain rights and obligations to persons, is possible without having a positive view of their personal characteristics. He believes such non-evaluative recognition of free and equal standing to be the kind of respect that is owed to persons *as persons*. His claim is also that self-respect is properly understood as a

---

[26] Stephen J. Massey, "Is Self-Respect a Moral or a Psychological Concept?," *Ethics* 93, no. 2 (1983): p. 247. A similar interpretation is that the psychologistic view "treats self-respect as a psychological phenomenon that gains support from whatever behaviour one engages in that one happens to deem worthy of oneself. Although this psychological variant of self-respect requires fulfilling one's plans and measuring up to one's ideals, one's plans and ideals are relative to individual beliefs and desires." (Diana T. Meyers, "Self-Respect and Autonomy," in *Dignity, Character, and Self-Respect*, ed. Robin S. Dillon (London: Routledge, 1995), pp. 222-23.)

[27] Stephen Darwall, "Two Kinds of Respect," *Ethics* 88, no. 1 (1977).

[28] Recognition respect 'consists in giving appropriate consideration or recognition to some feature of its object in deliberating about what to do.' (Ibid.: p. 38.)

recognition of one's place in the moral community. This acknowledgment of one's own rightful status is secure because, being independent of one's evaluative self-esteem, it is not susceptible to variances in self-esteem. Now, the psychologistic approach of self-respect, favoured by multiculturalists and nationalist authors, reverses this last claim by completely sinking self-respect into self-esteem.[29] Middleton writes along these lines that "Repeated injuries to our self-esteem will lead eventually to a loss of self-respect."[30] Otherwise put, on the psychologistic view recognition self-respect is a function of one's appraisal self-respect. Where persons are unable to regard themselves in a favourable light, due to institutional hindrances, their self-respect is impaired.

Rawls' treatment of self-respect lends itself to the psychologistic interpretation. His definition explicitly equates self-respect with self-esteem, and portrays both as one's rational perception of one's worth:[31]

> We may define self-respect (or self-esteem) as having two aspects. First of all... it includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out. And second, self-respect

---

[29] Self-esteem can be seen as the self-regarding equivalent to Darwall's appraisal respect. Robin Dillon defines self-esteem as follows: 'First, self-esteem has an evaluative dimension, which is identified as a favourable self-appraisal, as an attitude of self-approval, or as involving the belief that one is significant, worthy, capable, or successful... A second feature of self-esteem is affectivity: it is or it influences how we feel about ourselves. The person who has self-esteem is said to feel good about herself, to like herself, to have feelings of personal worth... The affective dimension of self-esteem is widely regarded as what makes it valuable to individuals and motivationally primary.' (Robin S. Dillon, "Introduction," in *Dignity, Character, and Self-Respect*, ed. Robin S. Dillon (London: Routledge, 1995), pp. 30-31.)

[30] Middleton, "Three Types of Self-Respect," p. 69.

[31] "Rational" here means "supported by the thin theory of the good". Even if Rawls understands self-respect as one's sense of self-respect, this sense is hence not a mere feeling. Rawls seems to conceive of self-respect as rational (in the above sense) appraisal of one's moral worth: one's reasons for respecting oneself are exhausted by one's reasons for regarding oneself respect-worthy. I discuss Rawls's thin theory of the good, which grounds reasons of this sort, below.

implies a confidence in one's ability, so far as it is within one's power, to fulfil one's intentions.[32]

It is therefore no surprise that Rawls is fairly commonly read as espousing the psychologistic view. Massey, for instance, explicitly attributes to Rawls the view (although his term for it is "subjectivist").[33] Sune Laegaard similarly argues that Rawls regards self-respect in purely descriptive terms, as a "subjective psychological state standing in causal relationships to the actions of persons."[34]

Rawls' putative psychologism about self-respect is thus a matter of great importance for advocates of a liberal theory of recognition. If Rawls endorses the psychologistic view, the notion of identity can be neatly plugged into his argument about the social bases of self-respect. The resulting claim is that recognition has uncompromising merit because it is a social condition of the primary good of self-respect. If, however, Rawls does not espouse the psychologistic view of self-respect, the prospects for a Rawlsian impartial defence of recognition will be more uncertain.

3.4. The place of self-respect in Rawls's theory

Rawls' classification of the social bases of self-respect as one of the primary goods is not without problems. He indicates that they occupy a special place within the category of primary goods by calling them "the most important" primary goods. However, I wish to claim that the social bases of self-respect are in one important respect unlike other primary goods. It is that the argument for social conditions of self-respect as a primary good does not impose any independent requirement on the distribution of wealth and liberties. The unique

---

[32] Rawls, *A Theory of Justice*, p. 440.

[33] Massey, "Is Self-Respect a Moral or a Psychological Concept?," p. 250.

[34] Laegaard, "On the Prospects for a Liberal Theory of Recognition," p. 339. For another psychologistic interpretation of Rawls on self-respect see Eyal, "'Perhaps the Most Important Primary Good': Self-Respect and Rawls's Principles of Justice."

nature of the argument explains, or so I will maintain, why it cannot support multiculturalist claims for recognition.

Rawls believes that principles that define deliberative rationality, embodied in his "thin theory of the good" provide sufficient support for the choice of primary goods in the original position. To recall, the original position is seen by Rawls as a device of representation, whose function is to model the conditions of fair choice. In order for the contract argument to succeed, the restrictions of the veil of ignorance must not be so harsh as to make choice inconceivable. They must leave room for deliberation by permitting some criterion of goodness to the contractors, albeit one that is severely conditioned by the purpose of modelling fairness:

> Since these assumptions [about the parties' motives in the original
> position] must not jeopardize the prior place of the concept of
> right, the theory of the good used in arguing for the principles of
> justice is restricted to the bare essentials. This account of the good
> I call the thin theory: its purpose is to secure the premises about
> primary goods required to arrive at the principles of justice.[35]

Rawls's idea is that the thin theory of the good is sufficient to justify the selection of primary goods in the original position.[36] In other words, the principles of deliberative rationality − to strive for deliberative consistency, to

[35] Rawls, *A Theory of Justice*, p. 396.

[36] Samuel Freeman takes a different view, arguing that for Rawls primary goods are a component of the thin theory of the good, rather than its forced implication. Freeman distinguishes between the formal and the substantive aspects of the thin theory of the good, where the formal aspects include principles of rational choice, the idea of deliberative rationality, and the idea of a rational plan of life. The substantive aspects include the account of the parties' higher-order interests, the primary goods, and the Aristotelian principle. (Freeman, *Rawls*, p. 147.) However, Freeman's reading is in conflict with Rawls' text, which for instance states that 'we need what I have called the thin theory of the good *to explain the rational preference for primary goods* and to explicate the notion of rationality underlying the choice of principles in the original position. This theory is necessary to support the requisite premises from which the principles of justice are derived. (Rawls, *A Theory of Justice*, p. 397, emphasis added.)

rank one's ends in an order of priority, not to prioritize one's interests at any particular point in time, to prefer having more than less goods, and so on – are alone meant to support the choice of primary goods. For example, in circumstances of radical uncertainty about one's actual condition, the rational principle of preferring more to fewer goods helps justify the focus on primary goods as the minimally required bundle of resources. This is why Rawls calls his conception of goodness under the veil of ignorance "goodness as rationality".

Rawls's focus on principles of deliberative rationality is illuminating because it provides the basis for his argument about self-respect. That argument can be approached by considering Massey's critical reception of it. Massey concedes that such goods as maximal basic rights and liberties, compatible with similar rights and liberties for others, or unconstrained access to all public offices, are necessary for the pursuit of any conception of the good. But the same reasoning does not seem to apply to the supposed primary good of self-respect. It appears dubious to assume that without self-respect no other goods could be attained or enjoyed. Massey remarks:

> Self-respect might be the most important primary good were it true that without it other goods have no value, or that one could enjoy nothing else without self-respect. Yet neither of these claims is true. It is false that Rawls's other primary goods, for example, income and opportunities, have no value for a person who has little or no self-respect. It is surely an excess of rationalism to claim that a person cannot enjoy going to the beach or to a baseball game unless he respects himself.[37]

The way for Rawls to counter Massey's criticism is to maintain that the *pursuit* of ends, rather than their enjoyment, would be impossible without self-respect. So, while it is true that one could still derive satisfaction from watching a baseball game without respecting oneself, Rawls' point runs deeper. It states that one would not bring oneself to go to the stadium in the absence of the belief that

---

[37] Massey, "Is Self-Respect a Moral or a Psychological Concept?," p. 259.

one's needs for entertainment and socializing deserve to be met. Engaging in any activity, Rawls plausibly claims, presupposes the agent's conviction that acting is worthwhile – or, since it is the agent that acts: it presupposes his conviction that he has worth. Without this conviction, "our sense of the value of accomplishing our aims,"[38] there would be no reason to do anything. By self-respect Rawls thus refers to the contractors' affirmation of the value of their practical reasoning.[39]

This clarification also reveals a further point about Rawls's concept of self-respect – one that ultimately turns out to be rather disappointing for multiculturalists. It is that in Rawls's usage that concept applies exclusively in the context of the original position. It picks out a feature of the *contractors*, stylized as they are, not ordinary persons in everyday situations. In this regard, Rawls uses the concept of self-respect much in the same way as the concept of autonomy, another notion that commonly has wider application. Rawls thus writes:

> The idea of the initial situation is central to the whole theory and other basic notions are defined in terms of it. Thus acting

---

[38] Rawls, *A Theory of Justice*, p. 181.

[39] The question of why parties rationally prefer practical reasoning to inactivity is unresolved in Rawls. In his contractarian framework, the value of self-respect must be constituted by the fact that it would be rationally affirmed by parties in the original position. But why is it rational for parties to protect the value of practical reasoning, rather than be content with passivity? The value of self-respect cannot be proven by appealing to standards of validity *in* deliberation – that is, norms (transitivity, consistency, etc.) for the rational conduct of practical reasoning. The required answer would need to demonstrate the rationality *of* deliberation. It might be thought that the question is addressed by Rawls' "Aristotelian Principle" which states: "Other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity. The intuitive idea here is that human beings take more pleasure in doing something as they become more proficient at it, and of two activities they do equally well, they prefer the one calling on a larger repertoire of more intricate and subtle discriminations." (Ibid., p. 426.) However, the purpose of the principle is only to favour more complex over less complex life-plans. This is evident in Rawls' use of the principle to suggest that parties in the original position would rationally prefer playing chess to playing checkers. The parties' rational preference for doing something, be it chess or checkers, over doing nothing remains unexplained.

autonomously is acting from principles that we would consent to as free and equal rational beings, and that we are to understand in this way.[40]

This strictly contractarian definition of autonomy runs counter to the more customary usage, in which the concept of autonomy is a tool for describing and explaining an aspect of the human condition. This is, for example, Raz's usage when he speaks of "the ideal of autonomy as a life freely chosen."[41] His definition aims to capture what self-mastery is generally. However this is not Rawls's purpose when he invokes the concept of autonomy – and something quite similar can be said of his treatment of self-respect. For Rawls, self-respect is equally to be understood as a feature of appropriately situated persons in the original position. It is useful to think of that concept as having a specific function, which rules out its application outside the original position. That function is purely constructive (rather than descriptive or explanatory): to play a part in the setup of the situation that must yield certain principles of justice. Its role is as part of the conceptual framework that Rawls uses in deriving determinate norms of justice from a basic commitment to individuals' freedom and equality. Specifically, it addresses the most fundamental question that Rawls's contractarian theory faces: not why we can assume that persons, under certain constraints, would choose some specific conception of justice, but why they would bother to do any choosing in the first place. Rawls's answer is that they would so because they place supreme value on their capacity to design their lives according to their own choosing.

Rawls's concept of self-respect must be understood in relation to its function in theory-construction. Now, since that function is essentially restricted to the original position, so is Rawls's concept of self-respect. It is then misplaced to ask about the real-world social bases of self-respect, as the multiculturalists do. This has implications for how one may understand the recognition that Rawls deems necessary for self-respect. In particular, Rawls's claim that was so promising,

[40] Ibid., p. 516.

[41] Raz, *The Morality of Freedom*, p. 371.

that self-respect relies on the mutual recognition of citizens, must also be interpreted in light of the original position. The recognition that is at issue here is the mutual acknowledgement of equal status by individuals behind the veil of ignorance. It is not a notion that is appropriate in speaking about flesh-and-blood citizens in everyday political circumstances. This helps explain why Rawls thinks that the recognition that is central to self-respect is sufficiently accomplished by common public adherence to his two principles of justice.[42] What this recognition based on justice delivers is the assurance of equal citizenship. The contractors' self-regard is unrelated to cultural belonging because knowledge of cultural identifications is ruled out from the original position. Therefore, the only recognition that can be the source of their self-respect is their mutual acknowledgement of equal civic status. And this does not entail recognizing each other as Nigerians, Muslims, or Francophones. It means nothing more than affirming each other's standing as a full participant in the enterprise for mutual advantage that is the state.[43]

The breakdown of the connection between the public affirmation of cultural identities and Rawls's argument about the social bases of self-respect is due to deep structural incompatibility. The link cannot be established without violating Rawls' commitment to the priority of the right over the good, which leads him to banish cultural identifications from the original position. As a result, this particular version of the argument that identities generate reasons grounded in self-respect fails. This does not mean that other versions of that argument cannot

---

[42] Samuel Freeman's reading of Rawls emphasizes the same point: "Equal basic liberties, fair equal opportunities, and political and economic independence are primary among the bases of self-respect in a democratic society." (Freeman, *Rawls*, pp. 186-87, original emphasis.) In *Political Liberalism* Rawls writes: "The social bases of self-respect are explained by the structure and content of just institutions together with features of the public political culture, such as the public recognition and acceptance of the principles of justice." (Rawls, *Political Liberalism*, p. 181.)

[43] As Joshua Cohen notes, within Rawls' framework others show me respect "by acknowledging and protecting my right to bring my sense of justice to bear on public affairs." (Joshua Cohen, "For a Democratic Society," in *The Cambridge Companion to Rawls*, ed. Samuel Freeman (Cambridge: Cambridge University Press, 2002), p. 109.)

work. However, the Rawlsian avenue for making this point, which is particularly promising as well as elaborated in multiculturalist writings, seems closed.

4. Frankfurt's volitional necessities

4.1. Introduction

In this chapter I will examine whether the idea of constraints on what reasons persons can come to have can be supported by Harry Frankfurt's argument about volitional necessities. That argument is made possible by the decisive role that Frankfurt attributes to the will in practical reasoning. For him, the will takes priority over the intellect in practical decision-making, in the sense that it sanctions the practical proposals of the intellect. This view amounts to a denial that reasons and reason itself, as the faculty that operates through the use of reasons, are essential to practical deliberation.[1] It also opens up a possibility that is of special interest here, the possibility that the will is unable to sanction any proposal of the intellect. Frankfurt terms such limitations on the range of reasons that a person can adopt volitional necessities. An essential aspect of the will's constitution, he suggests, is that it is limited in this way. Our choices are always and necessarily constrained, not because there are always external impediments to our choice, but because there are certain options that we cannot bring ourselves to choose. Since volitional necessities are inescapable for us, they fixate our personal standards of value:

> The necessities of a person's will guide and limit his agency. They determine what he may be willing to do, what he cannot help doing, and what he cannot bring himself to do. They determine as

---

[1] Cf.: "The supposition that people cannot make decisions or perform actions except for a reason strikes me as belonging to an excessively rationalistic conception of human life – a conception that is both theoretically gratuitous and false to the facts." (Harry G. Frankfurt, "Reply to Eleonore Stump," in *The Contours of Agency*, ed. Lee Overton and Sarah Buss (Cambridge, MA: MIT Press, 2002), p. 89.

well what he may be willing to accept as a reason for acting, what
he cannot help considering to be a reason for acting, and what he
cannot bring himself to count as a reason for acting. In these
ways, they set the boundaries of his practical life; and thus they
fix his shape as an active being.[2]

Before examining Frankfurt's argument for volitional necessities, I will clarify
what he means by the "will". An important question for understanding this
argument is whether it rests on a conception of willing as a mode of desiring, or a
conception of willing as distinct from desiring. I will turn to discuss Frankfurt's
views about volitional necessities, in sections 4.3 and 4.4, only after this question
is settled. I begin by examining the understanding of the will as desiring, which
is standardly attributed to Frankfurt.

4.2. Willing as desiring

Frankfurt's oft-cited exploration of the concept of personhood starts with his
observation that the central features of personhood are to be sought not in the
details of our physical constitution, but in the operation of our minds.[3] Non-
human aliens with bodily features vastly different to ours, should they exist, may
qualify as persons so long as they possess the right sort of psychic arrangement.
A necessary condition for this organization is the possession of second-order
desires, which serves to filter out non-human animals − organisms that are
capable of desire-based purposeful behaviour[4] but incapable of evaluating their

---

[2] Harry G. Frankfurt, *The Reasons of Love* (Princeton: Princeton University Press, 2004), p. 50.

[3] Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68, no. 1 (1971).

[4] Frankfurt goes so far as to maintain that humans are not unique in their capacity to act on reasons: "Insects have reasons, to which they respond, for their defensive scurrying about, as they do when someone comes after them with manifestly deadly intent. Their movements are neither random nor tropistic, they have alternatives, and they make mistakes. It is difficult to make sense of their behaviour without understanding it as in some way rational." (Frankfurt, "Reply to Eleonore Stump," pp. 61-62.)

desires and acting on these evaluations. Insofar as they cannot discriminate between their wants but are merely thrown around by them, their behaviour falls short of the standards of *acting*. However, although human beings possess the ability reflectively to detach themselves from their desires, this fact alone is not the basis of our referring to them as persons. Some classes of humans, such as very small children and the mentally ill, may on Frankfurt's account fall outside the remit of personhood if their reasons and desires are not appropriately organized. The necessary and sufficient condition for calling some organism "a person" – whether or not it belongs to the human species – is that it has second-order volitions.[5]

An agent's volition is, for Frankfurt, his effective desire – a "desire that is motivating or moving A to do what he is actually doing or that A will in fact be moved by... (unless he changes his mind) when he acts."[6] An example that Frankfurt uses to illustrate this analysis involves a physician committed to helping his drug-addicted patients, and wishing to gain – in the interest of this cause – a first-hand experience of their addiction. His motive for taking the drug is not focused on the elating effects of his consumption, as in his patients. Instead, it is related to his ambition to discover what it feels like to be addicted. This experience of addiction is not in itself related to the sensations felt under the influence of the narcotic. His endeavour to experience the addiction may, in principle, be fulfilled without his tasting the addictive substance. It is enough for this purpose that he feels the physical and psychological compulsion that torments his patients. In fact, the physician's only reason for taking the drug may be just to get himself into the grip of the addiction, after which the need for the drug, from his scientific standpoint, vanishes. Subsequent to this necessary initiation, argues Frankfurt, the doctor wants to have a desire for the drug without wanting this desire to be effective. He does not want the desire for the drug to constitute his will. Volition is on this account a "want to want to" do something (here: take the drug), which makes it impossible for there to be a volition without a desire it is directed at. It also allows that an agent can deem certain desires unfit

---

[5] Frankfurt, "Freedom of the Will and the Concept of a Person," p. 16.

[6] Ibid.: p. 8.

to govern his behaviour, and regard their grip on his actions – when they do so invade him – as usurpatory rather than rightful.[7]

Frankfurt's picture of personhood, based as it is on a definition of willing as a mode of desiring, is less than satisfactory. Problems with it begin right from Frankfurt's decision to explain agency in terms of a hierarchical ordering of desires. The hierarchical view of deliberation seems attractive as it fits well with the appealing "belief – desire" model of practical reasoning, sometimes associated with Hume. Consonant with this model, Frankfurt's approach accounts for the intentional content of practical reasoning with sole reference to desires about what there is to be.[8] This makes his claims parsimonious and metaphysically inoffensive. It is also the case, however, that the coherence of Frankfurt's picture of agency suffers as a result of this "desires-only" policy. To be noted, first, is that the regulative status of volitions requires that they be in some way *special* with respect to the desires that they regulate. This special status manifestly cannot be a reflection of any difference in kind, since volitions are themselves presumed to be desires. The required special character of volitions also cannot lie simply in the fact that they are of a higher-order than the desires they govern. It is not at all clear why, say, a "desire to recycle" should be automatically overruled by the "desire not to desire to recycle", nor why it should be automatically authenticated by "a desire to desire to recycle". The governing relation of volitions over desires is precisely what stands in need of explanation, and is obfuscated by their alleged structural similarity. As Gary Watson remarks, the problem with higher-order volitions is

---

[7] A special case is that of "wantons", individuals who have second-order desires but are indifferent as to whether they will act on any of them, and as to which of these desires will end up determining their behaviour. Frankfurt's wantons thus have no second-order volitions whatsoever corresponding to their second-order desires.

[8] I do not wish to imply that Frankfurt's account of personhood is in fact guided by the requirements of the belief-desire model, although Frankfurt is often read as a straightforward Humean. In fact, below I will argue that his position owes considerably more to Descartes than to Hume. Here I only suggest that a portion of the appeal of his position lies in its ability to accommodate the claims of the belief-desire model.

simply that they are just, after all, desires, and nothing about their level gives them any special authority with respect to externality. If they have that authority they are *given* it by something else. To have significance, the hierarchy must be grounded in something else that precludes externality.[9]

It is in anticipation of this type of criticism that Frankfurt develops his notion of identification. Volitions have a special commanding position over our generic desires because they express our identification with some of these desires. Although Frankfurt himself admits that this idea of identification is "mystifying", it does denote something like accepting a desire as representative of oneself.[10] This fact of acceptance alone accounts for their authority over other desires. It also "'resounds' throughout the potentially endless array of higher orders,"[11] meaning that it blocks the regress into desires of an ever higher order. Identification is decisive, insofar as it conclusively settles the question of what desires we care to act upon. Furthermore, Frankfurt regards the notion of identification as capturing the active aspect of agency. By endorsing a want as authentically ours, and committing ourselves to be guided by it, we stamp the mark of our authorship on our conduct. Identification moves one away from the status of a passive receptacle for desires, into the role of an active agent who "owns" his behaviour.

Frankfurt's answer to Watson-type criticism is that what lends second-order volitions their special commanding authority is, indeed, not the mere fact that they are of a higher-order. The authorizing fact about them is that they are grounded in the agent's satisfaction with being ruled by them.[12] However this answer too is contentious, principally because of the way in which Frankfurt

---

[9] Gary Watson, "Free Action and Free Will," *Mind* 96, no. 382 (1987): p. 149, original emphasis.

[10] For this point, as well as the relationship between identification and practical activity, see also Harry G. Frankfurt, "Three Concepts of Free Action," in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988).

[11] Frankfurt, "Freedom of the Will and the Concept of a Person," p. 16.

[12] See Harry G. Frankfurt, "The Faintest Passion," in *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999), p. 105.

construes "satisfaction". He conceives of it in purely negative terms, as the absence of unsettlement about the prospect of acting on some impulse. A desire that we identify with is just the one that we do not care to rebel against, one whose grip on our behaviour we have no inclination to loosen. The controversial conclusion that Frankfurt draws is that our behaviour stems from our practical activity when it is driven by desires that we submit to without resistance. Markedly absent from this account is any sort of evaluation as a criterion for endorsement.[13] Satisfaction that is at the basis of practical reasoning is a purely natural, non-normative phenomenon; it is "a state of the entire psychic system"[14] that is wholly analyzable by the disciplines of psychology and neurology.

However, Frankfurt's naturalized account of agency must confront at least two serious objections. The first, advanced by Joseph Raz, states that the satisfaction that is presumed to stop the regress into desires of an ever higher order itself cries out for explanation. If a play of desires is all that deliberation involves, the person's state of non-resistance must also be rooted in the fulfilment of some desire. Since there must always exist a desire that "authenticates the authenticator", Frankfurt fails to avoid the trap of a dizzying infinite ascent to higher-order desires.[15] The second potential criticism is that if the outcome of identification is a just another desire (in the form of a volition) it is unclear why having *higher*-order volitions should be considered the mark of personhood. After all, first-order desires are desires all the same, and equally eligible to be identified with. Frankfurt's position thus seems to render reflexivity – the detachment we have in relation to our impulses – superfluous as far as agency is concerned.

---

[13] Cf.: "[W]hat I have actually intended to convey by referring to "endorsement" is not that the agent *approves* of what he is said to endorse, or that he considers it to *merit* his support, but nothing more than that the agent *accepts* it as his own. The sense in which he accepts it as his own is quite rudimentary. It is free of any suggestion concerning his basis for accepting it and, in particular, it does not imply that he thinks well of it." (Frankfurt, "Reply to Eleonore Stump," p. 87, original emphasis.)

[14] Frankfurt, "The Faintest Passion," p. 104.

[15] See Joseph Raz, "When We Are Ourselves: The Active and the Passive," in *Engaging Reason: On the Theory of Value and Action* (Oxford: Oxford University Press, 2002), pp. 18-21.

4.3. Willing as a distinct mental faculty

If they are well-directed, the cited objections damage Frankfurt's conception of identification, as well as the resulting picture of personhood. Fortunately, it is plausible to think that his argument regarding volitional necessities does not rest on the conception of willing as desiring. In contrast to most commentators, I wish to suggest that the argument in question is strongly informed by a Cartesian conception of willing as qualitatively different from desiring. For the purpose of elucidating Frankfurt's views on volitional necessities it will therefore be useful to examine how they link up with Descartes's treatment of the will. A convenient entry into the topic is through Frankfurt's discussion of moral responsibility.

4.3.1. Inability to do otherwise and freedom of the will

In a much-discussed paper, "Alternate Possibilities and Moral Responsibility," Frankfurt argues that moral responsibility is compatible with over-determination of the agent's behaviour by external factors.[16] The paper targets a principle that Frankfurt thinks is held true by all parties in the long-standing debate on free will and determinism, the *principle of alternate possibilities* (PAP).[17] According to this principle, a person is morally responsible for what he has done only if he could have done otherwise.[18] By extension, a person A is not morally responsible for an act X if he could not do other than X. Frankfurt's assessment of the claims presented by PAP moves through two stages: first, he analyses what it means to say that A "could not do otherwise" than X and looks for a paradigmatic situation to which the qualification applies. Second, he asks whether in the pure case of inability to do other than X the agent is exempt from responsibility for doing X.

---

[16] Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66, no. 23 (1969).

[17] For a defence of PAP see Joel Feinberg, *Rights, Justice, and the Bounds of Liberty* (Princeton: Princeton University Press, 1980), pp. 36-40.

[18] Frankfurt, "Alternate Possibilities and Moral Responsibility," p. 829.

It would seem that coercion produces the clearest and most standard cases of inability to do otherwise. Suppose that a person A was presented with an irresistible threat — say, of being killed — unless he performs an act X. When A performs X, should he be held responsible for it, and is our assessment influenced by the drastic curtailment of his options? Sometimes but not always, answers Frankfurt. While it is true that A decides to do X in the presence of the relevant threat, it is nonetheless incorrect to assume that he necessarily decides so *because* of that threat. It is of course reasonable to expect that on some occasions this kind of external conditioning of the agent's behaviour will occur — namely, that the threat will perform the dual role of providing both the context and the reason for A's decision to do X. In that case the agent's deliberation will have been overpowered, or "stampeded", by the coercive threat and we would rightly think that he should take no blame or praise for doing what he did.

However, Frankfurt urges us to consider two distinct cases in which A may have decided to perform X "for reasons of his own", even if he was threatened do X.[19] First, it may be that when making up his mind A was recklessly indifferent to the threat, refusing to acknowledge its potential consequences and stubbornly persisting with his original intention to do X. In this case the threat exerted no effective force on A, even if only because A appears foolishly unreflective by most standards. Second, one may also imagine that A was fully aware of the unsavoury prospect of dying and was reasonably impressed by it, but equally had a pre-existing intention to do X. He counts himself lucky for having to do what he always wanted to anyway. In both scenarios A's behaviour stems from his own reasoning, and he should be held responsible for performing X. No doubt, it will be extremely difficult to ascertain whether A's compliance with the threat is a result of happy circumstance or the overpowering force of the threat. On the one hand, this means that assigning responsibility will often be difficult, even in what appear to be the clearest cases of coercion. However, it also means that such cases do not properly illuminate the inability to do otherwise, which is to

---

[19] Ibid.: pp. 831-33.

say that "the doctrine that coercion excludes moral responsibility is not a particularized version of the principle of alternate possibilities."[20]

A better illustration of the inability to do otherwise, in Frankfurt's view, is given by the following example: On a late evening, I see a man being robbed and beaten by several attackers right in front of my house. My initial inclination is to call the police, but it soon gets superseded by my desire to avoid tedious police interrogation, possible retaliation by the attackers, or just losing sleep. And so I decide not to call the police. However, unbeknownst to me, the telephone system in my entire area has been out of order, which means that I would not have been able to inform the police even if I had wanted to.[21] Am I morally responsible for failing to reach the police? Regardless of what my decision was, I would have ended up not reaching the police. The concealed element of the telephone malfunction ensures that events can take only one turn. In other words, from a wider objective viewpoint, there is no alternative possibility to my failing to inform the police, since the possibility of me reaching the police equals zero.

As far as the world is concerned there is only one way things will pan out, which leads van Inwagen to claim that I am not morally responsible for failing to reach the police. He calls the thesis underlying this claim the *principle of possible action* (PPA): "A person is morally responsible for failing to perform a given act only if he could have performed that act."[22] I cannot be responsible for not

---

[20] Ibid.: p. 833.

[21] The example is Peter van Inwagen's variation on Frankfurt's original scenario. (Peter van Inwagen, "Ability and Responsibility," *Philosophical Review* 87, no. 2 (1978): pp. 204-05.) Another variation on the same theme, featuring an electronic device secretly implanted in the agent's brain, is found in John Martin Fischer, "Responsibility and Self-Expression," *The Journal of Ethics* 3, no. 4 (1999). Frankfurt's original scenario runs as follows: Without Jones$_4$'s knowledge, Black is capable of perfectly second-guessing the choices that Jones$_4$ is about to make, and is prepared to effectively obstruct any of Jones$_4$'s actions that do not please him. Whatever it is that he decides, Jones$_4$ will thus always do what Black wants him to do. (Frankfurt, "Alternate Possibilities and Moral Responsibility," pp. 835-36.) I present van Inwagen's version of the same scenario because it forms the basis for his criticism of Frankfurt, which I discuss below.

[22] van Inwagen, "Ability and Responsibility," p. 204.

causing some state of the world (the police knowing about the assault) if causing it is impossible for me. The notion of responsibility does not apply to my situation, just as it does not apply when I contemplate my "failure" to stop the Sun from rising this morning. Van Inwagen's claims do not imply that I should be exempt from responsibility for having the mental states that I do. It may be that the inclinations that figured in my decision were altogether inappropriate, that I displayed an unacceptable degree of inertness and inconsiderateness when witnessing blatant wrongdoing. However, failing to *try* (or to want to try) to contact the police is not the same as failing to contact the police. I may be considered guilty of the former, but not of the latter.

In his response to van Inwagen, Frankfurt argues that his account obscures the moral aspect of responsibility.[23] According to van Inwagen, argues Frankfurt, whether one describes a person as merely "trying to call the police" or "calling the police" depends solely on the condition of the telephone system. One and the same physical activity − dialling the number − will count as trying to call the police if the telephones are out of order, and calling the police if the system is functional. My liability to praise or blame thus entirely hinges on what is the appropriate description of facts about the telephone network. Yet, Frankfurt observes, this position is inadequate because we aim to assess *my* responsibility, not that of the telephone system. This essentially agent-referential character of our moral assessments can only be retained if they depend, at least in part, on facts about agents − their intentions, motives, reasons and so on. In other words, the morally important aspect of statements about responsibility is that they pick out a person that falls under them. Once every reference to the person is removed from such statements, they completely lose their moral dimension. Last night's telephone malfunction would, after all, be morally uninteresting had it not coincided with my deliberation on how to respond to the perceived assault.[24] Therefore, contrary to van Inwagen, even the most correct and exhaustive

---

[23] Harry G. Frankfurt, "What We Are Morally Responsible For," in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988).

[24] Or someone else's mental happenings.

description of relevant non-mental facts alone fails to capture the grounds for assigning moral responsibility.

However, Frankfurt's most direct attack on the various versions of PAP is that a person's inability to do otherwise does not annul his freedom. To explain, the conditions for holding someone responsible for φ-ing are dependent on the conditions for holding him free in φ-ing. In other words, we can only blame or praise a person for acting in a certain way if he was free in so acting. The appeal of PAP seems to lie in the fact that it captures one necessary condition of this freedom that is required for responsibility. If there was no way in which the world could have conformed with an agent's intentions, then he could not have been free, or in turn responsible for what he did. Against this argument, Frankfurt claims that ability to do otherwise is not a necessary condition of freedom. Consequently, unlike van Inwagen, Frankfurt does not think that this inability annuls moral responsibility, even in its paradigmatic cases such as the robbery scenario. In support of his view, Frankfurt points out that both coercion and overdetermination by external factors constrain the power of the agent's will to initiate bodily movements. However, they do not infringe his freedom to choose a harsh punishment rather than obeying, or select the alternative that will turn out to be impracticable. As Frankfurt explains:

> Given that the freedom of a person's will is essentially a matter of whether it is up to him what he does, it is more a matter of whether it is up to him what bodily movements he makes than of what consequences he can bring about by his movements.[25]

These claims are clearly grounded in a particular understanding of the freedom that is central to our assessment of an agent's moral responsibility – on Frankfurt's view, it is the freedom to will that certain states of affairs obtain. Importantly, the freedom in question is not the freedom to cause that the willed states of affairs actually come into existence. Whether my willing that X can bring about X is irrelevant to the question of whether I am free to will that X. To

---

[25] Frankfurt, "What We Are Morally Responsible For," p. 103.

claim otherwise would be to confuse will's freedom with will's power.[26] This non-causal analysis of freedom gives rise to a compatibilist position: even if it turns out that my lack of power renders me incapable of actualizing *anything* that my will proposes, my will can still be free. By extension, even if every action that I set on performing gets frustrated by overwhelming coercive threats, subtle manipulation, or hidden facts about the natural world, I can still be held accountable for my behaviour.

Frankfurt's conclusions up to this point are negative, consisting in a rejection of PAP for the purpose of drawing moral assessments. Before moving on to discuss his positive account of responsibility, a possible objection against Frankfurt's treatment of PAP is worth noting. It focuses on the peculiar understanding of practical reasoning that underlies his account, which regards the working of practical reason as a purely mental exercise, explicable without reference to its effects in the external world. As will be discussed further below, so long as the agent can act on "reasons of his own" he is free. This however runs counter to our strong intuition that practical deliberation does bring about outcomes outside our heads – indeed, that seems to be the point of reasoning practically. It seems therefore that an explanation of free deliberation must answer rather than avoid the question of how our mental activity spontaneously produces non-mental happenings. In order to identify Frankfurt's resources for addressing this worry, one needs to appreciate the large debt that his practical philosophy owes to Descartes. His response to the problem of harmonizing the causal determination of nature with our spontaneous agency bears striking resemblance to Descartes's. From Frankfurt's perspective, the objection above is not too troubling as it is effective only against those who lay claim to, or hope for, a full understanding of human agency within the natural world. Frankfurt, however, has no such

---

[26] Cf.: "When we ask whether a person's will is free we are not asking whether he is in a position to translate his first-order desires into actions. That is the question of whether he is free to do as he pleases. The question of the freedom of his will does not concern the relation between what he does and what he wants to do. Rather, it concerns his desires themselves." Frankfurt, "Freedom of the Will and the Concept of a Person," p. 15.)

pretensions. To understand his position, it is necessary to consider at some length its Cartesian roots.

Descartes acknowledges that every part of the natural world is subject to causal necessitation. Since we initiate changes in the same world we must be capable of starting a new causal chain. Yet since we also belong to that world all our actions must themselves be caused. In conjunction, these last two claims render a full analysis of human freedom within a causally determined world outstandingly difficult to provide. While recognizing the problem, Descartes is not too troubled by our poor understanding of the place of agency in the system of nature. The pursuit of this understanding, he argues, is not of primary philosophical importance, nor is its attainment the ultimate objective of philosophical enquiry. The reason for this lies in the limited nature of human cognition in the face of God's omniscience. The full comprehension of the link between our mental processes and natural causality is, for Descartes, God's prerogative. Given the vast gulf between God's cognitive capacities and our own, even hoping for such comprehension is arrogant as well as unreasonable.[27]

Descartes's philosophizing does not begin by enquiring into metaphysical truths, precisely because he believes humans to be constitutionally ill-equipped for such an investigation. Our senses can, and indeed often do, deceive us. Sticks half-submerged in water appear to us bent even though they are not, buildings seem to us unrealistically small when viewed from a great distance, and so on. From

---

[27] Cf.: "But we shall get out of these difficulties if we remember that our mind is finite, while the power of God is infinite – the power by which he not only knew from eternity whatever is or can be, but also willed it and preordained it. We may attain sufficient knowledge of this power to perceive clearly and distinctly that God possesses it; *but we cannot get a sufficient grasp of it to see how it leaves the free actions of men undetermined.* Nonetheless, we have such close awareness of the freedom and indifference which is in us, that there is nothing we can grasp more evidently or more perfectly. And it would be absurd, simply because we do not grasp one thing, which we know must by its very nature be beyond our comprehension, to doubt something else of which we have an intimate grasp and which we experience within ourselves." (René Descartes, *The Philosophical Writings of Descartes*, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch, vol. 1 (Cambridge: Cambridge University Press, 1985), Prin. I:41, p. 206, emphasis added.)

these findings Descartes infers that no secure knowledge can be gained from the deliverances of the senses. So long as we remain in their hold, we are always liable to deception. Hence, he famously claims, bottom-rock certainty can only be had by provisionally denying, or bracketing, all that our experience has taught us about the world and ourselves.

Similarly to Descartes, Frankfurt opposes the speculatively metaphysical ambition to explain the fit between human agency and natural causation. Being free from Descartes's religious commitments, Frankfurt does not surrender the striving for a fully integrated account of human action. Quite modestly, however, he contends that the examination of the mind/world fit must start from that with which we are familiar – namely, the content of our mental states.[28] Only through a solid explanation of the psychic processes involved in agency can we hope to gain insights into how our agency causes changes in non-mental reality. As a result, Frankfurt maintains that his account of freedom, which leaves the effects of free agency unaddressed, is nonetheless not fatally vacuous. It merely reflects the fact that a much deeper understanding of the psychic component of agency is needed before we can move on to discuss its external effects:

> Determinism surely does not require us to think of ourselves
> merely as locales in which various events are caused to occur. The
> trouble is that no good account of the difference between being
> passive and being active is available. We are agents, even if it is a
> fact that everything in our lives is caused. But until it has been
> explained what being an agent means, the compatibility of
> determinism with our agency (and hence with our moral

---

[28] It is, presumably, this attitude of modesty that lies behind Frankfurt's claim that the aim of his enquiry into freedom is "primarily to locate the problem with which a person is most immediately concerned when he is concerned with the freedom of his will." (Frankfurt, "Freedom of the Will and the Concept of a Person," p. 19.) The metaphysics of agency is not only unclear, it is also of secondary importance in the heat of action.

responsibility) cannot be decisively established or fully understood.[29]

4.3.2. Volitional necessities

Frankfurt does not call on Descartes only to argue that his conception of freedom need not explain the place of freedom in the system of causation. He also explicitly uses some of Descartes's findings in developing his positive account of practical deliberation. Frankfurt's contention that his claims about the ground of practical normativity are intended as "significantly analogous to [Descartes's] argument about the ground of theoretical reason" is instructive in this regard.[30] Frankfurt's views about agency are therefore best approached by assuming that he understands the will, in Cartesian fashion, as qualitatively distinct from desires.[31] On the Cartesian conception, the will is a mental faculty that is independent from the intellect, and that governs the endorsement or rejection of beliefs (proposals of the intellect).[32] The will thus takes priority over the intellect in judgment-formation.[33] Now, Frankfurt explicitly professes his sympathies for Descartes's concept of the limitless will.[34] Noting this point is not useful just for the purpose of lineage-tracing. It also importantly clarifies that Frankfurt takes

---

[29] Harry G. Frankfurt, "Reply to John Martin Fischer," in *The Contours of Agency*, ed. Lee Overton and Sarah Buss (Cambridge, MA: MIT Press, 2002), p. 29.

[30] Harry G. Frankfurt, *Taking Ourselves Seriously & Getting It Right*, ed. Debra Satz (Stanford: Stanford University Press, 2006), p. 106n6.

[31] Unfortunately perhaps, Frankfurt's analysis of the concept of the person remains his most widely cited work.

[32] "[T]he will simply consists in our ability to do or not to do something (that is, to affirm or deny, to pursue or avoid); or rather, it consists simply in the fact that when the intellect puts something forward for affirmation or denial or for pursuit or avoidance, our inclinations are such that we do not feel we are determined by any external force." (René Descartes, *The Philosophical Writings of Descartes*, trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch, vol. 2 (Cambridge: Cambridge University Press, 1985), Med. IV, p. 40.)

[33] On this point see John Cottingham, "Descartes and the Voluntariness of Belief," *Monist* 85, no. 3 (2002).

[34] Harry G. Frankfurt, "Concerning the Freedom and Limits of the Will," in *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999).

the *freedom* of the will to be boundless, whereas its *power* is (necessarily) constrained. These constraints on the power of willing are just what he understands by volitional necessities.

The claim that wins Frankfurt's approval is Descartes's "breathtaking" contention, as Watson calls it,[35] that there is no occasion whatsoever in which our will can be unfree. Descartes argues that our freedom of willing is as great as can possibly be imagined, "so much so that it is above all in virtue of the will that I [Descartes] understand myself to bear in some way the image and likeness of God."[36] Interestingly, as Frankfurt points out,[37] for Descartes the freedom of the will is also as small as could ever be conceived. The reason for this is that the will is simple and indivisible, consisting as it does merely in assenting or not assenting to the propositions put forward by the intellect. The freedom of assenting or not assenting to these propositions therefore must be an all-or-nothing affair: where it is present nothing can be added to it or subtracted from it; where it is absent, it is absent completely. It is the very same reasoning that enables Frankfurt to maintain that, where it exists, the will is absolutely and perfectly active.[38] It is, however, still possible for Descartes to claim that the unbounded freedom of the will comes with its lack of perfect power. This is precisely the insight that Frankfurt lifts from Descartes in order to explain how persons deliberate practically. He writes: "The grip of volitional necessity may provide, in certain matters, an essential condition of freedom; indeed, it may actually be in itself liberating."[39]

Frankfurt is able to adopt from Descartes the notion of limits to the will's power because he advocates a non-metaphysical interpretation of Descartes's objective in *Meditations on First Philosophy*. Although controversial as a reading of Descartes, this is an interpretation that Frankfurt has favoured since his earliest

---

[35] Watson, "Free Action and Free Will," p. 163.

[36] Descartes, *The Philosophical Writings of Descartes*, Med. IV, p. 40.

[37] Frankfurt, "Concerning the Freedom and Limits of the Will," pp. 75-77.

[38] Ibid., p. 79.

[39] Harry G. Frankfurt, "Preface," in *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999), pp. ix-x.

work.[40] It can be explained by first noting that Descartes uses systematic scepticism to arrive at some ideas that cannot, regardless of our best efforts, be overturned by doubt. Once discovered, these unshakeable convictions can then be regarded as the one secure anchor in light of which our other beliefs can be explained and assessed. The dominant reading of Descartes sees him explaining immunity to doubt with reference to truth: if an idea is so entrenched within our minds that we cannot dissociate ourselves from it, it must be because it captures a slice of reality. The objective of the sceptical method is then what Bernard Williams called an absolute conception of reality, which denotes "knowledge of a reality which exists independently of that knowledge, and indeed (except for the special case where the reality known happens itself to be some psychological item) independently of any thought or experience."[41]

On Frankfurt's non-metaphysical reading, Descartes's aim in the *Meditations* is nothing more than confidence in his properly examined beliefs. In particular, this confidence is not based on a correspondence between his beliefs and the world, but on the absence of valid grounds for doubting his beliefs. For Frankfurt, Descartes is "indifferent to the question of whether the certain corresponds or fails to correspond with the real."[42] Certainty itself is his "fundamental epistemological concept."[43] The criterion of certainty is indubitability, which is to say that Descartes undertakes to accept as certain only those propositions that cannot be reasonably doubted. Although certainty of this sort offers no assurances about objective reality, achieving it is nonetheless useful. It delivers

---

[40] Harry G. Frankfurt, *Demons, Dreamers, and Madmen: The Defense of Reason in Descartes's Meditations*, 1st ed. (Princeton: Princeton University Press, 2008). Some other works that espouse the non-metaphysical interpretation of Descartes are: Louis E. Loeb, "The Priority of Reason in Descartes," *Philosophical Review* 99, no. 1 (1990).; Jonathan Bennett, "Truth and Stability in Descartes' Meditations," *Canadian Journal of Philosophy* Supplementary vol. 16 (1990).; Richard Smyth, "A Metaphysical Reading of the First Meditation," *Philosophical Quarterly* 36, no. 145 (1986).; Louis E. Loeb, "Sextus, Descartes, Hume, and Peirce: On Securing Settled Doxastic States," *Nous* 32, no. 2 (1998).

[41] Bernard Williams, *Descartes: The Project of Pure Enquiry* (London: Routledge, 2005), p. 48.

[42] Frankfurt, *Demons, Dreamers, and Madmen: The Defense of Reason in Descartes's Meditations*, p. 35.

[43] Ibid., p. 36.

the most confidence that humans, as limited in their cognitive powers as they are, can have about their mental states. And this confidence that is appropriate to the human condition attains the aim of the *Meditations*, which is to establish a foundation for the sciences that is "stable and likely to last".[44]

The non-metaphysical reading can be illustrated by considering Descartes's argument for the indubitability of his existence in the *Second Meditation*. Even if all of his perceptions are the work of a deceiving all-powerful demon, writes Descartes, the demon

> will never bring it about that I am nothing so long as I think that I am something. So after considering everything very carefully, I must finally conclude that this proposition, *I am, I exist*, is necessarily true whenever it is put forward by me or conceived in my mind.[45]

Descartes's point can be explained by emphasising that his claim "I exist" is supposed to hold *whenever he considers it*. Whenever it is expressed, argues Descartes, the claim holds. This can be contrasted with the statement "I am in pain." While we have a clear and distinct perception of that statement's truth when expressing it, this indubitability is indexed to the time of our utterance. This just means that there may be occasions when my report of pain will be false, since it is possible for me to be pain-free. The statement "I am in pain" is therefore inadequate as a general claim about me, since it is always vulnerable to reasonable doubt. But my statement that I exist is not inadequate in this way. As Frankfurt notes: "The certainty of beliefs concerning the content of consciousness is, as it were, contingent upon the occurrence of those contents. But the certainty of *sum* is not contingent in this way, since a person can never be aware that he does not exist."[46]

---

[44] Descartes, *The Philosophical Writings of Descartes*, Med. I, p. 12.)

[45] Ibid., Med. II, p. 17.

[46] Frankfurt, *Demons, Dreamers, and Madmen: The Defense of Reason in Descartes's Meditations*, p. 146.

Descartes conclusion is thus, argues Frankfurt, "a kind of necessary statement".[47] Yet it does not express an analytic truth. If it did, its denial would amount to a self-contradiction, which it does not. Moreover, its aim is to advance our knowledge, to prove that something that is provisionally doubted is true, which would not be possible if it were merely a piece of analytic reasoning. What makes it necessary is that upon honest and thorough examination, it cannot be denied. In other words, Frankfurt takes Descartes to have discovered a necessity of the will. This necessity is given by the fact that he cannot withhold assent from the proposition about his existence, that he quite literally cannot bring himself to think otherwise despite that proposition's logical contingency. Such unshakeable conviction leaves "no further questions for us to ask: we have everything that we could reasonably want"; therefore it is "clearly the same as the most perfect certainty."[48] Descartes's contemplation thus warrants his certainty regarding his existence, but it is a certainty that has to do with him, not the way things objectively are.

Frankfurt's notion of volitional necessities is an application to practical philosophy of Descartes's insight about the limited power of the will. One of Frankfurt's examples of volitional necessity concerns Luther's well-known explanation that he persists in his denunciation of the Catholic Church because he "can do no other".[49] Luther's firm conviction does not, suggests Frankfurt, primarily have to do with the strength of his reasons for breaking away from the governing religious institutions. We often find our reasons for performing some act, like donating to charity, quite incontrovertible yet fail to act on them nonetheless. To say that Luther had supremely good reasons for his actions does not properly account for his self-professed inability to do anything different. The only appropriate explanation of this notion of inability, according to Frankfurt, is a literal one. Luther's unwavering commitment reveals nothing more than a psychological fact about him, the fact that he quite simply could not stand the

---

[47] Ibid., p. 147.

[48] Descartes, *The Philosophical Writings of Descartes*, Sec. Replies, p. 103.

[49] Frankfurt, "Concerning the Freedom and Limits of the Will," pp. 80-81.

unrest that came with staying in the Catholic fold. Despite his best efforts, he could not identify with the idea of supporting the Pope, therefore the option of so doing was for him not alive. Consenting to it was for him *unthinkable*.[50] Examples such as Luther's lead Frankfurt to suggest as a constitutive characteristic of our will that it is not able to consent to just any practical proposal of the intellect. He writes: "There is a mode of rationality that pertains to the will itself... Logical necessities define what it is impossible for us to conceive. The necessities of the will concern what we are unable to bring ourselves to do."[51] And the guide to rationality of this sort is discovering what desires it is unthinkable for one to reject. Unthinkability is thus the practical counterpart of Descartes's notion of indubitability.

4.4. Unthinkability as a purely psychological notion

The most direct and far-reaching objection to Frankfurt's conception of volitional necessities would be to contest the dualism of the will and intellect that he inherits from Descartes. As noted, his claims only succeed on the assumption that there is a distinct mental faculty, the will, that sanctions the proposals of the intellect. Locke and Kant are just two of the numerous critics of this reification of the will. Locke's objection is a conceptual one, focusing on the inadequacy of speaking of "the will" as the source of "willing". What we mean by willing, he argues, is an activity conducted by the person as a whole, not any distinct faculty that is supposedly in charge of it – just as we say that a person performs the action of walking, and not his walking faculty. No doubt there is a "power", in Locke's terms, in virtue of which the person is capable of performing this act, but

---

[50] Yet, according to Frankfurt, this lack of power of the will did not entail any diminution of freedom, since the source of it lay nowhere else than inside Luther's own psychological constitution. Cf.: "The grip of volitional necessity may provide, in certain matters, an essential condition of freedom; indeed, it may actually be in itself liberating." (Frankfurt, "Preface," pp. ix-x.)

[51] Harry G. Frankfurt, "Rationality and the Unthinkable," in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), p. 190.

it is not the power that performs it but the person.[52] Kant's objection is more substantive in that he contests the entire metaphysical framework that stands behind the will/intellect dualism. For him there is no role for the will *in* practical reasoning because the will just *is* practical reason.[53]

Objections to the will as a distinct faculty would need to be considered in greater detail for the purpose of criticizing Frankfurt. It is not my aim to do this, but to determine whether Frankfurt's claims about volitional necessities can support the idea that there are limits to what reasons persons can come to have. I wish to suggest that they cannot. For the purpose of making this point it is not necessary to question the Cartesian conception of willing.

I have suggested that Frankfurt intends unthinkability to be understood as the practical analogue of Descartes's indubitability. Both concepts refer to certain constraints on the power of willing, and both are intended as guides to certainty. In Descartes certainty is the reasoned confidence in one's properly examined beliefs, which serves to put the claims of the natural sciences on a secure footing. Frankfurt equally aims at reasoned confidence, although of a different sort. The purpose of looking for desires whose rejection is unthinkable is not to found a science of psychological functioning, but to establish a firm starting point for identification. To recall, Frankfurt holds persons responsible for their actions if and when they act on reasons of their own – even when no alternative to their action was available. To accept (in the non-evaluative sense) a reason or a desire as one's own is to identify with that reason or desire. Now, the problem that unthinkability addresses for Frankfurt is that, ultimately, there must be some raw

---

[52] Cf.: "And so far as any one can, by preferring any Action to its not being, or Rest to any Action, produce that Action or Rest, so far can he do what he will. For such a preferring of Action to its absence, is the *willing* of it..." (John Locke, *An Essay Concerning Human Understanding*, ed. Peter Harold Nidditch (Oxford: Clarendon Press, 1975), II, XXI, 21, p. 244.) And especially: "Concerning a Man's Liberty there yet therefore is raised this farther Question, *Whether a Man be free to will*; which, I think, is what is meant, when it is disputed, Whether the *will* be free." (Locke, *An Essay Concerning Human Understanding*, II, XXI, 22, p. 244-45.)

[53] Immanuel Kant, *Groundwork of the Metaphysic of Morals*, trans. H. J. Paton (London: Routledge, 2005), II: p. 87 (412).

psychic material that persons draw on in deciding what desires to identify with. The brute fact that there are certain things we cannot bring ourselves to do is that ultimate anchor for identification. Limits to what a person can bring himself to do give him "volitional substance" without which "no choice he makes can be regarded as originating in a nature that is genuinely his."[54]

The question about the anchor for identification is important, maintains Frankfurt, because without it persons would be pulled apart by their competing desires. The reflexivity of our consciousness – our ability to *know* that we want certain things, in addition to wanting them – sets us a problem. The variety of desires making claims on our action puts us at risk to "inner fragmentation, dissonance, and disorder".[55] The difficulty must be solved by deciding what desires we can count as genuinely ours, which involves separating those impulses that are genuinely expressive of us from those whose hold on us is usurpatory or compulsive. To be able to make such decisions, argues Frankfurt, persons need confidence in their powers of discrimination. Discovering what desires they cannot shun despite their best efforts gives them confidence of this sort, since it tells them what must be true of them. Identification is thus grounded in acquiescence with how we must be, which "reestablishes the wholeness that was undermined by our elementary constitutive manoeuvres of division and distancing [from our desires]."[56]

Frankfurt's case for practical volitional necessities is strikingly original and commendable for its reliance on the philosophical tradition. However its claim that identities set limits to the individuals' power of willing, and thereby their reasons, is implausible. The point can be explained by noting an important *dis*continuity between indubitability and unthinkability. Descartes's volitional constraints are, in contrast to the ones that exercise Frankfurt, not merely a

---

[54] Frankfurt, "Rationality and the Unthinkable," p. 178.
[55] Frankfurt, *Taking Ourselves Seriously & Getting It Right*, p. 18.
[56] Ibid.

matter of psychological inability. Instead, they are grounded in logic.[57] There is a reason why he cannot withhold assent from the statement that he exists: it is that it is *inconceivable* for a non-existing entity to proclaim his existence. To say, as Frankfurt does in his non-exegetical work, that Descartes bases his anti-sceptical arguments on limitations on the will is hence only partially correct. He bases them on a very particular constraint on the power of willing, the inability of the will to reject propositions mandated by logic. There is, of course, no guarantee that such propositions correspond to objective reality. God and the all-powerful demon can make the inconceivable happen. Nonetheless, since we cannot understand *how* that can be, trusting logic is the reasonable thing to do. There is however no deeper explanation of Frankfurt's unthinkability than psychological inability: "What people cannot help caring about... is not mandated by logic. It is not primarily a constraint upon belief. It is a volitional necessity, which consists *essentially* in a limitation of the will."[58]

One implication of the fact that Frankfurt's unthinkability is a freestanding psychological notion is that the idea of volitional necessities cannot support the claim that there are reasons we cannot commit ourselves to. This is not primarily because that task would require taking on board a contentious conception of the will. Even if this conception is granted, Frankfurt's arguments cannot serve to show that the range of individuals' possible reasons is limited. To begin with, Frankfurt's psychological slant leaves him open to the criticism that he never establishes any real inability of the will. He makes much of Luther's profession that he can do no other than to rebel against the Pope. But since Frankfurt offers no reason *why* it really is inability that Luther runs up against, why should this assertion be taken at face value? It is equally justified to say that Luther finds it very difficult to obey Catholic dictates, not that he finds it impossible to do so.

---

[57] As Frankfurt himself notes, "Descartes requires a foundation that *can never* be subject to doubt. He must have statements for which reasonable grounds for doubt are logically impossible." Frankfurt, *Demons, Dreamers, and Madmen: The Defense of Reason in Descartes's Meditations*, p. 146, original emphasis.)

[58] Frankfurt, *The Reasons of Love*, p. 46, emphasis added.

As Watson notes, the "difference between being unable to bring oneself to act and simply giving up in the face of great difficulty... is obscure."[59]

There is also a deeper reason why unthinkability provides no evidence of constraints on the range of personal reasons. It is that statements about unthinkability express no necessity. This point holds even if one assumes, ignoring the argument I have just presented, that Frankfurt discovers a genuine inability of the will. Even on this generous assumption, the inability to reject certain desires remains only a psychological fact about us. Since there is nothing more to it than a feeling of supreme aversion, it does not prove that we could *never* reject these desires. In this regard, one's report of what is unthinkable for him resembles one's report of pain. The truth of the statement "I am in pain" is indexed to the time of its formulation, since it is possible to conceive of me in a pain-free state. Equally, the truth of my discovery of limits to what I can will is tied to the here and now. I may feel very assured that there are ends I cannot bring myself to pursue. However, if Frankfurt is correct in dismissing human pretensions to access metaphysical truths, this does not mean that I *really* cannot pursue them. More modestly, my discovery of volitional limits also cannot ground my reasoned confidence; in other words, I cannot take them as the secure anchor for identification. To recall, Descartes arrives at reasoned confidence about his beliefs by learning that there cannot be a proclamation of existence without there being someone to formulate it. Attaining confidence about one's desires would require a parallel argument, which is missing in Frankfurt. This would be an argument showing that a person cannot express any desire without acknowledging some particular limits to what he is able to desire. But that argument would rely on logic, which Frankfurt expressly rejects as a ground of practical normativity.

Frankfurt's complex and sophisticated account of action turns out to be incapable of bearing out the claim that there are constraints to what persons may come to will. Some of my remarks about it may also raise doubts about the plausibility of

---

[59] Gary Watson, "Volitional Necessities," in *The Contours of Agency*, ed. Lee Overton and Sarah Buss (Cambridge, MA: MIT Press, 2002), pp. 134.

Frankfurt's wider account of practical reasoning. They may suggest a mistake in thinking that finding one's volitional constraints puts an end to the battle of one's desires. But the plausibility of this suggestion cannot be explored here. T he important point here is that psychological inability is too weak a notion to ground a person's reasons. The claim that one has a reason to refrain from doing what he is unable to do might be plausible if it referred to necessary inability. But there is no necessity to be found in merely psychological facts about persons: a course of action that looks quite unthinkable at one point can, at another time, be a realistic option. Therefore, the connection that Frankfurt draws between identities and reasons is implausible.

My claim that statements about what we can bring ourselves to do express no necessity is inspired by one of the central theses in Kant's (as well as Hume's) philosophy: that experience, whether in the theoretical or the practical domain, cannot provide any insight into how things must be. In the next chapter I will examine Korsgaard's approach to identity, which ultimately aims to disprove this thesis but nonetheless claims Kantian ancestry. On the one hand, she construes identity as a psychological fact about an agent but, on the other hand, she also thinks that there is and identity that all persons must have: that of a Citizen of the Kingdom of Ends. Whether Korsgaard's derivation of necessity from the basis of empirical facts can, in the end, succeed is an interesting question, but I will not discuss it. My focus will be on the preliminary part of Korsgaard's argument – her account of identity as a person's evaluative self-description that is the "root" of a person's reasons. Although this account serves only a preparatory function in Korsgaard's overall project, it is nonetheless attractive in its own right. If it is plausible, it provides resources for an agreeable revision of Kant's purely formal – and supposedly empty – conception of practical reasoning.

5. Identity as the criterion of maxim-adoption

5.1. The normative problem and reflective endorsement

Korsgaard's concept of practical identity emerges as a part of her answer to the problem of justifying the normativity of morality. Moral concepts are normative in that they make demands on us: they prescribe what type of conduct is morally correct and ask us to behave accordingly. Korsgaard writes:

> [Ethical standards] do not merely *describe* a way in which we in fact regulate our conduct. They make *claims* on us; they command, oblige, recommend, or guide. Or, at least, when we invoke them, we make claims on one another. When I say that an action is right I am saying that you ought to *do* it; when I say that something is good I am recommending it as worthy of your choice... And it is the force of these normative claims – the right of these concepts to give laws to us – that we want to understand.[1]

The problem with the normativity of moral requirements, according to Korsgaard, is that it is possible to regard injunctions to keep one's promises, repay one's debts, care for one's parents in older age and so on, as nothing more than arbitrary dictates. The philosophical foundation of morality that Korsgaard seeks is a defence of the right of these injunctions to govern our lives. Her focus thus considerably differs from Plato's and Hobbes's, who aim to ward off the radically disaffected sceptic – one who cannot even hear the voice of morality,

---

[1] Korsgaard, *The Sources of Normativity*, pp. 8-9, original emphases.

thinking it is self-interest in disguise speaking.[2] Korsgaard seeks the conditions under which moral claims can be justified to the one who recognizes certain prescriptions as distinctly moral, but does not see them as normative.

Yet these statements only roughly reveal the nature of Korsgaard's approach to the problem with the normativity of morality. More precisely, the challenge that she confronts is that ethical considerations may fail to constitute reasons for action.[3] They may be requirements that persons meant to be governed by them have no reason to follow. It is this claim that leads Korsgaard to develop a general theory of reasons. The logic behind her foray into an examination of reasons can be presented as follows: Korsgaard's principal interest is in the question of whether there are any moral requirements that are normative for us. However, she also claims that for a moral demand upon some person to be valid, that person must have a reason to follow it. In order to determine whether any moral demands are valid, it must therefore be seen what would make them reason-giving. And to answer that question, it must first be explained what makes *any* consideration, moral or non-moral, reason-giving. Once the general conditions for the existence of reasons are clarified, it will then be possible to return to the issue of what moral norms can be supported by reasons.

One condition in particular is at the centre of Korsgaard's discussion, which is that no consideration can constitute a reason without being capable of *addressing* the person in question.[4] The basic idea here is that if a normative claim is to have any hold over a person it must be connected to him, the agent deliberating from the first-person perspective, in some special way. It must be directed at him, so that he feels that its command is issued to him and not anyone else, or no-one in particular. The requirement of address is therefore that for an injunction (of any sort) to be a reason for someone, that person must regard himself as a subject to

---

[2] Plato, *The Republic*, trans. Desmond Lee, 2nd revised ed. (Harmondsworth: Penguin, 1974), p. 47.; Thomas Hobbes, *Leviathan*, ed. Richard Tuck, student revised ed. (Cambridge: Cambridge University Press, 1996), especially pp. 86-129.

[3] Korsgaard, *The Sources of Normativity*, pp. 9-10.

[4] Ibid., pp. 16-17.

whom that injunction applies. The importance of that requirement should however not be overstated: it expresses a condition of the possibility of reasons, but not a sufficient condition for the existence of reasons. Even if some normative demand adequately addresses a person, the question still remains whether that person should obey it. Nonetheless, Korsgaard's point is that without this address, the question of the validity of the demand cannot even be raised. This is similar to the way in which the fact of political membership renders citizens liable to certain obligations, such as paying taxes, but is not enough to justify those obligations. Citizens may, for example, have good reason to challenge their obligation to fund their state's aggressive war. Yet they are only in a position to do so because they regard themselves as subjects of that state.

The requirement of address plays a central role in Korsgaard's assessment of the traditional approaches to normativity. Her objection to the meta-ethical theories of voluntarism and realism is precisely that they fail to yield moral norms that can address persons. The voluntarist position traces the validity of norms to the legislative authority of their maker: the fact that they issue from some suitable legislator's will is sufficient to justify the demands they make upon persons. In religious accounts, for instance, the place of the authoritative will is often taken by God. However, argues Korsgaard, the voluntarist position begs the question of why agents should regard themselves as the addressees of the legislator's commands. Even if a certain rule is God-made it does not automatically follow that we should obey it: we still need an explanation of why divine rules should concern *us*. The realist view in ethics, on the other hand, seems to leave implausibly little room for practical judgement, with its portrayal of deliberation as an exercise of discerning and comparing facts. The curious nature of inherently normative facts − for example, the fact of the wrongness of murder − is also sufficient to raise suspicion about ethical realism.[5] However Korsgaard's principal objection to the realist view, and one that stamps a strong mark on her

---

[5] It is however questionable whether the realist model can be outright refuted on account of its adherence to "queer" facts. For a famous argument that it can see J. L. Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1990).

project, is the same one that applies to voluntarism. Although the realist holds off the sceptical charge by positing a class of considerations that are normative in themselves, and thus not susceptible to further questioning, it lacks an explanation of how and why facts engage persons. Facts cannot determine agents' actions directly but only insofar as agents take them as normatively important. The claim that murder is intrinsically wrong then does not settle the issue of individuals' reasons. For that to happen, those individuals still need to view themselves as subjects who take the wrongness of murder as normatively important.[6]

Korsgaard's treatment of the "reflective endorsement" approach to normativity, espoused by David Hume, John Stuart Mill, and Bernard Williams, merits a more detailed discussion here. That approach seems well-suited to answer the question of how considerations address persons, due to its insight that "normativity is a problem for human beings because of our reflective nature".[7] By "reflexivity" Korsgaard means persons' ability to question the validity of their own beliefs and motives. According to Korsgaard, realism and voluntarism fail to account for the fact of reflexivity at the very basic level, since they are not at all concerned with the subjective experience of moral requirements. For both approaches, how persons regard normative claims has nothing to do with the validity of these claims. The commands of an authoritative lawgiver are thus supposed to be normative regardless of whether and how individuals perceive them. Equally, intrinsically normative facts are meant to issue reasons irrespective of whether we are aware of them. This neglect of reflexivity results in failure to address: it is difficult to see how norms that are wholly extraneous to the agent's consciousness can be accepted by that agent as directed at *him*. Reflective endorsement theorists deal with this issue rather better. For them, the principal

---

[6] Korsgaard, *The Sources of Normativity*, pp. 37-40. A position which escapes Korsgaard's criticism is *procedural* moral realism – the belief that there are correct and incorrect moral views, where the standard for making this judgement is not conformity with any objectively existing normative facts in the world but adherence to a correct procedure for arriving at moral views. Korsgaard herself ends up as a procedural moral realist, although with incorporated elements from several of the moral doctrines she criticizes.

[7] Ibid., p. 49.

question that an account of normativity must answer is the first-personal one – whether moral demands can survive the agents' own critical scrutiny. However Korsgaard claims that the reflective endorsement explanation of normativity also founders on the requirement of address. Although this claim is very important for understanding her ideas about practical identity as the source of reasons, her arguments in support of it are complex and often not very perspicacious. It is therefore useful to examine them more closely. Korsgaard's treatment of the reflective endorsement theory can be illustrated by looking at her appraisal of Hume, and especially Hume's discussion of justice.

The pivotal aspect of Hume's conception of justice is that justice is an artificial virtue, and as such opposed to natural virtues such as benevolence or compassion. The motives that guide us in beneficent or compassionate actions are natural inclinations and desires, those that belong to unadulterated human constitution as we would find it in the crude condition prior to society and political authority. These "passions" are the motives that we need not combat, reflect upon, or obey – they are the spontaneous workings of the animalist part of our personality. [8] Controversially, Hume thinks that passions are the only possible source of motivation. There can be no human action that is not, at some point in the chain of practical reasoning, traceable to original passions. So when we contemplate an action that can produce pain or pleasure, we are guided by an emotion of aversion or propensity towards the prospective outcome. [9] Since for

---

[8] A passion is an original existence, which within Hume's empiricist framework means that it is not a representation of anything outside itself. For example, the idea of anger is not a mental copy of any object or another idea; to be angry is just to be, irreducibly, in a certain state: "possessed" by anger. (David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, 1960), II.iii.3, p. 415.)

[9] "'Tis obvious, that when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, and are carry'd to avoid or embrace what will give us this uneasiness or satisfaction. 'Tis also obvious, that this emotion rests not here, but making us cast our view on every side, comprehends whatever objects are connected with its original one by the relation of cause and effect. Here then reasoning takes place to discover this relation ; and according as our reasoning varies, our actions receive a subsequent variation. But 'tis evident in this case, that the impulse arises not from reason, but is only directed by it." (Ibid., II.iii.3, p. 414.)

Hume reason cannot alone produce passions but consists merely in the discovery of relations between ideas, it cannot be sufficient to determine action.

However, our motives for adhering to norms of justice – principally, the laws of the contract – are not directly natural. Our reasons for not stealing and not defaulting on mortgage payments do not derive from any of the original pre-social passions and inclinations. This point means, on the one hand, that some requirements of justice cannot be explained in terms of natural motives. In many cases the requirements of justice do not perfectly align with our natural sentiments, such as compassion and beneficence. For example, one is required to pay back a debt even if the lender is wastefully rich and repulsively arrogant. Moreover, what justice demands can sometimes go against our natural motives. For instance, respecting others' property sometimes obviously goes against our selfish desires, such as to live a comfortable existence without being accountable to others. Therefore the motive to act justly must have its source in something other than natural inclinations. Hume locates this source in the "sympathy with public interest",[10] which on Korsgaard's construal amounts to something like enlightened self-interest. Individuals have an incentive to obey the system of justice even when its particular edicts pain them because the existence of the system is justifiable in view of their long-term well-being. To begin with, historically the system came into existence with the purpose of facilitating economic activity and advancing well-being, which provides an incentive for upholding it. Moreover, the system also fosters pleasures of its own – for instance, the agreeable feeling that comes with being recognized as the law-abiding member of the society, available only through observance of received norms. On Hume's account, the motive to be just is thus rooted in self-interest; as such it is not reducible to motivating passions, although it retains a connection with them.[11]

---

[10] Ibid., III.ii.2, pp. 498-501.

[11] Cf.: "Thus self-interest is the original motive to the establishment of justice: but a sympathy with public interest is the source of the moral approbation, which attends that virtue. (Ibid., III.ii.2, pp. 499-500.)

Agents in the Humean scheme acknowledge the authority of demands placed on them by the impartial system of justice, and act on them for the non-natural motive of sympathy with public interest. However, situations may occur in which the motive to be just comes into conflict with residual natural inclinations: this is the source of the "sensible knave" problem.[12] The sensible knave has a certain distance from his reasons insofar as he feels the impartial force of particular moral injunctions, but also recognizes the true basis of morality in self-interest.[13] This enlightened stance presents him with a deliberative difficulty when he needs to endorse or reject a course of action which would promote his own interests without threatening to bring down the system of justice. Suppose that he is in a panic rush to catch an airplane and is tempted to get his newspaper from the self-service stand without stopping to leave the change. He is well aware that the action would be wrong from the moral point of view, but he also knows that the ultimate justification of the moral viewpoint is that it promotes the satisfaction of natural desires. Furthermore, it is certain that his offence will pass unnoticed – which means that the person will not have to suffer the disapprobating attitude of others – and, being minor, will not shake the foundations of moral norms and habits. The question is whether the knave might think, in light of all this, that taking off without paying is justified.

---

[12] The sensible knave illustration appears in David Hume, *An Enquiry Concerning the Principles of Morals*, ed. Lewis Amherst Selby-Bigge and P. H. Nidditch, 3rd ed. (Oxford: Clarendon Press, 1997), IX.ii, pp. 282-84.

[13] This is a peculiarly Korsgaardian reading of the sensible knave problem, which relies on an individualist view of Hume's self. Since my interest is in how Korsgaard develops the idea of personal identity, I am here not questioning whether that reading of Hume is correct. However, there is room for doubting the plausibility of Korsgaard's individualist interpretation of Hume's self. A number of authors have argued that, for Hume, sympathy for others is a constitutive feature of persons, rather than just a constraint on their narrowly individualist pursuits. See for example Amelie Oksenberg Rorty, "'Pride Produces the Idea of Self': Hume on Moral Agency," *Australasian Journal of Philosophy* 68, no. 3 (1990).; Annette Baier, *A Progress of Sentiments: Reflections on Hume's Treatise* (London: Harvard University Press, 1991), chap. 6.; Pauline Chazan, *The Moral Self* (London: Routledge, 1998), chap. 1.; Christopher J. Finlay, *Hume's Social Philosophy: Human Nature and Commercial Sociability in a Treatise of Human Nature* (London: Continuum, 2007).

The aspect of Hume's theory that interests Korsgaard is not so much that it leaves open the possibility for the knave to decide that stealing is justified. She focuses rather on the knave's resources for deciding what to do. Hume's ambition, as we saw, is to portray the motive to act justly as a non-natural motive: it is an inclination that arises from one's consideration of the contribution of the system of justice to one's long-term wellbeing. And yet the knave in our example cannot be guided by such an inclination because he realizes that his wellbeing will not suffer as a result of his knavishness. No-one will see him steal the paper, and social life will go on as usual. Self-interest, which Hume regards as the proper, non-natural source of moral obligations, thus has no bearing on his decision. The knave will therefore have to arrive at his decision, whether it is to steal or to pay, by consulting some consideration other than his enlightened self-interest. One alternative is that he will decide by considering the utility of the particular act of taking the paper without leaving the change. However this response would attribute to Hume the kind of act-utilitarianism that is incompatible with his insistence on the benefits of the *system* of justice, and not of particular just acts. Another response, which is the one that Hume seems to take, is that the knave will make up his mind by choosing in line with his deepest dispositions of character. If he is a virtuous man he will be generally disposed to act justly, in which case he will also be disposed to do so in this particular situation. If, on the other hand, he lacks the virtue that results from proper social habituation, he will act on the desire for maximum comfort with minimum hassle.

Korsgaard's objection to Hume's virtue-centred response to the knave problem can be approached by saying that it arbitrarily stops the project of accounting for normativity in terms of the reflexivity of human consciousness. Its failure is that it does not explain why the knave ought to act in one way or another in terms of his assessment of his own motives. As such, it does not "push reflection as far as it will go".[14] The knave will not determine what he should do by critically examining himself but by surrendering to aspects of his psychic life that are independent of his reflection: desires and inculcated dispositions. If he is

---

[14] Korsgaard, *The Sources of Normativity*, p. 89.

virtuous, the knave will not need reasoning to confirm that he ought to do what his good habits incline him to. If he lacks moral virtue, critical reasoning will again play no part in his decision-making: on the one hand, his reasoning cannot bring him to want to act justly; proper habituation is required for that. On the other hand, if he has a natural desire to act unjustly, he will not need confirmation from reason to know that he should satisfy it (since the system of justice is not in danger). Against its proclaimed intentions, Hume's reflective endorsement model thus bases normativity on self-discovery rather than self-appraisal. However, writes Korsgaard, "If the reflective endorsement of our dispositions is what establishes the normativity of those dispositions, then what we need in order to establish the normativity of our more particular motives and inclinations is the reflective endorsement of those."[15]

In *The Sources of Normativity* Korsgaard's objection to Hume stops there. There she seems content to maintain only that Hume cannot show how normative claims can survive agents' own critical reflection all the way through. This critique remains squarely within the confines of Hume's philosophical framework: the only trouble with the reflective endorsement model is that it is not fully consistent. However it is misleading of Korsgaard to limit herself to an internal critique of that model, as it misrepresents the extent of her disagreement with Hume, Williams, and Mill. Her other works provide ample resources for developing an objection to the reflective endorsement model that is both stronger and in line with her assessment of other meta-ethical positions. That deeper objection is, once again, that Hume's conception does not explain how normative claims address persons. At first blush, this claim seems out of place. One of the strengths of Hume's approach is that it offers an explanation of how normative claims relate to persons: they do so by engaging their passions. It is this involvement of passions that sparks the activity of critical reflection: the person first finds himself naturally drawn towards certain courses of action, but cannot escape the question of whether his inclinations are worth satisfying. A successful normative claim will just be that call of his passions that he can accept upon reflection. Since the involvement of agents' passions is what initially triggers

---

[15] Ibid.

practical reasoning, any normative claims that a person considers are always intimately related to him.

However, Korsgaard has argued that this view of the relation between persons and normative claims is unsatisfactory. It explains how a person can be motivated to comply with normative demands, but a theory of normativity must explain something else. It must answer how it is that a person can be motivated to comply with normative demands *in virtue of* their normativity.[16] This point can be approached by noting that the fact that a normative demand moves a person to act cannot be taken as proof that the demand truly engaged the person. There is still the possibility that the motivation was unrelated to the demand itself. For example, one can behave morally not because it is the right thing but because it pleases him. And then it will not be correct to say that that person was moved by his obligation – the thought that there is something that he morally ought to do – but by the anticipation of pleasure. It just so happens that he is motivated to obey the moral requirement, perhaps because he needs a boost to his own self-image, or for some other reason. But there is nothing to say that he will always and necessarily be motivated to obey it – for instance when he regards himself in a more positive light. His motivation is thus only contingently associated with the moral requirement, dependent not on the content of that requirement but on factors extraneous to it. The trouble with Hume's account is that it leaves open the possibility of contingent motivation to follow normative demands. If so, it also allows that persons who are moved to comply with certain requirements may in fact not be addressed by those requirements. For all that Hume says, their passions may be incited by something else entirely. Explaining

---

[16] Korsgaard thus writes that in having reasons persons are "inspired to do things by the normativity of the reasons they have for doing them, by their awareness that some consideration makes a claim on them." (Christine M. Korsgaard, "Acting for a Reason," in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford: Oxford University Press, 2008), p. 209.) And further: "to have a reason is to be motivated by the consciousness of the appropriateness of your own motivation." (Korsgaard, "Acting for a Reason," p. 215.) Reasons are not only motivating but "normatively motivating".

motivation is thus of little help in understanding how normative claims address persons.[17]

Hume does not only leave open the possibility of contingent motivation to follow normative demands. He also seems committed to saying that in some cases, including the sensible knave's, persons *must* be so extrinsically motivated.[18] If Hume's knave decides to do what justice requires, it will not be because it is what justice requires. To be sure, the requirement of justice elicits a certain response from him, by connecting with one of his dispositions of character. And yet that connection does not amount to an address. A person that is properly addressed by a normative claim understands that the claim is directed at him – that he, and not anyone else or no-one in particular, should do what that norm asks of him. He regards himself as subject to its demands. However the reflection

---

[17] This line of reasoning can be regarded as an extension of Korsgaard's earlier critique of Williams's "sub-Humean" model of practical reasoning, especially the "internalism requirement" that she attributes to Williams. That requirement states that "Practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating rational persons." (Christine M. Korsgaard, "Skepticism About Practical Reason," *Journal of Philosophy* 83, no. 1 (1986): p. 11.) Williams argues that this principle can serve as a grounding for a theory of practical reasons, and especially to rule out those conceptions of reasons, such as the Kantian, that fail to honour it. The requirement is meant to place "independent constraints, based solely on motivational considerations, on what might count as a principle of practical reason". (Korsgaard, "Skepticism About Practical Reason," p. 8.) However Korsgaard maintains that the internalism requirement cannot play this critical function, as it assumes a separation between the motivational and the normative aspects of a reason. Defending this separation in turn requires, she suggests, a more substantial backing than Williams's thin "motivational scepticism" about practical reason. It requires one to endorse a particular view of what reason it and how it operates – as, for example, Hume himself does. The plausibility of the internalism requirement will then depend on the plausibility of the underlying understanding of what reason is. Korsgaard's aforementioned two objections to Hume – that his conception is inconsistent, and that it fails to explain how normative claims address persons – can be seen as completing her critique of the internalism requirement. They reveal that the requirement cannot be salvaged because Hume's thick understanding of reason that supports it is implausible.

[18] If Hume is indeed committed to this claim he qualifies as a motivational externalist, meaning that he accepts that an explanation of why some consideration moves a person to act can be distinct from an explanation of why that person ought to act on that consideration.

of the would-be knave leads him to quit regarding himself in this way: once he realizes that the source of the validity of moral obligations is self-interest, and that his self-interest will not be harmed by defaulting on payment, the requirement to pay for bought goods stops speaking to him. With this new knowledge, that requirement does not concern him anymore. It will concern those who think that *all* moral norms are *always* in one's self-interest. But he is no longer one of those persons. Hume's account thus ultimately does not explain what it takes for a person to be addressed by a normative claim. When pushed to its limits, it states that persons can sometimes be *un*addressed by normative demands but be motivated to act on them all the same. However, explaining how they can be so motivated is irrelevant to the question of address: what is important, this explanation leaves out.

5.2. Freedom and maxims

Since Korsgaard's argues (or might be taken to argue) that the traditional conceptions of normativity founder on the requirement of address, her own conception is built around it. Its seeks to explain normativity starting from the question of what it takes for persons to regard themselves as subjects to normative demands. Her answer, as will be seen, is that in order to properly address a person a norm must appeal to his practical self-conception, or identity. That conception borrows aspects of Kant, but also contains some important departures from Kant's ethical theory.

Korsgaard makes a turn towards Kant by reframing the issue of reflexive endorsement in terms of human freedom.[19] In order to explain her strategy it is useful to briefly retrace her discussion. She argues that persons confront a normative problem insofar as they are able to question the validity of any normative claim that they experience. The reflective endorsement approach seeks to answer how persons can deem any normative claim justified by consulting their first-person perspective – by asking which of the claims they experience

---

[19] Korsgaard, *The Sources of Normativity*, pp. 94-97.

they can regard as justified. Korsgaard accepts this recourse to the reflexivity of human consciousness. However she also maintains, unlike Hume for example, that reflexivity raises the issue of freedom. In making this claim she, first, follows Hume in maintaining that the initial candidates for reasons are desires and inclinations. Once we reflect upon our own mental states we find that we are already, prior to reflection, drawn to various courses of action. Korsgaard writes:

> I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have a problem. Shall I act? Is this desire really a *reason* to act?[20]

It is at the point of deciding whether we can accept our desires as reasons that, for Korsgaard, the issue of freedom arises. What enables Korsgaard to make this claim is her endorsement of a particular view of desires, and a particular view of freedom. She conceives of desires as essentially foreign to the mind that contemplates them — as "biddings from outside" the consciousness.[21] Korsgaard's guiding idea here is that all desires are affections; as Kant puts it, they are modifications of the "receptivity belonging to inner sense".[22] In other words, we are "struck" by desires and play no part in their production. It is due to their alien nature, and to their unmediated grip on our psyche, that desires may turn out to be invasive: unless the person makes a conscious intervention into their hold on him, they will determine his behaviour without his active participation. The person would then be prevented from acting freely by being prevented from acting at all. If he surrenders to his desires, the movements of his

---

[20] Ibid., p. 93, original emphasis.

[21] Ibid., p. 94. Cf. also: "Anything outside of the will counts as an alien cause, including the desires and inclinations of the person." (Korsgaard, *The Sources of Normativity*, p. 97.)

[22] Immanuel Kant, *Critique of Practical Reason*, trans. Mary J. Gregor (Cambridge: Cambridge University Press, 1997), 5: 58, p. 51. I will return, in chapters 6 and 7, to the issue of whether Kant can indeed be taken to be claiming that all desires are affections.

body will not have their source in him, but in overpowering impulses that are foreign to him. The things he does will then just be things that *happen* to him.[23]

Although the potentially invasive force of desires present a real threat to human agency, and thereby human freedom, Korsgaard argues that they nonetheless do not rule out the possibility of free agency. This claim is part of her broader compatibilist argument against the idea that external determination of human behaviour annuls freedom. On her view, it is possible for individuals' choices to be conditioned by factors outside those individuals' control — such as evolutionary natural selection, an all-powerful God, or incursive desires — and still be free. She presents her argument for this view as Kantian, which is reinforced by her insistence on approaching the problem of freedom in practical rather than theoretical terms. The problem, she says, is not to establish from a speculative point of view whether we are free. What the speculative approach misses out on is an essential feature of human life: that the question about freedom arises for persons looking to decide how to conduct themselves. This is not to say that one cannot philosophize about freedom from outside the first-person practical perspective; only that a theory of freedom must take account of the human practical interest in deciding what to do.[24] With this change of viewpoint, Korsgaard suggests, the question about freedom becomes the question of whether we, as persons who must decide how to act, can regard ourselves as free.

Her answer is that, as practical agents, we must view ourselves as free, which means that we must also view ourselves as free from invasive desires. This statement is supported by a claim about the ineliminability of reflexive consciousness. Even if all our choices were somehow determined from outside

---

[23] Christine M. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press, 2009), p. 18.

[24] Cf.: "But that doesn't mean that I am claiming that our experience of our freedom is scientifically inexplicable. I am claiming that it is to be explained in terms of the structure of reflective consciousness, not as the (possibly delusory) *perception* of a theoretical or metaphysical property of the self." (Korsgaard, *The Sources of Normativity*, pp. 96-97, original emphasis.)

our consciousness, Korsgaard suggests, this would not remove our need to make up our minds on what to do. The "simple inexorable fact of the human condition" would remain: we would still need to choose and act.[25] This means that we cannot imagine ourselves fully invaded by desires, because although desires affect us from the outside, they cannot conquer us without our being conscious of it. Once we do become conscious of their hold on us, we cannot sink back into the position of the unreflective spectator. Even if impulses come to dominate us, the reflective structure of our consciousness still gives us a choice – whether to accept their rule over us, or not. One of these options asserts our subjectivity, while the other resigns us to passive submission. However, and this is Korsgaard's point, this submission is not complete because we still get to choose it. Whatever it is that we consciously decide, argues Korsgaard, we exercise our agency in the act of deciding. This has an implication for her view about freedom. We are agents because we are reflexively conscious and reflexive consciousness is tied up with the necessity of acting. However it is impossible to be an agent without thinking of oneself as a free agent. The conception of freedom that underlies this Korsgaard's claim is one in which freedom is the absence of external determinants of one's practical deliberation. Once freedom and agency are understood in the way that Korsgaard proposes, it is clear that we cannot simultaneously be agents and unfree. To act just *is* to act freely: it is to assert oneself over and above foreign influences on the mind, by choosing whether to follow or reject them. Since this choosing is something that we must do, we must also be free.

Korsgaard's view that all persons are free in the described sense shapes her theory of normativity by steering her towards a particular understanding of normative claims. For ethical realists, the considerations that are meant to possess normativity are facts, or states of affairs; for voluntarists, these considerations are commands of some external lawgiver, like God; for reflective endorsement theorists, they are one's endorsed desires. For Korsgaard, however, the considerations that enter practical deliberation are *maxims*, or "subjective

---

[25] Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*, pp. 1-2.

principles of action".[26] Maxims can be understood as first-person, law-like precepts, such as "I will work extra hours to get a promotion" or "I will drink less coffee". Their content is universal, in the sense that they do not recommend some particular action (like "working late *today*") as worth doing. Rather, they pick out a class of actions ("working late *generally*") as normative. Particular actions are then justified for an agent by their falling under the class of actions for which he has adopted a maxim. Within Korsgaard's picture of practical reason all action is principled in this way. On the face of it, this claim sounds highly counter-intuitive as a description of the phenomenology of decision-making.[27] Nonetheless, Korsgaard maintains that we are required to accept it if we are committed to regarding persons as free. Only if all action stems from maxims is there room for freedom of the will. It is the freedom that comes with our ability to resist our natural impulses by bringing them under a principle. This notion of self-legislation allows us to see how it is possible for agents to have reflective distance from their motives: desires, or passions, or impulses can never provide sufficient ground for acting. They can only be *incorporated* into maxims; but by then "although we may do what desire bids us, we do it freely".[28]

5.3. Identities as grounds for maxim-adoption

Understanding practical reasoning as a matter of considering and adopting maxims leaves room for persons' freedom from their desires. However it also reactivates the problem of address. If normative claims are maxims, in virtue of

---

[26] Kant, *Groundwork of the Metaphysic of Morals*, II: 51 (421).

[27] However, the claim is not without contemporary advocates. For Scanlon, for example, maxims provide the general framework that enables persons to figure out which considerations are relevant to a given decision. They are "principles specifying the adequacy or inadequacy of various considerations as reasons for one or another judgment-sensitive attitude" (Scanlon, *What We Owe to Each Other*, p. 53.)

[28] Korsgaard, *The Sources of Normativity*, p. 94. This view of the relationship between desires and maxims – commonly known as the "Incorporation Thesis" – is widely cited in discussion of Kantian practical reason. Henry Allison has called it "the centrepiece of Kant's conception of rational agency". (Henry E. Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1991), p. 189.)

what do maxims appeal to persons? What makes this question difficult is the assumption of individuals' perfect detachment from their attachments and projects. Free agents can only be guided by some attachment or project if and when they choose to endorse a maxim that incorporates it. In this sense, they are contingently and not constitutively tied to all of their concerns. But if persons stand so aloof from all of their attachments, what grounds do they have for choosing *any* maxim? Stated in terms of the requirement of address, the question is: in virtue of what can any maxim, enjoining the person to perform or refrain from specific actions, address a free chooser? The solution, according to Korsgaard, is to be found in the standpoint from which we deliberate practically, the standpoint of our *identities*.

Korsgaard's treatment of identity can be introduced by considering how the idea of identity emerges in the course of Korsgaard's response to Hegel's "empty formalism" objection to Kant.[29] In short, that objection states that moral prescriptions cannot be derived from a justificatory procedure that fully abstracts from the content of moral reasoning. Purely formal principles of morality are indeterminate, in that they do not elect any actions as morally correct. Korsgaard approaches this objection by focusing on the categorical imperative, a normative

---

[29] See Georg Wilhelm Friedrich Hegel, *Elements of the Philosophy of Right*, ed. Allen W. Wood, trans. Hugh Barr Nisbet (Cambridge: Cambridge University Press, 1991), pp. 162-63. The best-known contemporary version of the objection is Rawls's criticism that Kant's "Categorical Imperative procedure" cannot yield determinate assessments of the moral probity of maxims. (John Rawls, *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, MA: Harvard University Press, 2000), pp. 172-75.) On Rawls's reading of Kant, whether a maxim is to be regarded as right or wrong is answered by considering the hypothetical world in which each person acts on that same maxim. If this "corrected social world", governed by one's initially particular rationale as a universal rule for all, can be willed consistently then this rationale is morally appropriate; if not, it is morally improper. In the *Groundwork* Kant thus claims that that hypothetical social world which has no place for acts of compassion and mutual assistance cannot be willed because unsympathetic acts of others will sometimes go against what Kant calls our self-love. However, Rawls thinks that the CI-procedure fails to authorize not only the world which universalizes acts of mutual disinterest, but also the world in which sympathetic acts are the norm. On his view, the procedure lacks a substantial standard that would sanction *any* course of action.

requirement that, following Kant, she regards as entailed by freedom. On Korsgaard's interpretation, that requirement states that if we view ourselves as free creators of maxims, we must accept that we ought to only create maxims that are universalizable. Both Kant's argument for this claim and Korsgaard's rendition of that argument are complicated and controversial, but they need not be discussed here. What is important is that Korsgaard accepts the categorical imperative as "the law of a free will... It describes what a free will must do in order to be what it is. It must choose a maxim it can regard as a law."[30] This demand is purely formal as it imposes no limits on *what* we must choose in order to be free, only on *how* we must do the choosing. Nonetheless, Korsgaard also admits that the categorical imperative is too abstract to serve as a guide for determining the moral probity of maxims. Its flaw is that it "doesn't settle the question of the *domain* over which the law of the free will must range."[31] Persons may adopt all sorts of maxims that can be both conceived and consistently willed as universal laws, but that are nonetheless incompatible with morality. For instance, they may undertake to act on the principle of acting on whatever desire has the strongest grip on them at the time of decision-making. Their behaviour will then hardly be labelled as moral, especially when they happen to be gripped by jealousy, or lust, or hatred.

For Korsgaard the categorical imperative, with its absence of restrictions on the content of maxims, is insufficient as a test of morality. The objective of her revision of Kant is, first, to import some such substantive constraints into the categorical imperative. Second, she aims to do so without compromising the idea that agents' choosing must be regarded as free from external determinants, including desires. This latter qualification rules out one quick and easy fix for the indeterminacy problem. That fix would be to simply add a reference to some determinate desires or projects to the universalizability requirement: in other words, to say that maxims must not only be capable o f being laws but also

---

[30] Korsgaard, *The Sources of Normativity*, p. 98.

[31] Ibid., p. 99, original emphasis.

further some specific interests, or exclude some others.[32] For instance, one could stipulate that universalizable maxims must not promote only the agent's own well-being. However, for Korsgaard such a solution is unacceptable because it clashes with the assumption of agent's freedom. If persons are free, then their choosing cannot be constrained by external factors – including the categorical imperative. If there are certain maxims that they must not have, that just means they are *not* free after all: they are not free to choose those maxims.

Korsgaard's response to the indeterminacy problem is to supplement the categorical imperative with another requirement, what she calls the "moral law".[33] Since the moral law is substantive, it can remedy the empty formalism of the categorical imperative. The moral law states that in order to qualify as moral, maxims must be such that all rational beings (in what Kant calls the "Kingdom of Ends") could agree to be governed by them. Unlike the categorical imperative, the moral law has content: it is "a substantive command as long as we have *some*

---

[32] This is Rawls's strategy of adding substantive constraints to the categorical imperative. He writes: "First, we must give more content to the will of ideal agents in deciding whether they can will an adjusted social world... Second, we must specify further the point of view from which these decisions about social worlds are made..." (Rawls, *Lectures on the History of Moral Philosophy*, p. 173.) The idea that accomplishes both tasks, and for which Rawls claims support from Kant's *Metaphysics of Morals*, is that of "true human needs". The notion is not elaborated very thoroughly but it does state that the fulfilment of certain requisite conditions is necessary if persons are to lead fully satisfactory lives. True human needs constitute the standpoint from which the validity of maxims can be assessed, insofar as these maxims further or frustrate their satisfaction. (Rawls, *Lectures on the History of Moral Philosophy*, p. 174.)

[33] Kant does not distinguish between the categorical imperative and the moral law in the way that Korsgaard does. For him, the two principles that Korsgaard discusses are two formulations of the same principle that governs all sensible beings possessed of reason. It is that underlying principle that Kant calls the categorical imperative, rather than just the requirement of universalizability. Because it governs *sensible* beings possessed of reason, whose sensible nature may pull them away from obeying reason itself, the categorical imperative takes the form of a command. By contrast, the moral law governs reason considered in abstraction from sensible affection – for example, we can imagine it applying to God. Since the moral law is the law of the "holy" will, it is not expressed in imperatival form. (Kant, *Groundwork of the Metaphysic of Morals*, II: 36-39 (412-14).)

way of determining what those laws [for the Kingdom of Ends] would be."[34] This Korsgaard's claim is unclear and unduly compressed. However, one can venture an explanation of what Korsgaard means by it. The moral law is presumably substantive insofar as it appeals to the notion of rationality, which in turn requires the presence of substantive motives and projects. To call a person rational is to say something about his relation to his motives or projects – something about the way in which he manages, or pursues, or combats them. Without the existence of such substantive concerns, the concept of rationality finds no application. With this interpretation of rationality in place, it can be seen why Korsgaard thinks that the moral law has content. The command to adopt only maxims that can serve as laws in the republic of rational beings mandates a class of acceptable maxims: those maxims that we can all rationally will given our underlying cares and concerns. However that command is open-ended, as it does not specify *which* those underlying cares and concerns must be. It only requires that there must be some.

Following Korsgaard's identification of the moral law as a suitable solution to the empty formalism objection, she needs to show how it can be appended to the categorical imperative. The two requirements are not obviously linked, and "the argument that shows that we are bound by the categorical imperative does not show that we are bound by the moral law."[35] Hence a further argument is needed that connects the two requirements. Korsgaard's argument to this effect is that free agents must regard and value themselves as rational agents.[36] As she puts it, they must have the *identity* of Citizens of the Kingdom of Ends. This necessary self-conception of free agents brings a restriction on the content of maxims allowed by the categorical imperative. Given that they must regard themselves as rational, free persons must accept that their universalizable maxims must be such that all persons can rationally will them given their underlying cares and concerns. The moral law is, in this sense, implicit in the categorical imperative. The plausibility of this amendment of Kant depends on the plausibility of the

---

[34] Korsgaard, *The Sources of Normativity*, p. 100, original emphasis.

[35] Ibid., p. 100.

[36] Ibid.

thesis that we must regard and value ourselves as rational agents. Korsgaard defends this thesis by means of a transcendental argument from the conditions of the possibility of any practical self-conception.[37] However, it is not important here whether that argument succeeds. Within the context of my discussion, the interesting aspect of Korsgaard's revision is that it relies on a claim about identities as the "sources of our reasons".[38]

Korsgaard maintains that free agents, detached as they are from their inclinations, nonetheless have a basis for evaluating their maxims. This basis is their evaluative self-conception, "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking".[39] Korsgaard's point here must be understood in the context of her claims about the reflexivity of human consciousness. Since we are beings that are able to reflect on our natural motives, we must decide whether to accept or reject those motives. Once confronted with impulses that incline us towards some courses of action we cannot shrink away, but must determine whether to follow them. Since we are free, we must make that determination by assessing maxims that include those impulses. This will in turn be possible only if we have some fixed standard for assessing our maxims – some criterion against which maxims can be weighed. This standard must be capable of yielding a *conclusive* assessment of maxims, which is to say that it must be able to stop the regress of questions about justificatory grounds. In other words, it must equip persons to make decisive appraisals of maxims. For example, if the maxim is "I will quit smoking" my justification for adopting it may be "Because it is expensive", which meets with the subsequent question "Why should I care about money?" If I can figure out why money is important to me – say, because it gives me leisure time – this further justification can also be questioned, and so on infinitely. This kind of infinite regress would be paralyzing. By swamping the agent with never-ceasing questions about justification, it would sever practical reflection from its action-guiding objective.

---

[37] This argument is given, and explicitly termed "transcendental", in Ibid., pp. 120-25.

[38] Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*, p. 21.

[39] Korsgaard, *The Sources of Normativity*, p. 101.

But we cannot escape acting, suggests Korsgaard, which means that we *do* have means of stopping the regress. Those means are provided by my evaluative self-conception, understood as the way in which I value myself. Persons do not wonder whether to accept a maxim simpliciter; what they need to know is whether to accept a maxim *given their particular identity*. Here, finally, is Korsgaard's answer to the question of what condition normative claims must satisfy in order to address persons: they must appeal to their evaluative self-conception. The real practical question for the smoker is "Should I make it my maxim to quit smoking given that I think of myself as a marathon runner – or a rock star, or a Buddhist, a soldier, or anything else?" That maxim will be acceptable to me – which means that it will become my *prima facie* reason – if it is appropriately related to the description under which I value myself: for example, since quitting would be beneficial to my stamina. Finally, identity-based endorsement of maxims is conclusive as it leaves no justificatory questions to be asked. The answer to the question "Why should I care about my lung capacity?" is contained in the standpoint from which I deliberate, that of a dedicated athlete.

5.4. Two problems with Korsgaard's account

Korsgaard's account of identity-based normativity is bold and ambitious. Not only does it show how a certain view of freedom leaves agents with resources for practical deliberation, it also lays the foundation for an agreeable revision of Kant's moral philosophy. However, that account is not without problems. To begin with, Korsgaard espouses what some, including Kant, would regard as an unacceptably "psychological" conception of human freedom. She defines freedom with sole reference to the contents of one's consciousness, as the ability of the subject to exert control over his mental states. It is the freedom to will what one wants to will. This position is meant to draw on Kant's idea that we must act under the Idea of freedom since "we cannot possibly conceive of a

124

reason as being consciously directed from outside in regard to its judgements".[40]
However, in Kant, the concept of freedom refers to spontaneity that is absolute,
and not just relative to one's psychological states. This is the ability to start a
new intelligible chain of events outside the conditions of space and time,
achieved by acting on the law of pure practical reason, or the categorical
imperative. Without this absolute spontaneity, Kant claims, freedom would be
"nothing better than the freedom of a turnspit, which, when once it is wound up,
also accomplishes its movements of itself."[41]

The criticism that Korsgaard's view of freedom is psychological will not be
universally accepted. Kant's own attempts to show that we are free in the
absolute sense have not met with much approval. In particular, the doctrine of
transcendental idealism that grounds these attempts has been criticized for being
too close to the straightforward idealism that it seeks to dislodge. Even with this
concession, however, Korsgaard's account of identity-based cannot be
considered successful. The unsatisfactory aspect of it is Korsgaard's explanation
of how we come to have identities. In the *Sources of Normativity* Korsgaard
understands identities in a naturalistic way, as unchosen self-descriptions that we
simply find ourselves having. This can be seen in her claim that one's reasons
express one's "nature",[42] and even more clearly in her admission that her
argument "grounds normativity in certain natural – that is, psychological and
biological – facts".[43] However, this thread of naturalism is in tension with a
central part of her Kantian outlook, her focus on maxims as the staple of practical
reasoning. It is appropriate to think that all action is maxim-based because
principled action allows freedom. Exacting a maxim that incorporates thirst-
satisfaction differs from simply responding to the sensation of thirst. By doing
the former a person stamps his authorship on his behaviour; doing the latter
lowers him to the level of a mere respondent to physiological influences and
conditioning. By the same token, the claim that identities directly deliver reasons

[40] Kant, *Groundwork of the Metaphysic of Morals*, III: 101 (448).
[41] Ibid.
[42] Korsgaard, *The Sources of Normativity*, p. 101.
[43] Ibid., p. 160.

denies the possibility of persons having any distance from their contingent designations and memberships. What Simon Blackburn intended as a criticism of Korsgaard is then, after all, an accurate description of her position: we can only accept or reject their inclinations on the basis of yet *other* inclinations.[44] On this view, an agent's identities and not the agent himself determine the grounds of his actions.[45]

In her most recent work Korsgaard appears to give a different account of how we come to have identities, perhaps in an attempt to distance herself from the objectionable naturalism of the *Sources*. She now maintains that persons create all of their identities, as this is their only way of dealing with the "human plight" – the necessity of choosing that comes with reflexivity. She writes:

> We must act, and we need reasons in order to act. And unless there are *some* principles with which we identify we will have no reasons to act. Every human being must make himself into someone in particular, in order to have reasons to act and to live. Carving out a personal identity for which we are responsible is one of the inescapable tasks of human life.[46]

Our identities do not consist in any facts that we must accept, but are generated through our self-definition. As Korsgaard writes, "in the relevant sense there is

---

[44] Simon Blackburn, *Ruling Passions* (Oxford: Clarendon Press, 1998), p. 252.

[45] It can be argued that by espousing this view Korsgaard once again parts company with Kant. As Timmerman notes, for Kant "a person's character is not entirely a matter of naturalistic formation. Nature herself is thought to be within our control as far as our actions are concerned. We are thus fully responsible even for our character; our freedom consists in being able to act on maxims that are both firm and rational." (Jens Timmerman, "Kant's Puzzling Ethic of Maxims," *Harvard Philosophy Review* 8 (2000): p. 43.) Contrast this with Korsgaard's statement that "A view of what you ought to do is a view of who you are." (Korsgaard, *The Sources of Normativity*, p. 117.)

[46] Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*, pp. 23-24, original emphasis.

no *you* prior to your choices and actions, because your identity is in quite a literal way *constituted* by your choices and actions."[47]

However the view of identities as the product of the necessity of choice brings back the concern about justificatory grounds: what basis do we have for selecting some specific identity for ourselves? Note that the question concerns the substantive criterion for opting for some *particular* self-conception over others. It is the question of what can justify our thinking that some distinct and unique self-description is more acceptable to us than others. What is needed is an explanation of what gives us reasons to settle upon some specific identity, in contrast to indiscriminately plumping fo r any identity. Why choose to regard oneself exactly as a "Buddhist" rather than anything else: "chess player", "procrastinator", "lover", "samurai", "smoker", and so on? This is a question that Korsgaard never adequately answers. She comes closest to addressing it in her discussion of what she calls the "paradox of self-constitution".[48] This paradox states that one cannot constitute oneself unless one already exists; but if one exists, there is no need for self-constitution. Her way out of the paradox is to claim that it rests on a misconception about the practical self as a fixed entity.[49] Instead of being a finished product, Korsgaard maintains, the self is an ongoing project that consists in continuous self-integration – in continuously choosing identities for oneself and living up to the standards they impose. Following a long and complicated argument, she then also claims that this ongoing self-constitution is not lawless, but must conform to Kantian principles of practical reason.

Yet this argument still does not answer the question about the grounds for choosing particular identities. Even if it is true that in our efforts to integrate ourselves we must obey Kantian principles, those principles do not tell us which specific identities to adopt, or why. By Korsgaard's own admission, "The Kantian imperatives are principles that instruct us in *how* to formulate our

---

[47] Ibid., p. 19, original emphases.
[48] Ibid., p. 20.
[49] Ibid., pp. 43-44.

maxims; autonomy and efficacy set standards for the form of our maxims."[50] In the context of self-integration, this point means that these principles govern how we should go about choosing our identities, but do not inform us why any particular identity is worth choosing. The criteria for individuals' reasoned choice of their identities still remain unexplained. One may worry whether this point is not unfair to Korsgaard, especially given that she thinks that Kantian moral imperatives are substantive. The objective of her revision of the categorical imperative is to establish this very point. If the principles of self-integration are substantive then perhaps they do warrant choosing some particular identities and steering clear of others. Unfortunately, this defence cannot be accepted. As noted earlier, for Korsgaard the moral law is not only substantive, it is also in an important sense *open-ended*. It commands us to adopt those maxims that we can all jointly and rationally will given our underlying cares and concerns. However it does not specify which those underlying cares and concerns must be; it only requires that there must be some. By extension, all that the moral law can tell us about choosing our identities is that we ought to choose ones that we could all will jointly and rationally. But that norm also assumes that we have some identities already. And the question then is, what grounds can we have for choosing those?

In sum, even if Korsgaard's conception of identity can be made compatible with her account of freedom, it does not illuminate the reasoned basis for deciding upon specific identities. This shortcoming is evident in Korsgaard's explanation of what warrants her self-identification as an American:

> Someone might say to me: okay, sure, I see that you must do that insofar as you identify yourself as an American citizen, but why must you take that way of identifying yourself so seriously? It is only an accident that you were born in America. And here part of the answer is that I must take *some* ways of identifying myself seriously, or I won't have any reasons at all.[51]

---

[50] Ibid., p. 131, original emphasis.

[51] Ibid., p. 24, original emphasis.

The trouble with Korsgaard's explanation is that it does not address the issue. The question is what justifies her valuing herself under the description "American", not what justifies her valuing herself under *some* description. Since the Kantian principles that supposedly govern the choice of identities cannot help in settling that question either, it must be concluded that she cannot answer it. For that reason, Korsgaard's view of identity should be resisted.

6. Universality of reasons

6.1. Introduction

My discussion in chapters 3-5 has shown that some of the most prominent current attempts to show that persons' identities play a part in determining their reasons fail. In the rest of this thesis I will discuss an alternative argument to this effect. The concept I have in mind is tied to a particular view of what reasons are, which I attribute to Kant. The justification for adopting this view of reasons is not that it is plausible *tout court* but that it suitably coheres with, and extends, the theoretical framework of reasons-based justification. Specifically, its strong point is that it accommodates the idea of the universality of reasons, required by the project of reasons-based justification, better than rival accounts of reasons. I will discuss Kant's understanding of reasons in this chapter, and the novel conception of identity that can be derived from it in the next chapter.

As I discussed in chapter 1, in reasons-based liberal theories reasons are understood in the normative sense, as considerations that count in favour of doing something. From this perspective, when we ask about someone's reasons for doing something we are asking about what makes that action sensible or worth doing – we are asking about the *justification* for his behaviour, not an *explanation* of why he acted as he did. A shipwreck survivor's act of drinking saltwater can be explained by describing his distressed psychological state – by saying, for example, that this desperate act is a result of his prolonged solitude and a lack of hope in rescue. But we may still want to now whether there was anything about his situation that made the choice to drink saltwater the proper one in the circumstances. This is the question about normative reasons.

The reason-based justification of political principles is that citizens have reason to accept certain normative principles. An important aspect of this justification is the breadth of its scope. It does not single out, or discriminate against, any specific persons: all citizens, rather than just some particular ones, are meant to have reason to endorse the relevant norms. This has a direct implication for the kind of reason, or reasons, that can be given in support of political principles. Since they have to hold for all citizens, they cannot be *particular* in the sense of being valid only for some specified citizen or group of citizens. Rather, they must have *universal* validity.[1] However, it should be noted that contemporary liberals commonly understand reasons-universality with reference to the members of a bounded political society, rather than humanity at large. The liberal requirement is typically not that reasons behind political norms must be strictly universal, in the sense of having absolutely no restriction of scope. They must apply to all members of the political society, not to all persons without qualification. One way of showing that they are valid for all citizens is indeed to show that they are valid for all persons, but this is not an argumentative route that liberals must, or typically do, take. More commonly, the universality they require of public reasons is confined to the domain of the relevant citizen body. For example, Rawls notes that "the correct regulative principle for anything depends on the nature of that thing."[2] Since the "thing" that his principles of justice aim to govern is a determinate society that regards persons as free and equal, and not the world, their justification needs to be valid only for all members of that society.

I will begin this chapter by examining two ways in which contemporary liberals aim to achieve the universality of reasons that support their normative

---

[1] It is in light of this condition of universality that Rawls requires of principles of justice to be "general", in the sense that "it must be possible to formulate them without the use of what would be intuitively recognized as proper names, or rigged definite descriptions. Thus the predicates used in their statement should express general properties and relations." (Rawls, *A Theory of Justice*, p. 131.)

[2] Ibid., p. 25. For an interesting argument that criticizes attempts to extend liberal accounts of justice to the global domain based on this Rawls's principle see Thomas Nagel, "The Problem of Global Justice," *Philosophy & Public Affairs* 33, no. 2 (2005).

principles.[3] One of them, associated with Rawls and O'Neill, is to purge these reasons of partiality. The other, espoused by Scanlon and Nagel, is to define reason-judgements in such a way that all reason-judgements are necessarily universal. I will then turn to Kant's account of reasons-universality, which will serve as the basis for the transcendental concept of practical identity to be developed in the next chapter.

6.2. Impartiality and universality

One liberal approach to reasons-universality starts from the assumption that non-empirical speculation offers no satisfactory way of arriving at universal justifications. This conviction is exemplified by Rawls, when he remarks that "The analysis of moral concepts and the a priori, however traditionally understood, is too slender a basis. Moral theory must be free to use contingent assumptions and general facts as it pleases."[4] As he suggests in his discussion of Kant, prescriptions that would arise from such pure enquiry would be "purely transcendent and lacking explicable connections with human conduct".[5] In order for a normative theory to be "realistic", it needs to "start from men as they are".[6] A consequence of this repudiation of non-empirical metaphysics is that universal justifications cannot be derived from a pure enquiry into the concept of reasons-universality. Rather, the initial point in devising such justifications must be *partial* reasons that favour the interests and projects of individuals whose reasons they are. As will be seen shortly, there is a connection between partial and particular reasons, in light of which a move towards impartiality can be regarded as a move towards universality. This move towards impartiality is made, and universal justification "constructed", through imposing a set of carefully selected

---

[3] In the further text I will use "universal" to refer to reasons valid for all members of the relevant society, and the term "particular" to refer to reasons that are valid for any individual or group of individuals short of the whole citizen body.

[4] Rawls, *A Theory of Justice*, p. 51.

[5] Ibid., p. 256.

[6] John Rawls, *The Law of Peoples* (Cambridge, MA: Harvard University Press, 1999), pp. 12-13.

constraints on individuals' reasons. The expected outcome is a suitably universal conception of justice that is nonetheless sufficiently realistic.

Before explaining how partiality and particularity may be connected, it is first necessary to distinguish them. The notion of particularity says something about *for whom* reasons justify certain actions, whereas the notion of partiality says something about *how* reasons justify certain actions. So, a reason is particular if it is valid only for some person or persons. A reason is, on the other hand, partial if it justifies the pursuit of certain interests and concerns at least partly on the basis of their being the interests and concerns of some particular person or persons. For instance, it is constitutive of friendship that friends not only put the interests of their mutual relationship before their relations with strangers, they do so at least partly because the friendship is *theirs*, not because it has cosmic significance or value. Such (partly) subjective grounding of not only this, but various other reasons – to care for one's parents more than for other elderly people, to cheer for one's own country in the football World Cup, to send more help to one's compatriots in the wake of an earthquake than to other victims – is not usually found objectionable. A person's entirely contingent relations to other people and things are thought to play a legitimate part in determining his reasons. Moreover, it is thought that these contingent relations can sometimes fully determine a person's reasons without help from further, less subjective considerations. This is one of the points that Bernard Williams makes in his discussion of the husband deciding whether to save his life from deathly peril.[7] The thought that it is his wife that is in danger, Williams argues, is enough to make it sensible for the husband to rescue her. No additional, bias-free principles or facts (such as, for example, that one ought to help drowning people generally) are needed for him to have this reason.

Since partiality and particularity refer to different aspects of reasons, it is not conceptually necessary for partial reasons to be particular. The way in which these two notions are defined leaves open the possibility of reasons that contain

---

[7] Bernard Williams, "Persons, Character and Morality," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), pp. 17-18.

an irreducibly subjective component but are nonetheless universally valid. However, contemporary liberals are committed to a further principle that closes this possibility. According to this principle, a part of what makes political justifications valid for all citizens is that they are *acceptable* to all citizens. Universal acceptability is a necessary condition of the universality of reasons, and any justification that could not be endorsed by all members of the society does not really apply to them.

The source of the demand for universal acceptability of political justifications is what may be termed the perspectival approach to reasons, according to which reasons are to be defined starting from the first-person perspective of reasoning agents. The claim here is that a person cannot have a reason to do something without having some appropriate conscious state: for example, unless he has some desire and a belief about how to satisfy that desire, or unless he has internalized a principle that warrants acting in that particular way. Facts or considerations play a part in determining his reasons but they do so indirectly, by eliciting some appropriate conscious states in him. It is in virtue of the person's states of this sort that he has what can be called *his* reasons. Without this grounding in the person's own point of view, so the argument goes, his reasons would be implausibly detached from him. Jeremy Waldron expresses the point well when he says that

> intelligible justifications in social and political life must be available in principle for everyone, for society is to be understood by the individual mind, not by the tradition or sense of a community... If there is some individual to whom a justification cannot be given, then so far as *he* is concerned the social order had better be replaced by other arrangements, for the *status quo* has made out no claim to *his* allegiance.[8]

---

[8] Jeremy Waldron, "Theoretical Foundations of Liberalism," *The Philosophical Quarterly* 37, no. 147 (1987): p. 135, original emphases.

A political justification that does not make a claim to some citizens' allegiance is problematic in light of the link between justified norms and constraints on individuals' freedom. The link is that if there is a valid universal reason for adopting certain principles of political organization, then there is also a reason for *enforcing* those principles. As we saw in chapter 3, Barry argues that to limit a person's freedom on the basis of reasons that he cannot accept is to coerce him. It is to constrain the person's choices illegitimately, without a basis in his (possible) consent. It then follows that if state actions are to be more than bare coercion, they must be rooted in reasons that are acceptable to all citizens. The demand for universal acceptability can thus be regarded as a legitimacy-based constraint on what kind of justifications can be admitted as publicly normative reasons.[9] Individuals may, and most often do, appeal to justifications that fall short of this standard in their private, everyday lives. However, in matters that involve the exercise of state power such justifications are improper.

The requirement of universal acceptability precludes partial reasons from having universal validity. A justification that is not only given from some person's standpoint but also essentially bound to that standpoint will have trouble persuading those who occupy other standpoints. Suppose that a person X proposes a conception in which justice consists of promoting certain human interests, but justifies that conception on the basis that these interests are important to *him*. This justification is partial because it gives preference to X's own point of view for no other reason than that point of view is his. This reasoning will be insufficiently compelling to others, who inhabit other points of view and may not see any special significance in the fact that some interest is important to X. For some persons, like X's friends and family, this fact may indeed be especially significant, and for them X's reasoning will be acceptable. As an extension of their affection for X, they will be prepared to think that X's concerns, merely in virtue of being his, are worth promoting. However it cannot

---

[9] The liberal criterion of legitimacy, as Rawls calls it, allows only those uses of state power that are "in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational." (Rawls, *Political Liberalism*, p. 217.)

be expected that this affection for X will be shared by all citizens, or that those who do not share it will see anything especially valuable in X's point of view. The same, as we saw in the earlier discussion of Barry, holds for justifying the public protection of a cultural practice by saying simply "It is a part of my culture". In both cases, partial reasons leave some citizens unengaged.

In order to be publicly acceptable reasons need to be impartial, which means that they must justify certain actions without privileging the position or point of view of any particular persons. This is not to say that impartial reasons cannot include a *reference* to some particular person or persons. However, the justification that includes that reference must not rely on the special status of any specific individual. For example, one can have an impartial reason to support one's national football team: that it plays the most beautiful football. That reason picks out a specific national team as worth supporting, but nonetheless remains impartial because it justifies its selection by a principle that does not favour any team, but the beautiful game itself. Similarly, one can advocate the public recognition of one's own cultural practices on impartial grounds. As I suggested in chapter 3, one can claim that the self-respect of all individuals depends on the survival and flourishing of their culture. This claim would then warrant protecting any given person's particular culture, but on the basis of a principle that does not privilege any culture over others.

Rawls's defence of his two principles of justice in *A Theory of Justice* is a good example of the liberal striving to purge political justifications of partiality. He attempts to ensure the impartial grounding for his norms at two levels: at the level of the design of the original position, and at the higher-order level of achieving reflective equilibrium. At the former level, Rawls presents his two principles of justice as the outcome of rational choice, in which each party is interested in maximizing his own interests. The tendency of the parties to privilege certain projects and attachments just because they are theirs is assumed. However, Rawls expunges this element of partiality by radically restricting the range of information available to the contractors. The veil of ignorance, brought down upon the contractors, "excludes the knowledge of those contingencies that

set men at odds and allows them to be guided by their prejudices".[10] Its function is to "nullify the effects of specific contingencies which... tempt them to exploit social and natural circumstances to their own advantage"[11] In other words, it can be said that the veil disables the parties' tendency toward favouring their own projects by depriving them of any knowledge of what interests and concerns really *are* theirs. Without this necessary condition of partiality, the parties are forced to choose without special regard for their own point of view. As a result, reasons that guide each contractor are not valid only for him, but for all members of the society.

At the level of the reflective equilibrium, the impartiality of Rawls's justification is ensured by the way in which he understands considered convictions about justice. To recall, Rawls thinks that principles yielded by the original position must be set against the citizens' stable pre-theoretical judgements about what is just. The ultimate justification of his two principles is that they cohere with such judgements, where the coherence is understood in the "Socratic" way. This means that the principles and the considered convictions influence each other, and each element is open to revision in light of the other until agreement between them is reached. What is important in this context is that the content of pre-theoretical judgements about justice is not entirely undetermined, since they embody the requirement of impartiality. They represent our best current (albeit intuitive) understanding of what would be just from an impartial standpoint.[12] The original position and considered convictions thus equally aim at answering what would be impartially just. This is why Rawls can say that justice as fairness is "the hypothesis that the principles which would be chosen in the original position are identical with those that match our considered judgments and so

---

[10] Rawls, *A Theory of Justice*, p. 17.

[11] Ibid., p. 118.

[12] As Peri Roberts notes, "We can think of our considered convictions as those in which we are currently most confident of their objectivity. If we regarded them as obviously wrong or as merely selfish or biased, then we would not find them convincing. This does not mean that they directly express moral facts or principles but rather that they are those we currently think are best justified and could therefore underpin reasons for everyone." (Roberts, *Political Constructivism*, p. 28.)

these principles describe our sense of justice."[13] Since discrepancies between principles and pre-theoretical judgements will inevitably exist, both elements will have to be modified in search for the reflective equilibrium.

What guarantees the universality of Rawls's justification is that it is the closest we can come, given the most plausible theory of justice and our stable intuitions about justice, to an impartial conception of justice. However, there are several problems with this claim. One is that our best effort at devising an impartial justification still does not amount to an impartial justification. This point is not at all trivial. The idea behind providing a reason to accept a certain conception of justice is that all citizens are supposed to have that reason. If it cannot be shown with certainty that the best justification that can be produced really does apply to all citizens, the whole enterprise is under threat. And Rawls really does not show that about his reasoning. To call a justification our best shot at an impartial conception is to concede the *possibility* that the justification is not impartial after all – or, by extension, universal. With this concession, the reason-based approach to justification fails to live up to its own standards of validity. It strives to ground a conception of justice in reasons that hold for all citizens, but delivers only a conception for which we cannot be sure whether it is grounded in reasons that hold for all. That this conception represents our best effort at impartiality is beside the point. Even if it is the best effort that we can muster, trying one's best is no guarantee of success: the impartial perspective may still be beyond us, and that is what matters. And the same is true if the elimination of partiality is the *only* viable method of justification, as Rawls suggests in the wake of his critique of "empty" metaphysical speculation. The point is still that we cannot be sure that Rawls's approach really does eliminate partiality.

But the main failing of Rawls's approach to universality lies deeper, in the implausible idea that to ensure the impartiality of a political justification *is* to ensure its universality. To prevent misunderstanding, I am not suggesting that impartiality is not enough for universality, as if something more is needed from an impartial reason to make it universal. Rather, my point is a conceptual one. It

---

[13] Rawls, *A Theory of Justice*, p. 48.

starts from the thought, mentioned earlier, that the notions of impartiality and universality are distinct, where one refers to the method of justification and the other to its scope. Striving for impartiality therefore yields one class of reasons (those derived in a certain way), and striving for universality yields another class of reasons (those with a certain scope). Now, my point is that there are no logical grounds for thinking that these two classes necessarily coincide. What gives the false impression that they do is the negative relation between partiality and universality: the fact that partial reasons cannot have universal validity because they cannot be accepted by all citizens. On account of the requirement of acceptability, the class of partial reasons falls outside the class of universal reasons. But it does not follow from this fact that all *im*partial reasons fall *within* the class of universal reasons, or that there is an exact correspondence between the two classes. Logic leaves open the possibility that reasons that do not privilege anyone's point of view nonetheless do not hold for all. Therefore, fixing the partiality of political justification, in the way that Rawls does, is inadequate for ensuring its universality.

6.3. Universality of reason-judgements

The second approach to reasons-universality, exemplified by Thomas Scanlon, concentrates on what it is to make a judgement about reasons, or to "take something as a reason". A necessary feature of such judgements, argues Scanlon, is universality of scope, so that one cannot take a consideration to count as a reason for him only. He must concede that his judgement expresses a claim that applies for everyone. Reasons are "the sort of things, picked out by 'that' clauses, that are the contents of beliefs",[14] where these "that" clauses can be true or false. For instance, I may be correct or incorrect in thinking that I have reason to buy new running shoes. Whether or not I am correct in so thinking, the same will hold for all reasoners, so that if I have reason to buy shoes in my circumstances then so would everyone else, and if I do not then no-one else would. By proclaiming something to be a reason we are thus referring to a domain that is

---

[14] Scanlon, *What We Owe to Each Other*, p. 56.

independent of our judgement, and that exists equally for all reasoners. According to Scanlon, we do not need to appeal to mysterious metaphysical entities or natural facts to account for this domain. The possibility of ascribing truth-value to reason-judgements is enough to convince us of the existence of an objective practical reality. This claim makes Scanlon's approach to universality rather different from Rawls's. For Rawls, as we saw, universality must be constructed from reasons that are initially particular (because they are partial). In Scanlon, by contrast, universality is presumed at the outset: any claim about what there is reason to do is a claim about what anyone would have reason to do in relevantly similar circumstances.

In order to appreciate Scanlon's position, it is necessary to notice that the thesis of universality refers to reason-judgements, not to reasons themselves. Although Scanlon defines reasons merely as considerations that count in favour of acting, it is clear that these considerations have at least something to do with natural facts. For example, as he notes in relation to one's reason for buying a pink hat, "what is relevant is something about the hat, not about my state of mind".[15] It appears therefore that Scanlon does not subscribe to the perspectival view of reasons, which we saw characterizes Rawls. However, Scanlon also advances another claim that restores the first-personal element of practical reasoning, the claim that we never act on reasons themselves but only on what we *take* to be reasons. To understand this point, consider again the shipwreck survivor scenario from the beginning of this chapter. What consideration can justify his drinking saltwater (as opposed to explaining what caused the movements of his body)? One such consideration may be that the island he is stranded on is controlled by vicious and sadistic bandits, who threaten to kill him unless he drinks saltwater. For Scanlon, the survivor's reason for obeying is then given by this fact about his circumstances, a fact about the world that is independent of his perspective on the world. This is his *real* reason.

---

[15] Ibid.

However, the survivor can only act on what he understands to be his reason: on what Scanlon calls his *operative* reason.[16] This distinction between real reasons and operative reasons, between what is sensible for an agent to do and what the agent believes is sensible for him to do, makes possible a discrepancy. Suppose that what our survivor thought were guns in the hands of the bandits were actually water guns. If so, his decision to drink saltwater in order to preserve himself was not backed by a reason – simply because the toys did not endanger his life. He was mistaken about his reasons, and should not have engaged in the behaviour that he did. Otherwise put, his operative reason failed to correspond to his real reason, which was to desist. A person's operative reasons thus fit the bill of reasons only inasmuch as they reflect the reality that is outside his mind. The possibility of a gap between what a person regards as sensible and what is sensible from outside his perspective nonetheless remains. On the one hand, this gap appears easily explicable: it is a simple and unremarkable consequence of flawed perception of the physical reality. Our survivor yielded to bandits waving guns; but had he looked better he would have realized that they are really toys. On the other hand, however, the gap *is* problematic because it is not clear how subjective representations can connect with real reasons. The question is, how can individuals' judgements about reasons be the sort of thing that reflect real reasons? How is it that first-personal practical reasoning can be about the world rather than being about the subject?

Scanlon's resolves this problem with a special understanding of subjective judgements about reasons. If Scanlon were to admit just *any* evaluative attitudes into the class of reason-judgments, then the relation between reason-judgments and the world would be obscure. Spurious reason-judgements that are unsubstantiated by anything in the real world would then be entirely possible. For example, Jane could then decide that helping her neighbour to shovel snow off his driveway is warranted by the fact that Magellan first circumnavigated the Earth. This justifications is clearly ludicrous, but Scanlon wishes to be able to explain *why* it is ludicrous. His explanation comes down to an especially qualified definition of reason-judgments: they are not any evaluative attitudes but

---

[16] Ibid., p. 19.

attitudes that an ideally rational person would come to have whenever that person judged there to be sufficient reasons for them and that would, in an ideally rational person, "extinguish" when that person judged them not to be supported by reasons of the appropriate kind.[17]

The view that Magellan's voyage justifies helping out a snowed-under neighbour would not be held by an ideally rational person. The restrictive definition of reason-judgments thus rules out judgements that do not aim at the real world. Subjective views of reasons are just representations of the perspective of the ideally rational person: they are "concerned with what an individual takes to be reasons in this primary sense".[18]

Furthermore, this definition of reason-judgements also grounds Scanlon's claim that reasons-judgements are universal. Returning to the snow-shovelling example, let G signify the set of facts in virtue of which Jane takes herself to have reason to help her neighbour. According to Scanlon, accepting that G warrants helping the neighbour commits Jane to accepting that *anyone* presented with G has a reason to help a neighbour in need.[19] To put it differently, Scanlon thinks that Jane holds the judgement that she should help her neighbour because a certain abstract condition is satisfied. For example, that condition may be that her neighbour is an elderly person that cannot maintain his house on his own. Scanlon's claim is that if Jane affirms that her neighbour's falling under this description gives her a reason to help him, she is then also affirming that anyone falling under it is entitled to help. The abstract condition grounds a reason-judgement whenever it is met. Whenever we make judgments about our own, we thus consider what an ideally rational person would have reason to do. As a result, "there is fundamentally no question of why we should be concerned with

---

[17] Ibid., p. 20.

[18] Ibid., p. 19.

[19] Ibid., p. 73.

the reasons that other people have."[20] The universality of reason judgements is a straightforward "formal consequence"[21] of the way in which Scanlon defines reason judgements.

Scanlon's idea of the universality of reason-judgements is conceptually tied to his view of how persons deliberate – that is, how they form opinions about their reasons. One might object that this view is phenomenologically misguided. It states that when deliberating what to do, we ask ourselves what an ideal reasoner would regard as sensible in our circumstances. However, this looks like an implausible picture of what happens when persons deliberate. Persons do not make up their minds by second-guessing, as it were, the perspective of the ideally rational agent. They wonder about what *they*, situated as they are, should do. They do not try to jump out of their skins and place themselves in the shoes of the all-seeing observer. In other words, our reasons are not what remains after we correct our views for perspectival error. The perspective that would be so filtered out is the only practical perspective we can have. So, it would have been pointless for the survivor from the example above to wonder whether or not complying with the threats would make sense from some external standpoint that takes all the relevant facts into account. That is not, and cannot be, *his* standpoint, and so is irrelevant to him. He had to decide "from the inside", judging by the facts that he could access.

I appreciate that some will not find the criticism from the phenomenology of practical deliberation persuasive. The thought that persons form judgements about their reasons by transcending their own perspective is a well-reputed one, and may be difficult to resist.[22] However, Scanlon's view of how persons deliberate, which grounds his account of universality, is also liable to another, less controversial objection. That objection is internal to the context of the

---

[20] Ibid.

[21] Ibid., p. 74.

[22] For example, it is advocated by Thomas Nagel in numerous works including Thomas Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970).; and Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986).

reasons-based approach to justification, and states that Scanlon's account defeats the purpose of offering other persons reasons to accept normative principles. The source of the problem lies in Scanlon's preoccupation with reason-judgements at the expense of reasons themselves. His claim, which sets him apart from Rawls, is that we do not act on reasons but on what *seem* to us as reasons. Initially, this shift of focus appears appealing because it allows Scanlon to defend universality in the sphere of practical reasoning, while accommodating the diversity of individuals' circumstances. People find themselves in various contexts and therefore have different reasons; yet their claims about their own reasons have universal validity.

However, maintaining that practical deliberation deals with what seem to be reasons, rather than reasons themselves, has an important consequence. It commits Scanlon to denying that reason-judgements are in themselves action-guiding. He concedes as much when he divides the activity of practical reasoning into three successive stages: seeming, assessing, and opting.[23] At the first of these stages, individuals are confronted with considerations that appear to them to be reasons. The person then needs to determine, in the stage of assessing, whether he accepts to take them as reasons. Finally, he still has to decide, or opt, to act on the consideration he takes as reason-giving. A striking fact about this three-stage sequence is that taking something as a reason is not enough to move an agent to action. On this view, as Watson usefully puts it,

> practical deliberation involves making up my mind *twice*. Making up my mind about what is best to do is coming to a judgement: deciding *that* such and such is the thing to do. Making up my mind about what to do is forming an intention: deciding *to* do such and such.[24]

---

[23] Thomas M. Scanlon, "Reasons and Passions," in *The Contours of Agency*, ed. Lee Overton and Sarah Buss (Cambridge, MA: MIT Press, 2002), p. 169.

[24] Gary Watson, "The Work of the Will," in *Weakness of Will and Practical Irrationality*, ed. Sarah Stroud and Christine Tappolet (Oxford: Clarendon Press, 2003), p. 176, original emphases. Wallace similarly remarks: "The question of what action we are going to perform is not

It is difficult to see how Scanlon could argue anything else than that reason-judgements do not issue in action. To recall, he thinks that we form reason-judgements by asking whether or not a consideration would be regarded as a reason by an ideally rational agent. Even if the answer is affirmative, it is still only a statement of fact: the fact that our subjective opinion conforms with the ideally rational perspective. Something extra needs to be postulated to explain how reasoning translates into doing. For Scanlon, that something extra is a further act of the will, a decision on the part of the agent to be guided by the consideration he can accept as a reason. Now, the troubling aspect of this further decision is that it cannot be made in the same way as the prior decision (that a consideration can be taken as a reason).[25] The criterion of the ideally rational perspective has already been applied, and the consideration has passed it. It is not altogether clear what criterion guides "opting". What can be said is that Scanlon's view undermines a key aspect of the reasons-based approach to justification: its ambition to get people to accept a certain normative conception. Surely, the point of showing that some principles of justice have a valid grounding is for people to embrace them. Its intended aim, in Watson's terms, is a decision *to* act in accordance with those principles, not a decision *that* acting in accordance with them is the thing to do. However, on Scanlon's view this aim cannot be achieved. The "decision to" must be based on reasons that are distinct from the grounds for the "decision that" — reasons that are, moreover, quite unclear. All this would make the reasons-based approach to justification radically incomplete. Hence, the underlying conceptions of practical deliberation and universality must, from the standpoint of that approach, be deemed unacceptable.

necessarily answered by our having determined to our own satisfaction what it would be best to do." (R. Jay Wallace, "Normativity, Commitment, and Instrumental Reason," in *Normativity and the Will: Selected Papers on Moral Psychology and Practical Reason* (Oxford: Clarendon Press, 2006), p. 94.)

[25] In Albritton's poignant formulation, "It isn't for *reasons*, in the end, that we act for reasons." (Rogers Albritton, "Freedom of Will and Freedom of Action," *Proceedings and Addresses of the American Philosophical Association* 59, no. 2 (1985): p. 248, original emphasis.)

6.4. Practical universality in Kant

Rawls's and Scanlon's treatment of practical universality are, for different reasons, unsuited to the enterprise of reasons-based justification. Rawls's approach fails to show that his putatively impartial principles also have universal scope. Scanlon's account, on the other hand, is unacceptable because it renders the reasons-based justifications radically incomplete. I would like to suggest that Kant's conception of reasons-universality succeeds where the discussed views fail: in explaining how it is possible to offer universally valid reasons to accept certain normative principles. This will require a detailed investigation into Kant's view of reasons, which departs some way from most current treatments of reasons. I will not try to argue that this view of reasons is the most convincing without qualification – only that it allows a plausible conception of reasons-universality, required for the project of reasons-based justification. As such, it provides a plausible platform for devising a reasons-based account of identity, which is a task I will take up in the next chapter.

One obstacle in investigating practical universality in Kant is that Kant does not use the terminology of "reasons". It is therefore important to identify which concept plays a role comparable to that of "reasons" in his philosophical system. On a well-established interpretation, for Kant a person's reasons are given by his maxims. More precisely, reasons are maxims that the person endorses in accordance with certain "objective" second-order requirements, but they are maxims nonetheless. Barbara Herman thus writes:

> In the most basic kind of voluntary action, a rational agent
> determines a course of action appropriate to promote or bring
> about a state of affairs she has adopted as an end. In so acting we
> say the agent acts for reasons; Kant says the agent has a maxim of
> action.[26]

---

[26] Barbara Herman, "Leaving Deontology Behind," in *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1993), p. 217.

For Rawls, a Kantian maxim "reflects the agent's actual reasons (as the agent would truthfully describe them) for the intended action."[27] Rüdiger Bittner clearly has maxims in mind when he states that in Kant "a reason for which one does something is a principle on which one acts. Thus it says that reason and action are so related that the latter is a case of the agent's complying with the former."[28] Noell Birondo similarly remarks that the standard interpretation has it that "the concept Kant employs when he appeals to the notion of a maxim is precisely the concept of an agent's reason for action."[29] I would like to argue against the standard reading that identifies reasons as maxims selected in accordance with certain second-order requirements. An alternative understanding of reasons in Kant that I will propose explains, unlike the standard reading, how reasons can be universal.

The issue can be approached by supplementing the remarks on Kant's view of agency that were given in chapter 5. There I discussed an influential interpretation – endorsed by Allison and Korsgaard, among others – of Kant on practical agency. On that interpretation, Kant takes rational agency to consist in an agent's incorporating their inclinations into their own principles of acting, or maxims. This reading begins with the observation that for Kant an inclination or desire is insufficient to count as reason giving.[30] Desires rationally underdetermine action because for any of our desires we may ask whether it is worth satisfying. This is what Korsgaard and Frankfurt refer to as the reflexivity of our consciousness. Now, the question about the worth of a desire cannot be

[27] John Rawls, "Themes in Kant's Moral Philosophy," in *Kant's Transcendental Deductions*, ed. Eckart Förster (Stanford: Stanford University Press, 1989), p. 83.

[28] Rüdiger Bittner, *Doing Things for Reasons* (Oxford: Oxford University Press, 2001), p. 43.

[29] Noell Birondo, "Kantian Reasons for Reasons," *Ratio* 20, no. 3 (2007): p. 266. Richard McCarty notes that "Most English-language interpreters of Kant's moral theory regard maxims as principles or policies expressing the reasons upon which rational agents act." (Richard McCarty, "Maxims in Kant's Practical Philosophy," *Journal of the History of Philosophy* 44, no. 1 (2006): p. 65.) For Roger Sullivan, "an agent's maxim normally could be stated by that agent if he or she were asked to set out the reason for acting in a particular way." (Roger J. Sullivan, *Immanuel Kant's Moral Theory* (Cambridge: Cambridge University Press, 1989), p. 28.)

[30] Allison, *Kant's Theory of Freedom*, p. 40.

answered by stating something about *that* desire because our reasoning takes the form of judgements. One may not, for instance, decide whether to yield to one's desire for a vacation solely by considering that it overpowers all other desires. Citing the strength of the desire already traces the grounds for its-satisfaction to something outside the desire – something like the intrinsic goodness of overpowering desires, or a principle that one should always act on one's strongest desire. Also, the question about the worth of a desire cannot be answered by stating something about *other* desires, because satisfying those needs to be justified as well. However, argues Allison, Kant's system allows inclinations to become reasons for acting. They do so "with reference to a rule or principle of action, which dictates that we ought to pursue the satisfaction of that inclination or desire."[31] And that rule or principle of action, as Allison suggests, is an agent's self-imposed maxim.[32]

This view of practical deliberation, as we saw, has the considerable merit of allowing room for the agent's freedom. It offers a picture of persons free of the grip of their desires, no matter how strong that grip is, by the creation of their own maxims. They thus become self-governing, a law unto themselves, regardless of their affection by sensuous motives.[33] However, the presence of an agent's maxim that calls for φ-ing is only a necessary condition of the agent having a reason to φ. Having that maxim is not enough to give the person a reason to φ; as Kant says in various places, practical principles are not "mere maxims".[34] What he means is that the activity of practical deliberation is governed not only by individual's maxims, but also by certain requirements that have their basis outside his perspective. For Kant, thus, practical reasoning is guided by two kinds of principles, distinguishable by the different grounds of their validity. On the one hand there are maxims, which are *subjective* in that the basis for their validity is the individual's own choice. That is to say, a person's

---

[31] Ibid.

[32] See especially Ibid., p. 86.

[33] A s Allison puts it, practical reason creates an "order of ends or ought-to-bes"; and it is spontaneous in so doing because by framing it in terms of maxims "it goes beyond what is dictated by the sensible data". (Ibid., p. 40.)

[34] See for example Kant, *Critique of Practical Reason*, 5: 19, p. 17.

maxims govern his deliberation because he imposes them upon himself. On the other hand there are principles that also apply to his reasoning but are *objective*, in that the basis for their validity is independent of his choosing. They govern his deliberation regardless of his decision or, more generally, of his perspective. Kant writes:

> A *maxim* is a subjective principle of action and must be distinguished from an *objective principle* – namely, a practical law. The former contains a practical rule determined by reason in accordance with the conditions of the subject (often his ignorance or again his inclinations): it is thus a principle on which the subject *acts*. A law, on the other hand, is an objective principle valid for every rational being; and it is a principle on which he *ought to act* – that is, an imperative.[35]

The idea of objective principles of practical deliberation can be explained by noting that, for Kant, all practical deliberation is done "under the guise of good".[36] This means that maxims are sanctioned or rejected in view of their goodness: "we will nothing under the direction of reason except insofar as we hold it to be good or evil".[37] For this to be possible, it is necessary to have standards for making determinations about the goodness of maxims. Crucially, these standards for assessing the goodness of maxims cannot themselves be (higher-order) maxims. The reason for this is that as Kant explains, maxims do not have the required fixity to serve as any kind of standard.[38] The precepts we make to determine our conduct may be comprehensive and versatile, and our

---

[35] Kant, *Groundwork of the Metaphysic of Morals*, II: 51 (421), original emphases.

[36] The phrase is borrowed from J. David Velleman. (J. David Velleman, "The Guise of the Good," *Noûs* 26, no. 1 (1992).)

[37] Kant, *Critique of Practical Reason*, 5: 60, p. 52.

[38] "[T]he principles that one makes for oneself are not yet laws to which one is unavoidably subject, because reason, in the practical, has to do with the subject, namely with his faculty of desire, which by its special constitution can make various adjustments to the rule." (Ibid., 5: 20, pp. 17-18.)

commitment to them may be deep, but none of this guarantees that we will not break them. They are vulnerable to the vagaries of our moods, the appeal of other projects, our forgetfulness or carelessness, weakness of the will, and so on. We can never truly *bind* ourselves to our own laws, as we imagine that an apple is bound to obey the law of gravity. Since there is always the possibility of deviating from a self-imposed principle, criteria for maxim-selection must be independent of my making. For example, if there is a reason for me to go on holidays this December, this cannot be the outcome of my undertaking to only take vacations in December, joined with my desire for time off this year. A part of my reason – the part that explains the *goodness* of my holidaying this year – must be some principle that I have not made for myself, a principle under which my maxim falls stably and reliably. This principle is the objective ground of my practical reasoning.

Kant goes on to make an even stronger point. Not only must my maxims fall under objective evaluative principles stably and reliably, they must do so necessarily: "Every practical law represents a possible action as good and therefore as necessary for a subject whose actions are determined by reason."[39] In other words, if I deem one of my maxims as worth pursuing, it must be impossible for me to deem it *not* worth pursuing. There is no room for ambiguity when it comes to assessing maxims: I can only call a maxim good if, in my situation, it must be called good and nothing else. This looks like a forbiddingly rigoristic view of practical reasoning, but it is not. It ties in well with the commonplace idea that if I think that my circumstances give me a reason to do something, then I am not free to also think that those same circumstances do not give me that reason. They either do or they do not, but not both. Of course, I can come to see my situation differently, which may lead me to conclude that I was initially mistaken about my reasons. But this is not tantamount to claiming that something both is and is not a reason. It is just a sign that I have discovered *new* facts, which warrant a different judgement about my reasons. Had my situation, or my perception of it, remained the same, my reason would necessarily have stayed the same, too.

---

[39] Kant, *Groundwork of the Metaphysic of Morals*, II: 39-40 (414), emphasis added.

According to Kant, the only evaluative principles that maxims can fall under necessarily are *rules of reason*, which include rules of skill, those of prudence, and those of pure reason.[40] Unlike with self-imposed evaluative principles, if a maxim satisfies any of these rules, it is impossible to conceive of it not being good. Rules of reason necessarily govern the goodness of maxims.[41] However, this necessary relationship between rules and maxims comes in two sorts. In rules of skill and of prudence this relationship is necessary *a posteriori*, which means that the necessary connection depends on practical experience – specifically, the presence of an agent's ends. Both types of rules specify the rationally correct way of pursuing ends. Rules of skill govern the goodness of maxims pursuing known, or determinate ends – for example, they govern a stonemason's deciding on the most efficient way to carve a granite slab. Rules of prudence govern maxims that pursue unknown, or indeterminate ends – principally happiness, the concept of which includes the satisfaction of all desires. Although all humans naturally pursue happiness, argues Kant, they are unable to determine how to realize it because their desires conflict with one another. Striving to be happy therefore requires prudence rather than skill. Nonetheless, the commands of skill and prudence are alike in that both reduce analytically to the requirement to do whatever is needed for the attainment of the chosen end. It is this means-end principle that explains the irrationality of a stonemason who chooses an inefficient way of cutting stone, and a wanton who disregards what might make him happy.

Finally for Kant there is also a rule of pure reason, the moral law. This law governs the goodness of maxims entirely *a priori* – independently of agents' experience, and as a precondition of that experience. The fact that the moral law

---

[40] Ibid., II: 41-43 (415-16).

[41] As Allison notes, there is a helpful analogy with Kant's claims about the conditions of the possibility of theoretical cognition: "the relationship between maxims and objective practical principles is analogous to the relationship in the theoretical realm between empirical concepts as first-order rules for the unification of the sensible manifold and the pure concepts or categories as second-order rules governing the formation of empirical concepts." (Allison, *Kant's Theory of Freedom*, p. 88.)

applies to agents' willing regardless of any willed ends makes it unsuitable for analytical investigation, as "we do not enjoy the advantage of having its reality given in experience and so of being obliged merely to explain, and not to establish, its possibility."[42] The existence of the moral law must be proven synthetically, by means of a metaphysical argument. Kant approaches this task in more than one way, and the success of his arguments is debatable. It is however clear that the commands of the moral law are embodied in the categorical imperative, expressed variously in the Formula of the Law of Nature, the Formula of the End in Itself, the Formula of Autonomy, and the Formula of the Kingdom of Ends.

As we have seen, for Kant there are thus two kinds of requirements of practical reasoning, the subjective (maxims) and the objective (rules of reason). An important matter that still needs to be explained is how rules of reason figure in deliberation about the good. It is clear how maxims determine practical reasoning: persons first represent an action or desire as good, and then design a principle that regulates the pursuit of that action or desire. But rules of reason hold independently of the person's choosing or perspective. Therefore, it makes sense to wonder how they come to bear on first-person deliberation, and what is their connection with the agent's self-imposed maxims. On a widely held view, rules of reason are "second-order principles that specify the norms for maxim selection and action."[43] Whereas maxims are to be regarded as first-order requirements that govern the person's inclinations, rules of reason are requirements of a higher order that govern the final endorsement of maxims. I may, for example, choose to regulate my present desire for a vacation by making it my maxim to only holiday in December. However, my maxim is not enough to give me a reason to wait until December to take a vacation because the maxim itself needs to be validated, or confirmed. This validation is ensured by passing the maxim through a special objectivity-yielding procedure. For example, Rawls has influentially argued that the moral probity of maxims in Kant is tested by

---

[42] Kant, *Groundwork of the Metaphysic of Morals*, II: 49 (420).

[43] Allison, *Kant's Theory of Freedom*, p. 88.

running them through a "Categorical Imperative procedure".[44] Whether a maxim is to be regarded as right or wrong is answered by considering the hypothetical world in which each person acts on that same maxim. If this "corrected social world" could be willed consistently by rational and reasonable persons then this rationale is morally appropriate; if not it is morally improper.

However my interpretation of rules of reason, if correct, rules out regarding them as second-order principles for maxim-selection. On the accepted view, maxims acquire the status of reasons upon passing a special test, such as Rawls's CI-procedure. The assumption is that prior to that test, maxims need not be (and most often are not) governed by objective principles such as the rules of pure reason. Maxims enter the procedure as purely subjective, and objectivity is implanted into them by judging them against objective principles. In this respect, there is a sharp divide between the nature of maxims and higher evaluative requirements. On the reading I have proposed there is no such divide: all maxims are formed on the basis of a representation of some action or desire as good, which representation can only be made with reference to rules of reason. This means not only that without objective rules no action or desire could be called good, without which no maxims could exist. Rules of reason give maxims their distinctive character by providing indispensable reference points for evaluative judgement. Therefore, purportedly subjective maxims cannot as much as be *thought* without recourse to objective requirements. Kant thus remarks that

> instead of the concept of the good as an object determining and
> making possible the moral law, it is on the contrary the moral law
> that first determines and makes possible the concept of the good,
> insofar as it deserves this name absolutely.[45]

Although Kant is here talking about the moral law, his point applies to all rules of reason. They can all serve as the conditions of the possibility of maxims

---

[44] Rawls, "Themes in Kant's Moral Philosophy," pp. 82-90.

[45] Kant, *Critique of Practical Reason*, 5: 64, p. 55.

because they are all necessary standards of goodness.[46] The only difference is that the moral law holds *a priori*, regardless of what ends are pursued and whether they are determinate or not. Rules of skill and prudence have more limited application, depending on the kinds of ends picked out by the relevant maxims. As a result, the moral law governs all possible maxims and no person can beg off from it by saying that he does not value some particular end. However, this should not be taken to mean that, for Kant, only moral reasons exist. It just means that the moral law is basic in the sense that its validity does not depend on empirical conditions. Being basic in this way, it overrides other practical rules in cases of conflict. In all other cases, however, technical and prudential principles find legitimate application, presenting valid claims to guide action.

If rules of reason are the conditions of the possibility of maxims, viewing them as second-order principles for maxim-selection is implausible. That view is founded upon what turns out to be an unacceptable assumption, that maxims are entirely subjective. If they are not, but make essential reference to rules of reason, subjecting them to a special objectivity-yielding test is out of place. Objective principles figure in first-person deliberation not as second-order constraints on maxims, but as their inescapable components. If so, it must be possible to define the notion of a reason without appealing to the notion of maxim-validation. This can be done by considering what it is to form a maxim. Adopting a maxim involves proclaiming some action as good, which is nothing else than proclaiming that the action conforms with one or more rules of reason. Each maxim thus advances a claim about the goodness of its object. For example, by undertaking to only holiday in December I am implicitly maintaining that it is worth for me to take winter holidays. Given this implicit claim, the criterion of the correctness of my reasoning is contained in my very

---

[46] For example, when discussing the prudential precept of saving for old age Kant says: "Reason, from which alone can arise any rule that is to contain necessity, does indeed put necessity even into this precept (for otherwise it would not be an imperative), though it is only a subjectively conditioned necessity and cannot be presupposed in the same degree in all subjects." (Ibid., 5: 20-21, p. 18.)

maxim: that criterion is whether my claim about goodness is indeed true. Since, for Kant, an action can be good only if it agrees with one or more objective principles of goodness, we arrive at the following definition of reasons: an agent's reason consists in the conformity of his maxim with one or more rules of reason. In the case of ends-directed maxims, such as the holidaying maxim, the demand is for conformity with the requirement of maximum expediency. Additionally, there is also a further condition of objectivity that all maxims must satisfy regardless of whether they aim at end-satisfaction, and that is abidance with the categorical imperative.

With this definition in hand, it is possible to enquire whether reasons in Kant have universal scope. Regarding this question, it is useful to notice the way in which my interpretation of reasons in Kant differs from the standard view. To recall, the accepted reading identifies reasons as *maxims* – specifically, those maxims that satisfy second-order objective requirements. In Rawls, for example, a person's (moral) reasons are those of his maxims that pass the CI-procedure. By contrast, I am proposing a thoroughly formalistic understanding of reasons, in which reasons are given by the *fit* between maxims and rules of reason. Having a reason is not a matter of the content of practical reasoning: it does not depend on what actions a maxim enjoins the person to do, and whether those actions could be universalized in some "corrected social world". Rather, it consists in the appropriate relation between one's maxim, whatever it is, and objective principles. It is this kind of formality that Kant has in mind when he writes that reason must be supposed to be practical "of itself and alone", which means that it is

> able to determine the will by the mere form of a practical rule without presupposing any feeling and hence without any representation of the agreeable or disagreeable as the matter of the faculty of desire, which is always an empirical condition of principles.[47]

---

[47] Ibid., 5: 24, p. 22.

If reasons were constituted by maxims themselves, there could be no reasons with universal scope. No two persons could ever share the same reason because, as I noted earlier, it is in the nature of maxims that they are subjective. Every maxim is always and necessarily a precept that some particular person makes for *himself*, to regulate his own inclinations and guide his own actions. An agent's maxim thus cannot be simply reduced to some propositional content to which others can have equal access. For example, one cannot think that my maxim to care for my pet dog expresses the proposition "All persons should care for their dogs", where I just happen to be the person that is presently uttering that proposition, but the same could be done by any other person. The important aspect of my maxim that is lost in this construal is the fact that it is essentially bound to my standpoint. For Kant this is the very definitive characteristic of maxims, and that which makes them a necessary precondition for having reasons. Because maxims are always someone's maxims, it is possible for particular persons to have reasons. If maxims were shareable, the maxims of no-one in particular, there could be no reasons that hold for anyone in particular. In fact, there could be no reasons at all because objective practical principles (rules of reason) could not find application. To understand this point it is enough to consider why rules of reason govern some specific person's deliberation. They do so because by formulating a maxim he implicitly proclaims the conformity of his deliberation with them. He forms a maxim by first calling an action good, and thereby becomes liable to objective criteria of goodness. Now, if *he* had not made this claim to goodness, *he* would not be accountable to rules of reason. Even the categorical imperative issues a command to us only because we bring our maxims under the concept of goodness, and would not apply if we could deliberate in some other way. In this regard, the personal dimension cannot be extricated from the concept of a maxim.

Things are different if one takes Kant to be holding a formal conception of reasons, as I have suggested. If a reason consists in the relation of agreement between a maxim and a rule of reason, it is possible for the same reason to apply to different persons. This relation of agreement requires the presence of two elements, a maxim and an objective principle, but it does not require the presence of some specific, unsubstitutable maxim. A potentially unconstrained number of

156

different maxims can be part of this reason-giving relation of fit with the reason staying the same throughout. As a result, an unconstrained number of maxim-creators can have that same reason. For instance, suppose that I want to have my breakfast outdoors, and that this will only be possible if I clean up my balcony. My reason to clean up the balcony will consist in the conformity of my maxim of cleaning up with the rule of skill that instructs me to pursue the means to my desired ends. Now suppose that my neighbour wants to cycle to work, but his bicycle has a punctured tyre. His reason to repair his bicycle will consist in the conformity of his maxim to repair it with the rule of skill enjoining him to pursue the means to his desired ends. It must be said that the two of us have the very *same* reason, since our maxims stand in the same relation to the same rule of skill. That the two maxims have different content, and come from different authors, makes no difference to the matter.

In sum, in Kant's conception reasons can apply universally because the reason-giving relation of fit is not tied to any particular person or maxim. A reason that holds for someone else may hold for me even if, on the subjective side, its maxim is based on a desire that I do not share, and enjoins that person to actions that are alien and unimaginable to me. Reasons are a matter of the correct form of practical reasoning, not its content. To have a reason is to be rationally required to deliberate in a certain way – so that one's deliberation conforms with one or more rules of reason – not to pursue any specific ends. Therefore, it is possible to offer reasons that apply across persons, and moreover across persons that do not share any ends. Kant's account thus answers the liberal demand for a conception of reasons that allows for reasons-universality. It does not draw the questionable connection between impartiality and universality, in the way that Rawls does. It also shows how it is possible to offer universal reasons that are sufficient to guide action – in contrast to Scanlon's conception, which does not illuminate the grounds for "opting" to act on reasons. On account of the requirement of universality, Kant's conception of reasons is a plausible candidate for employment in liberal justifications of political principles. Some may, of course, regard Kant's conception of reasons implausible on other accounts – for example, its basis in a thick metaphysical framework. This criticism cannot be addressed here. However, it should be remembered that less metaphysically

invested views of reasons must also fulfil the task of accounting for reasons-universality – a task in which, as we saw, Rawls and Scanlon fail and Kant succeeds rather well.

7. Identity as a transcendental concept

7.1. Introduction

In the previous chapter I portrayed Kant's approach to reasons as formal, meaning that it locates reasons in a certain relation between maxims and rules of reason. The account gives an answer to how reasons can be universally valid, which is a precondition for the viability of reasons-based justification. In this chapter I will show how Kant's conception of reasons gives rise to a transcendental concept of practical identity. To see what is involved in investigating the identity of a Kantian self, it is useful to notice a concern that Kant's formalistic conception is liable to. Since it pays vital attention to the first-person perspective of deliberating agents, it must explain how persons can act on reasons of a purely formal kind. On Kant's view, as I noted earlier, what makes something a reason for me is not the fact that it happens to be my reason, but that it could be anyone's reason in relevantly similar circumstances. Reasons consist in an appropriate relation between maxims and rules of reason, and it is irrelevant to that relation whose maxim stands under it. So, the fact that a reason involves *my* maxim falling under a rule of reason makes no difference to the reason being a reason. Someone else's maxim, if it can also satisfy that rule of reason, can constitute the same reason. And yet it is, after all, *I* who has to act on my reason. It must be clarified how this is possible, given the obliteration of particularity that Kantian reasons involve.

The problem needs to be distinguished from a similar issue, which is to show that universal reasons are not necessarily *impersonal*. Contractarian liberals have been particularly keen to argue that point, insisting that their position does not

entail an impersonal view of individuals and their deliberation.[1] Strictly speaking, contractarians reject impersonality as an approach to deriving principles of justice, not as a view of reasons. In Rawls, for example, the rejection of impersonality serves to ground the argument that the principle of greatest average utility would not be chosen by contractors behind the veil of ignorance.[2] This outcome of the choice-scenario would be preferred by what Rawls calls classical utilitarianism, which he traces back to Hume and Adam Smith. Its distinctive feature is the adoption of perspective of the impartial spectator that stands outside the society for which moral norms (in Rawls's case: norms of justice) are to be decided. Morally proper norms are those that would be accepted by an external observer who takes into account the interests of all members of the society without privileging any of them. It seems natural to assume that the impartial observer perspective would favour an arrangement that delivers maximum utility to each person, compossible with its maximum utility to every other person. Since the interests of no-one are privileged, the result would be a system of maximum average utility.

Rawls's objection to the impartial spectator method of deriving principles of justice is that it fails to honour the separateness of persons. In the classical utilitarian interpretation of the contractual situation, "The principle of rational choice for one man is taken as the principle of social choice as well."[3] Rawls's point is not that some contractors' well-being will remain unaccounted for in the utilitarian calculus. Since for the spectator every member of the society "counts for one and no more than one", this criticism is off the mark. Rawls's idea is, rather, that the interests of all will be accounted for in the wrong way: utilitarian principles are guided by what is good *for* each person, but fail to respect what is considered to be good *by* those same persons. Instead of running roughshod over persons' plans and commitments, principles of social organization must be an expression of their ideas about what constitutes a good life. For Barry, similarly,

---

[1] For an extended discussion of this debate see Susan Mendus, *Impartiality in Moral and Political Philosophy* (Oxford: Oxford University Press, 2002).

[2] Rawls, *A Theory of Justice*, pp. 183-92.

[3] Ibid., p. 187.

the elementary idea of impartiality that must inform any plausible theory of justice is that "we live in a world that is full of other people with different conceptions of the good, and they should have a fair chance to pursue them even if we have the power to stop (some of) them."[4]

It should be clear from my earlier depiction of the Kantian perspective on reasons that it does not fall prey to the charge of impersonality. The only standpoint from which judgements about the goodness of maxims can be made is the first-personal one. To say that a person has a reason to do something is to say that *his* maxim, incorporating *his* inclinations, conforms with objective rules of reason.

Nonetheless, this position generates a different puzzle, which will be the topic of this chapter. For Kant, a reason arises as the solution to some specific person's problem of determining what to do. However, and seemingly paradoxically, it solves this problem without reference to anything about the content of that person's maxim. Its normative validity is grounded solely in a fact about the relation between his maxim and objective standards of goodness. It follows that no person can claim to stand in any special relationship to his reasons. He adopts a maxim, which is a necessary but not a sufficient condition for the emergence of a reason (the other being a rule of reason). And even in that function he can be replaced by other maxim-adopters, so long as their maxims conform to the same rule of reason. All this makes for a rather wide chasm between reasons and reasoning agents, with their particular circumstances and particular concerns. What is then unclear is how any self can act on Kantian reasons. I will present and assess this concern with reference to some of Bernard Williams's famous objections to Kant's deontological ethics. Ultimately, the points that will emerge from this discussion will be useful in determining how one may think about the practical identity of a Kantian self.

---

[4] Brian Barry, "Something in the Disputation Not Unpleasant," in *Impartiality, Neutrality and Justice*, ed. Paul Kelly (Edinburgh: Edinburgh University Press, 1998), p. 237.

7.2. The problem with categorical duties

The principal target of Williams's attack on Kant's ethics is its postulation of categorically overriding moral duties. He writes:

> There can come a point at which it is quite unreasonable for a man
> to give up, in the name of the impartial good ordering of the world
> of moral agents, something which is a condition of his having any
> interest in being around in that world at all. Once one thinks about
> what is involved in having a character, one can see that the
> Kantians' omission of character is a condition of their ultimate
> insistence on the demands of impartial morality, just as it is a
> reason for finding inadequate their account of the individual.[5]

Assessing Williams's claims in the context of my discussion requires some care, as they are not strictly cast in terms of individuals' *reasons*. In fact, Williams does not use the notion of a reason as a technical term, instead treating it interchangeably with terms such as "motivation", "project", and sometimes even "desire". His charges against the Kantian view of the self can therefore be interpreted in various ways. For instance, his argument can be read as saying that the thesis of the categorical priority of morality founders on purely motivational considerations. The thesis requires persons to extinguish some goals and projects that are vitally important to them, which they could never bring themselves to do. It can then be said that the moral duty never really gets a grip on them. For example, it is not difficult to understand how the duty to refrain from bribing medical staff can fail to move a daughter whose mother is in desperate need of scarce medication. The duty is too taxing on her emotions and personal attachments, which together constitute her character, and by proclaiming its unreserved sovereignty the Kantians fail to appreciate the important motivational aspect of the human condition.

---

[5] Williams, "Persons, Character and Morality," p. 14.

162

However, if Williams's objection is given this strictly motivational interpretation, it fails to engage with Kant's position on practical reasoning. We saw earlier that for Kant practical deliberation is not a matter of considering desires themselves, but maxims which incorporate desires. If this is so, one's desires, no matter how strong they are, have no direct bearing on the issue of someone's reasons. Inclinations cannot conflict, or defeat, reasons because they are not part of the evaluative process that issues in action. To be sure, desires have a decisive indirect influence on individuals' reasons, since our maxims can only be constructed from the material of our desires. We are, so to speak, motivationally situated at the basic level. For the daughter above, this means that her reasons will have to take vital account of her deep attachment to her mother's well-being. But this attachment cannot simply *be* her reason. In order for it to impact her actions, she will have to fit it into a principle she makes for herself – and it is precisely in so doing that she becomes accountable to certain objective rules of goodness.

This response might be considered too quick. It may be claimed that Williams has more resources at his disposal than just to claim that duties run up against desires. Indeed, it seems plausible to assume that Williams's conflation of motivational and normative aspects of practical reasoning is a not a result of conceptual confusion. Instead, it can be taken as indicative of a thought that he develops more fully elsewhere.[6] The thought is that motivational potential constitutes a condition of a consideration's normative force. On this view, for some considerations to become rationally compelling for a person, they must first be capable of moving him to action. Desires thus determine practical reasoning more immediately than in Kant, and the fact that a duty conflicts with a person's desire does have a bearing on the issue of his reasons. However, Williams can only establish the relevance of desires to reasons by straying considerably from Kant's position. His motivational condition (sometimes called the "internalism requirement") is not a part of Kant's approach to practical reasoning. For Kant, reasons are considered and employed in the heat of action; as such, they are

---

[6] Bernard Williams, "Internal and External Reasons," in *Moral Luck* (Cambridge: Cambridge University Press, 1981).

always action-guiding. His position can be contrasted with what Onora O'Neill calls "a retrospective, spectator perspective on action or ethics". On that view contemplating one's reasons is, like theoretical speculation, concerned with what facts there are, and therefore can be undertaken an essentially impassive audience. But Kant's position is different. For him, "The basic task of practical reasoning is to guide action rather than to adjudicate past acts."[7] Considering reasons has a profoundly practical function: to answer the first-personal question "What am I to do?" Because this is so, motivational potential is a necessary element of what it is to be a reason. It is not, as Williams would have it, a condition that applies to *pre-existing* normative claims, sorting the legitimate from the illegitimate normative demands. For Kant, since every normative claim aims to settle a practical problem, the action-guiding function is built into normative demands from the very start.[8]

In order to assess the impact of Williams's criticisms of Kant, it is necessary to frame them in terms that are compatible with Kant's theory. This requires casting his arguments exclusively in terms of reasons. Although this reinterpretation is at odds with the spirit of Williams's claims, it is essential for determining to what extent they relate to Kant. Accordingly, Williams's argument may be restated as saying that a person may sometimes have a *reason* to resist moral commands: specifically, when moral duties violate one or more of his "ground projects" – pursuits which are "closely related to his existence and which to a significant degree give a meaning to his life."[9] The concept of a ground project does not necessarily refer to only one distinct objective or ambition; rather, it can encompass a "nexus" of projects, "and it would be the loss of all or most of them that would remove meaning".[10] The trouble with categorical moral duties is that

---

[7] Onora O'Neill, "Rationality as Practical Reason," in *The Oxford Handbook of Rationality*, ed. Alfred R. Mele and Piers Rawling (Oxford: Oxford University Press, 2003), p. 95.

[8] As Korsgaard notes, "the Kantian *supposes* that there are operations of practical reason which yield conclusions about actions and which do not involve discerning relations between passions (or any pre-existing sources of motivation) and those actions." (Korsgaard, "Skepticism About Practical Reason," p. 8, emphasis added.)

[9] Williams, "Persons, Character and Morality," p. 12, original emphasis.

[10] Ibid., p. 13.

they may present demands that stamp over this basic framework of commitments. Since ground projects give purpose to a person's life, accepting to sacrifice them would be truly insufferable, and hence unreasonable.

7.3. The assessment

In assessing Williams's (restated) objection it is necessary to note that it relies on a particular view of how moral duties present themselves to the person. On this view, persons experience moral norms as "side-constraints" on their otherwise morally neutral pursuits.[11] The important aspect of this view is that it regards moral obligations as distinct *in kind* from the considerations that spring from our ordinary attachments to parents, friends, and lovers. As Williams puts it, moral and non-moral motivations have "deeply disparate characters".[12] The chasm can be compared to the utilitarian distinction between a principle of publicly correct conduct – that of maximum utility – and the various principles that guide individuals' private deliberations. It is not a part of the utilitarian position to require that persons comport themselves on the basis of a calculation of which action will maximally benefit all members of the society. They can be concerned with their own selfish interests as legitimately as they would be with the happiness of all. The utilitarian schema leaves open the grounds of private practical reasoning.[13] However, there is one distinct principle, the utility

---

[11] The term "side-constraints" was introduced in Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 28-33.

[12] Williams, "Persons, Character and Morality," p. 2.

[13] Cf.: "[I]t is a misapprehension of the utilitarian mode of thought, to conceive it as implying that people should fix their minds upon so wide a generality as the world, or society at large. The great majority of good actions are intended, not for the benefit of the world, but for that of individuals, of which the good of the world is made up; and the thoughts of the most virtuous man need not on these occasions travel beyond the particular persons concerned, except so far as is necessary to assure himself that in benefiting them he is not violating the rights that is, the legitimate and authorized expectations of any one else. The multiplication of happiness is, according to the utilitarian ethics, the object of virtue: the occasions on which any person (except one in a thousand) has it in his power to do this on an extended scale, in other words, to be a public benefactor, are but exceptional; and on these occasions alone is he called on to consider

principle, that guides public deliberation – for instance, deliberation concerning the distribution of state resources. Importantly, this public criterion of right is not reducible, or in other ways related, to the multifarious standards of private practical reasoning. It is not, for example, somehow distilled from the heterogeneous private reasons, nor is it implied by them. The principle of utility would still be normatively valid for a society of egoists (even though it might be difficult to implement).

Similarly, Williams claims, the justification for Kant's categorical imperative has nothing to do with the justification behind the less exalted considerations that move us – considerations such at the loveliness of a lover's smile, or the preciousness of a friendship, or the soundness of saving for one's retirement. If both kinds of considerations derived from the same normative basis, it would be possible to imagine a situation in which prudential, romantic, filial and other such reasons can present a challenge to the primacy of morality. Putting the point crudely, the numbers or the intensity of "low-worth" considerations might add up to contest the dominance of the "high-worth" ones. A loved one's health might be so bad that bribing the medical staff might look like a reasonable option; a freedom-fighter's cause might be so valuable that one might be willing to lie in order to save his life. Yet, according to Williams, Kantian duties never fall under such a challenge because their validity derives from a distinct and unique normative basis. It is due to this normative distinctness that the demands of morality can categorically override other considerations. This special normative status has a consequence for how persons experience moral norms: persons can only regard them as limiting their range of acceptable options – a range that would be larger if moral boundaries were not in place.

Williams is not alone in understanding Kantian moral duties as side constraints on the pursuit of non-moral ends. In fact, this reading is a very common feature

---

public utility; in every other case, private utility, the interest or happiness of some few persons, is all he has to attend to." (John Stuart Mill, *Utilitarianism*, ed. Roger Crisp (Oxford: Oxford University Press, 1998), pp. 65-66.)

of contemporary Kant interpretation.[14] However, if the account of Kantian reasons I gave in the previous chapter is correct, that reading of Kant should not be accepted. To begin with, I wish to suggest that moral and non-moral considerations are not as deeply disparate as Williams claims because both are framed in terms of individuals' *reasons*.[15] The basis for the validity of both sorts of considerations is hence the same: conformity with objective rules of goodness. It is important not to misunderstand this point. It does not state that moral obligations can be somehow inferred from ordinary practical pursuits – for example, by reducing all of our non-moral considerations to a set of shared commitments, or by using a special algorithm to purify our ordinary motives of selfishness and short-sightedness. In this regard, Williams is correct to claim that there is no way of getting to morality from the non-moral starting point. But this need not mean that duties therefore have nothing in common with our particularist concerns. Even if moral commands cannot be derived from non-moral interests, it is still possible to maintain that both owe their validity to a shared source.

This, I have suggested, is what Kant maintains. Whenever we turn to practical reasoning to determine our actions, Kant argues, we consider and employ reasons. This is not tantamount to saying that we always act on reasons. There are very many occasions in which we do not conduct ourselves in the light of practical reasoning. For instance, it is not common to find people dispassionately judging their maxims against objective rules of reason when deciding whether to ask for a lover's hand in marriage, or whether to doze off in the afternoon sun, or whether to avenge an insult. However, according to Kant, only practical reasoning yields itself to rational investigation since its determining grounds are communicable. Unlike acting from love, or habit, or lust, practical reasoning is a *conceptual* activity, consisting in bringing maxims under the concept of

---

[14] For instance, Barbara Herman sees Kant as maintaining that "principles of right constrain our pursuit or particular conceptions of the good". (Herman, "Leaving Deontology Behind," p. 210.)

[15] Since Kant himself does not speak of reasons, this claim should be understood as shorthand only. It summarizes the thought that for Kant, the standards of correct practical reasoning are given by an appropriate relation between one's maxims and one or more rules of reason.

goodness. Kant's focus on reasons is hence compatible with the (alleged) fact that most of our everyday decisions can be traced to the influence of emotions, moods and habits. A father may, for instance, be utterly consumed by love for his children. Kant should then not be understood as reducing, quite implausibly, all his feelings to reasons. His point is merely that *for the purpose of rational examination* it must be assumed that our consummate father's behaviour is grounded in his representation of his actions as good. Only on this assumption can his actions be "appraised by reason and hence through concepts, which can be universally communicated, not through mere feeling, which is restricted to individual subjects and their receptivity".[16] Once this point is accepted, the unified normative basis of moral and non-moral considerations becomes clearer. A maxim of any kind is worthy of acceptance if and when it complies with rules of reason. Moral norms are not an exception in this regard. The concept of a duty just denotes a special case of this compliance – the case in which an endorsed maxim conforms with the rule of pure reason, or categorical imperative.

If duties are themselves reasons, they do not conflict with or, strictly speaking, override particularist interests. Rather, it can be said that moral obligations express reasons that we necessarily have – that is, reasons that we have by the mere virtue of engaging in evaluative deliberation, regardless of what it is that we pursue. If this is so, the view of morality as a system of side-constraints, upon which Williams's critique rests, is implausible. Yet, it may be objected that this view neglects the fact that we sometimes feel resistance towards doing what duty requires. Without a doubt, living morally can sometimes feel impossibly hard. In the example above, the daughter required to refrain from bribing the medical staff at the expense of prolonging her mother's pain will almost certainly think so. For Williams, compromising one's ground projects for the sake of the moral law is so taxing as to be insufferable, which only seems to confirm that morality limits us – sometimes oppressively so.

However, this line of reasoning should be resisted. In assessing it, the crucial question is just *what* is so hard about the moral life. Kant gives a surprising reply

---

[16] Kant, *Critique of Practical Reason*, 5: 58, p. 51.

that is not only plausible but also consistent with the interpretation that I have given. He writes:

> Man feels in himself a powerful counterweight to all the commands of duty presented to him by reason as so worthy of esteem – the counterweight of his needs and inclinations, whose total satisfaction he grasps under the name of 'happiness'. But reason, without promising anything to inclination, enjoins its commands relentlessly, and therefore, so to speak, with disregard and neglect of these turbulent and seemingly equitable claims (which refuse to be suppressed by any command). From this there arises a *natural dialectic* – that is, a disposition to quibble with these strict laws of duty, to throw doubt on their validity or at least on their purity and strictness, and to make them, where possible, more adapted to our wishes and inclinations; that is, to pervert their very foundations and destroy their whole dignity – a result which in the end even ordinary human reason is unable to approve.[17]

Crucially, Kant maintains that the unease we sometimes feel when confronted with a moral duty is *not* directed at performing what duty requires. We never can, or do, resent doing the moral thing – even if it means compromising the well-being of one's parents, or giving up a cherished project, or suppressing a deep disposition of character. This claim sounds wildly radical, but in fact it is not. Kant's point focuses on what grounds we may legitimately rebel against morality. Motives that one might cite in support of refusing to do what duty requires can be emotive, or habitual, or temperamental: that he loves his parents, or that he has always played football with friends, or that he is just so constituted to take his anger out on others. But morality speaks in a different voice, the voice of reasons. It tells us that by making judgements about the goodness of our maxims, we become accountable to objective standards of goodness. Facts about our habits, feelings and temperaments are irrelevant to this justification of moral

---

[17] Kant, *Groundwork of the Metaphysic of Morals*, I: 23 (405), original emphasis.

requirements, and therefore cannot be summoned to challenge it. This is why it cannot be sensibly denied that we should do what obligations ask of us. The only possible objection against duties is "to throw doubt on their validity or at least on their purity and strictness". But this quibble concerns the moral law – namely, whether it really exists as a law – not the particular things we must do. It is hence unrelated to our perception of how hard it is to follow the moral commands. Instead, the natural dialectic arises from the worry that the principles that govern the operation of practical reason rest on an insecure foundation. In this way, argues Kant, "the *common reason of mankind* is impelled, not by any need for speculation... but on practical grounds themselves, to leave its own sphere and take a step into the field of *practical philosophy*."[18]

7.4. The deeper objection

Williams's objection to Kant's view of the self fails because it relies on the mistaken view of Kant's moral requirements as side constraints. However, in the course of making this objection Williams also hints at the possibility for a different criticism of Kant, one that need not rely on a flawed interpretation. This criticism is underdeveloped in Williams's article, and must be reconstructed. The task is nonetheless worthwhile because the new objection asks rather more penetrating questions of Kant, which will be helpful in working out a Kantian concept of identity.

The new objection is given in Williams's claim that the Kantian perspective on reasons provides "ultimately too slim a sense in which any [of my] projects are mine at all."[19] Originally, the claim is closely tied to the view of morality as side constraints. Williams's idea is that moral constraints are insufferable because they undermine practical character. Yet the suggestion that each person has a practical character that makes his actions uniquely *his* is separable from the notion of moral constraints, and hence merits attention. It can be introduced by

---

[18] Ibid., I: 23-24 (405), original emphasis.

[19] Williams, "Persons, Character and Morality," p. 12.

briefly considering Williams's critique of utilitarianism, which is strictly parallel to his attack on Kant. The utilitarian position, with its postulation of the uncompromising objective of public utility, misrepresents the nature of practical reasoning. It portrays the reasoning agent as

> a channel between the input of everyone's projects, including his own, and an output of optimific decision; but this is to neglect the extent to which *his* actions and *his* decisions have to be seen as the action and decisions which flow from the projects and attitudes with which he is most closely identified.[20]

In other words, there is no sense in which the conclusion of utilitarian practical deliberation belongs to the reasoning agent, or is expressive of his particular nature: it is merely the preferred outcome of a procedure of utility-maximization. But if this is so, it is difficult to see why *he* should live by it. A similar obliteration of character is suffered, argues Williams, by a person surrendering himself to Kant's moral duties. In an example he made famous, Williams considers a hypothetical scenario in which a person must decide whether to save his wife from a drowning ship.[21] Kant's ethics certainly requires that the person save his wife, which accords with our intuitions that he really ought to save her. But the Kantian view of the situation is plausible only superficially, as it asks of the husband to come to his wife's rescue for the wrong reason. In order for his behaviour to qualify as moral, he is required to act in complete disregard of what looks like the only appropriate consideration to move him: the fact it is *his wife* that is in peril. Instead, the husband must arrive at the decision to rescue in a strangely detached fashion: by treating his own wife as one among many persons that deserve help on account of their rational nature. By bracketing all of his particular attachments, the husband is thus required to act on "one thought too many".[22] As a consequence, his choice will not be recognizably his.

---

[20] J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (London: Cambridge University Press, 1973), pp. 116-17, original emphases.

[21] Williams, "Persons, Character and Morality," pp. 17-18.

[22] Ibid.

From this point, Williams goes on to supplement his critique of utilitarianism and Kant with a positive theory of what makes a person "own" his actions. According to that theory, an agent's choices are expressive of his character if and when they are mandated by his "ground projects". In order to understand the notion of a self defined by its core commitments it is useful to appreciate the context in which Williams proposes it. The notion first emerges in the context of his discussion of some of Derek Parfit's arguments about (metaphysical) personal identity. Williams appeals to the notion of ground projects to support his claim, against Parfit, that we have reason to care about our future well-being (more precisely, that we have reason to care about what happens to our future selves). He argues that the thesis that persons do not care about their future can only be sustained if it is possible for persons to imaginatively project themselves into the future. This imaginative leap, in turn, requires or persons to possess a concept of themselves. They have to be able to conceive of *themselves* in the future so as to determine that *their* interests beyond the present are not worth caring about.

And yet, argues Williams, once persons form a concept of themselves, it becomes impossible for them to be indifferent about what will happen to them. Self-conceptualization is inseparable from discerning certain core concerns that make one the person that one is. To think of oneself, as a distinct agent, just is to detect a certain unity of purpose and commitment – a unity that, at the most fundamental level, endows persons with some particular character. Since one can only think of oneself as a person who cares about certain things, to imagine oneself in the future is also to envisage a person with certain basic commitments. It might be objected that this argument only shows that the future version of myself will have *some* basic concerns. They need not be concerns that I now recognize as mine. The disconnect between my present and future selves hence still seems possible: if a future self is committed to altogether different projects than me, why should I care about happens to him? However, as Williams notes, to maintain that a future self has different ground projects than me is to maintain that he is not *my* future self. If one accepts Williams's understanding of a self as defined by his basic commitments, the volitional continuity of a self cannot be

denied. And if the self indeed is so temporally extended, one must have an interest in one's future states.[23]

The views about the self that Williams raises in opposition to Kant are an extension of his critique of Parfit. The idea of character that is marked out by certain constitutive concerns is here not used in support of the volitional continuity of the self. Instead, it is meant to explain individuals' practical investment in the world. Against Parfit, Williams aims to show why we are not indifferent towards our future; against Kant, he wants to show why we are not practically indifferent generally. The questions he sets out to answer are: Why do we bother making practical choices? How come it makes a difference to us that things turn out in some specific way, rather than another? Why is it that we ponder, let alone worry about, what to do – why not do *whatever*, or just nothing at all? Williams finds the unifying answer to these questions in

> the idea that my present projects are the condition of my existence, in the sense that unless I am propelled forward by the conatus of desire, project and interest, it is unclear why I should go on at all: the world, certainly, as a kingdom of moral agents, has no particular claim on my presence or, indeed, interest in it.[24]

Should Kantians accept Williams's thick notion of practical character that is constituted by ground projects?[25] Everything depends on Williams's claim that

---

[23] As Williams notes, if a person "clear-headedly knows that his present projects are solely the projects of his youth, how does he know that they are not *merely* that, unless he has some view which makes sense of, among other things, his own future? One cannot even start on the important questions of how this man, so totally identified with his present values, will be related to his future without them, if one does not take as basic the fact that it is his own future that he will be living through them." (Ibid., p. 10, original emphasis.)

[24] Ibid., p. 12.

[25] I cannot here discuss whether Williams's critique of Parfit succeeds. Let me just note that that critique ultimately rests on a particular understanding of the self that might not be compatible with Parfit's theory. In particular, Parfit rejects the understanding of a self as temporally unified in any way, preferring to regard it in terms of the Humean bundle of perceptions. Much of the

they *need* to accept it in order to make sense of agents' practical investment in the world. If this claim is correct, the Kantian outlook will need to undergo drastic revision. Most importantly, it will have to abandon the view that considering reasons is a purely formal activity that consists in discerning relations between maxims and rules of reason. Practical deliberation will then be guided by substantive ends after all – ends related to the preservation and furtherance of one's core concerns.

However, the problem to which the conception of practical character supposedly offers a solution is unclear. It is not at all easy to understand what Williams means by saying that we are "propelled" forward in life, and that this requires an explanation. A plausible assumption is that the sense of Williams's statement is motivational. Presumably, it means that in order for one to be susceptible to the normative influence of any consideration, one must first be able to be moved by it. It can then be said that this ability to be moved by certain considerations and not others can only be explained if one's self has a determinate and fixed structure. Without this structure, constituted by one's ground projects, everything in this world would leave us cold. As Williams puts it, "unless such things exist, there will not be enough substance or conviction to a man's life to compel his allegiance to life itself."[26] However, once again, this purely motivation-centred interpretation of Williams does not relate to Kant. On the Kantian view, our ability to be moved by considerations does not need to be explained. What "propels" persons forward is reasons. The basis for caring to act on a maxim is just that it presents a rationally plausible answer to the question of what to do. Action-guiding potential is an aspect of normative claims from the very start. Since this is so, there is no need to postulate a substantive notion of practical character to explain the action-guiding nature of practical reasoning. Williams's notion thereby becomes superfluous.

---

point of his treatment of metaphysical identity lies in this rejection. It is then unclear why Williams's criticism, coming from a very different philosophical perspective, should exercise him.

[26] Williams, "Persons, Character and Morality," p. 18.

Perhaps Williams's claim about us being "propelled" forward can be understood differently, in a way that does relate to Kant. Earlier in this discussion I noted that this is only possible if Williams's theses are restated in terms of reasons. Consequently, it may be argued that ground projects constitute higher-order reasons: they are the source of *reasons for reasons*. Practical character would then offer a response to a seemingly puzzling question: if particular reasons (to perform this or that action) express what behaviour is rationally warranted in some specific circumstances, what makes acting on reasons rationally warranted? The appealing possibility is to say that every reason owes its normative dimension to the fact that it expresses the agent's ground commitments. In this regard, ground concerns make up the primordial material from which reasons are built. They situate agents within the world, which is a precondition for any action at all to be rationally warranted. Without this "stuff" that fuels practical reasoning, the formal activity of discerning relations between maxims and rules of reason would be completely vacuous.

However, this line of reasoning runs two considerations too closely together. One of them is that we are, through some feature of our natural constitution, disposed toward seeing some actions as rationally warranted. We are, so to speak, naturally attuned to find reasons for doing certain things. The other consideration is that we *have* reasons for doing certain things. On my restated interpretation, Williams wants to explain the latter fact by means of the former. The argument is that since we have to be somehow situated in the world in order to have reasons, this situatedness also explains *why* we have reasons. But this conclusion does not follow from the premisses. Structurally, it is identical to explaining the cause of a patient's illness by stating that he has a disposition to get ill – for instance, due to a weakened immune system. This explanation cannot work because potentiality is not the same as actuality. It takes something to move a person from a state in which he merely has the potential for sickness to the state of being sick. That extra factor might, for example, be exposure to a virus. By the same token, citing persons' basic-level investment in the world only illuminates a necessary, not a sufficient condition of their reasons. The lacking ingredient is an account of what it is that transforms mere potentiality for reasons – agent's receptivity to the world – into actuality.

Kant has an explanation of how this transformation occurs, and is therefore not pressed to accept the unsubstantiated account of the self that Williams proposes. For him, it is the agents themselves that refashion, in accordance with rules of reason, their sensibly conditioned desires into reasons. This also enables us to see why, in the context of Kant's theory, the idea of higher-order reasons is fundamentally misguided. Kant has no need for the notion of practical character because his framework rules out the concern whether persons have any reason for acting on reasons. A maxim's conformity with one or more objective criteria of goodness leaves no further questions to be asked. This favourable relation signals that the maxim fully complies with the standard that, by making a claim about goodness, it sets itself. Such a maxim is as perfectly justified as it can be. The only way to deny this is to contest the validity of rules towards which maxims must orient themselves, the most basic of which is the categorical imperative. But, as in the case of reservations about moral duties, this challenge concerns *rules of reason*, not reasons themselves. Since the worry is about the reality of laws that structure evaluative deliberation, the appropriate response to it is philosophizing, not doubting that we ought to do what we have reason to do.

7.5. Practical identity as a transcendental concept

This chapter started off by identifying a potential problem with Kant's account of reasons, which is that it is unclear how particular persons, with particular cares and concerns, can act on reasons of a purely formal kind. My discussion so far has illuminated one way in which this concern cannot be framed. This is to insist, as Williams does, that no person would have a reason to act on such reasons. The Kantian approach to morality, he argues, "has never succeeded, and could not succeed, in answering the question, *by what right* does it legislate to the moral sentiments?"[27] As I have shown, this objection disregards the fact that for Kant practical reasoning sets its own standards of objective validity. The criteria

---

[27] Bernard Williams, "Preface," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), p. x, original emphasis.

against which maxims are to be judged are entailed by the very activity of adopting maxims – specifically, the activity of practical evaluation. Once these criteria are satisfied, the resulting reasons need no further justification.

However, some of Williams's claims also point to a better way of understanding the initial question. It can be taken to mean not how it is possible for persons to have formal reasons, but how it is possible for persons to be *disposed* to have formal reasons. The query, in this restated form, is a variation of Williams's concern about explaining agents' practical investment. It asks how it is that we can be responsive to the world in a way that enables us to have formal reasons. Since this is the only question that can be asked about the nature or character of the Kantian practical self, it may be called the question about Kantian *identity*. Discussing this topic will require a more decisive foray into Kant's metaphysics than what I have done so far. To begin with, it will be necessary to explain the thesis that basic-level receptivity to the world provides the matter of reasons. This thesis, as will be seen, has some rather surprising consequences with regard to the possibilities for self-knowledge.

First, it is useful to dispel a possible misunderstanding of Kant's claims about practical judgement as essentially evaluative. In particular, it is worth noting a dissimilarity between the role of the concepts of good and evil in practical judgement, and that played by the categories in theoretical judgement. Kant defines judgement in general as "the faculty of subsuming under rules, i.e., of determining whether something stands under a given rule (*casus datae legis*) or not."[28] In the case of theoretical judgements, this means bringing two or more concepts under the rules of synthesis that Kant calls categories.[29] It might seem that the concepts of good and evil are the functional equivalent of the categories in practical judgement. A maxim, as we saw, can only be adopted on the basis of

---

[28] Immanuel Kant, *Critique of Pure Reason*, trans. Allen W. Wood and Paul Guyer (Cambridge: Cambridge University Press, 1998), A132/B71, p. 268.

[29] Categories can be understood as different ways in which concepts can be connected in thought, and include forms of quantity, quality, relation, and modality. In accordance with the categories, for example, the judgement "All candies are sweet" is universal, affirmative, categorical, and assertoric.

a representation of some action as good. Hence, it might be assumed, the overarching concept of goodness is what allows our concepts of objects and actions to be synthesized, or conjoined, in practical thought.

However, this assumption would be incorrect. The categories partly constitute the objects of theoretical reason by imposing a certain structure on the deliverances of the senses, or the "manifold of sensible intuitions". Since they order experience, the categories cannot themselves be derived from, or present in, experience. They are pure concepts of the understanding. By contrast, argues Kant, the concepts of good and evil take no part in constituting the objects of practical reason; instead, they *presuppose* them as given.[30] The role of the pure concept that synthesizes empirical data is, in practical reasoning, played by the Idea of free causality. That is to say, in order for us to represent to ourselves an action as a possible object of our choice, we must first regard ourselves as an uncaused cause. As is the case with the categories, the Idea of freedom is pure, or non-empirical. On the other hand, the concepts of good and evil are empirical in that they can find no application without the presence of some deliverances of our senses. The idea of the empirical nature of evaluative judgements can be reframed in terms of Kant's oft-used distinction between form and content.[31] Although practical judgement is purely formal, consisting in the discernment of the relation between maxims and rules of reason, it must have a basis in some content. Both form and matter are necessary for the activity of judging to be possible.[32] Therefore, in order for it to be possible for persons to have reasons, they must enter the process of practical deliberation with some pre-existing material for judgement – material that can be "subsumed under a rule".

---

[30] Kant, *Critique of Practical Reason*, 5: 65, p. 56.

[31] The dualism of form and matter is one of the cornerstones of Kant's entire critical system. He writes: "These are two concepts that ground all other reflection, so inseparably are they bound up with every use of the understanding. The former signifies the determinable in general, the latter its determination (both in the transcendental sense, since one abstracts from all differences in what is given and from the way in which that is determined)." Kant, *Critique of Pure Reason*, A266/B322, pp. 69-70.)

[32] As Kant memorably puts it, "Thoughts without content are empty, intuitions without concepts are blind." (Ibid., A51/B75, pp. 193-94.)

Material for practical judgement is provided by the agent's desires, which ultimately depend on a modification of the "receptivity belonging to inner sense".[33] At the bottom of each desire lies the person's affection by factors external to him, with regards to which he is passive rather than active. I say "ultimately" and "at the bottom" to avoid exaggerating the point about the affective nature of desires. I am not espousing the view that for Kant desires are straightforward affections. On this view, to say that a person has a desire is to say that he is "struck" by some external object in a way that elicits in him an expectation of pleasure or agreeableness. Importantly, this expectation is not of a reasoned kind: for example, we do not crave chocolate because we recognize that we have a reason to crave it. Rather, desires *impress* themselves on our minds, and we play no part in their creation. Yet Kant does not think that desires invade us from without, but that we rationally conceive them. He thus writes: "The *faculty of desire* is the faculty to be by means of one's representations the cause of the objects of these representations."[34] It is clear from this Kant's remark that he attributes the "pull" of desires to the agent's activity of conceptualizing external stimuli, rather than to his passive surrender to them.

However, it is also implausible to think that for Kant desiring is a pure, or non-sensible, activity, which would render the origins of the content of practical deliberation mysterious. The task for Kant interpreters is therefore to explain how the sensible strain of desiring in Kant's can be maintained without thinking that all desires are just affections. One such explanation has been given by Andrews Reath. He suggests that in Kant pathological affection must be taken to figure in the causal history of every desire.[35] However that does not mean all desire are directly sourced in affection. A pleasure-based desire often leads to activities, interests and forms of human interaction that are then enough to

[33] Kant, *Critique of Practical Reason*, 5: 58, p. 51.

[34] Immanuel Kant, *The Metaphysics of Morals*, ed. Mary Gregor (Cambridge: Cambridge University Press, 1996), 6: 211, p. 11.

[35] Andrews Reath, "Hedonism, Heteronomy, and Kant's Principle of Happiness," in *Agency and Autonomy in Kant's Moral Theory* (Oxford: Clarendon Press, 2006), p. 38.

stimulate desires on their own, without any agreeable modification of the subject's senses. In Reath's example, one's intimate familiarity with the nuances of baseball and a sense of community with other baseball fans may be sufficient to give one grounds for desiring to follow the game, after and independently of the enjoyment one felt on his first visit to the stadium.[36] It is not important here whether Reath's argument succeeds. What matters is the point that underlies his reasoning: that it cannot be assumed that for Kant desires contain no affective element whatsoever. Even if not all desires must be immediately rooted in the modification of the person's sensibility, indirectly affection plays a part in the formation of every desire. For instance, although in Reath's example the fan develops a liking for baseball that does not derive simply from the pleasure of watching the game, it was that initial pleasure that created the conditions for his rational affection. In that sense, there can be no desire without a basis, however remote, in affection.

This last thought offers a way of thinking about practical identity. Since desires are ultimately rooted in sensible affection, there must be an explanation of how this impingement is possible. One possibility would be to say that the explanation can be found in certain features of the objects that entice us. However, for a Kantian this explication does not reach far enough. The fact that *we* feel the tug of desires cannot be properly accounted for if one only assumes that things outside the mind have a certain property – some sort of ability to affect our minds. What would remain unexplained is the *interaction* between the world and our consciousness – that is, the precise way in which objects exert pressure on our consciousness, so as to give rise to desires. Accounting for desires requires postulating an extra factor that enables this union. One way of doing this is to say that the invasive power of external objects must be matched by a corresponding susceptibility to invasion on the part of person. The notion of a Kantian practical identity would then refer to a certain arrangement of an agent's receptive apparatus that allows him to experience desires.

---

[36] On the basis of similar arguments, Thomas Nagel has argued that "many desires, like many beliefs, are *arrived at* by decision and after deliberation." (Nagel, *The Possibility of Altruism*, p. 29, original emphasis.) Desires, on his view, can be motivated as well as motivating.

The concept of identity as the basis for desires will seem disappointing to anyone hoping for a more immediate connection between identity and reasons. However since the basis for the validity of reasons is clear, and has nothing to do with the particularity of any person, the desire-centred concept of the practical self looks like the only one possible. Furthermore, the Kantian perspective requires taking an even more unfamiliar step: regarding identities, including one's own, as perfectly *inscrutable*. The structure of receptivity that underlies a person's desiderative engagement with the world is out of bounds of his cognition. Since this thesis can be understood in two senses it is important to note which one is intended here. On the one hand, the thesis of inscrutability can just mean that a person's identity cannot be represented in his consciousness. This claim relies on the fact that we do not bring our desires into existence. They depend on the initial impingement of our senses by the world, which we cannot spontaneously achieve ourselves. Therefore, the only evidence upon which we can base our knowledge of our identities is the bare sensation of desiring – a sensation which we did not bring about, but found ourselves having. From this bare sensation, in turn, nothing can be inferred about the desiring subject. For instance, learning that I have a desire for chocolate is no basis for self-knowledge because it still leaves me wondering *why* I have this desire – that is, what features of my self make this desire possible. Our passivity with regard to our desires thus cuts off any possibility for self-knowledge.

While this line of reasoning is correct, I am proposing the thesis of inscrutability in a stronger sense: that a person's identity *necessarily* cannot be represented in his consciousness. The difference is that this new reading does not turn on our inability to attain self-knowledge. The weaker interpretation presents the thesis of inscrutability as dependent on the fact that we are mere bystanders to our natural motives and inclinations. By contrast, the stronger claim I am proposing does not rely on what we empirically can or cannot do. It hangs on a requirement of judgement – namely, the fact that the existence of "content" is a condition of the possibility of practical judgements. It is especially important that this content, being a precondition for making judgements, must precede the activity of judging. As Kant notes, the understanding "demands first that something be

181

given (at least in the concept) in order to be able to determine it in a certain way."[37] It then follows that for a person to have a representation of his own identity, he would first need to be presented with some content. However this is impossible, because identity is presumed to refer to the most basic structure of receptivity, the structure that accounts for all given content. In other words, there is no higher source from which the sensibility can receive the substance that is required for representing one's identity. What is the condition of all experience cannot itself be experienced.

Let me conclude by briefly considering how this transcendental concept of identity fits with, and explains, Kant's claim that the concept of happiness is indeterminate. Happiness, Kant maintains, is the sole end that can be predicated of all persons, "so far as they are dependent beings to whom imperatives apply".[38] However, an obstacle to understanding the claim of indeterminacy is Kant's apparent inconsistency in defining happiness. H.J. Paton has for example argued that Kant held two views of happiness that cannot be readily reconciled: one is that happiness consists in the individuals' enjoyment of the maximum amount of pleasure and avoidance of pain; the other holds that happiness refers to the satisfaction of all of his ends, considered as a mutually compatible set.[39] The objects picked out by the two definitions are not the same, and pursuing them may result in different actions. In other words, seeking maximum pleasure can be distinct from trying to fulfil the totality of one's ends, considered as a compossible whole. Sometimes the two pursuits can even be inimical, as when, in one of Kant's examples, a gout sufferer refuses to "kill the enjoyment of the present moment" at the expense of long-term damage to his health.[40] In the choice between Kant's two definitions of happiness Paton squares firmly with the one that emphasises the satisfaction of all of one's ends, rejecting the

---

[37] Kant, *Critique of Pure Reason*, A267/B322-23, p. 70.

[38] Kant, *Groundwork of the Metaphysic of Morals*, II: 415, p. 83.

[39] H. J. Paton, *The Categorical Imperative* (Philadelphia: University of Pennsylvania Press, 1971), pp. 85-87.

[40] Kant, *Groundwork of the Metaphysic of Morals*, I: 399, p. 67.

hedonistic alternative as "inadequate for the description of action".[41] This definition renders the imperative of prudence, which governs the pursuit of happiness, as a principle of "integration": to aim at the satisfaction of one's desires "in a whole organised and systematic life".[42]

There is reason to doubt whether there really are two incompatible views of happiness in Kant. Daniel O'Connor has convincingly argued that there is no clash because while the "integration reading" view represents the definition of happiness, the "maximum enjoyment" view gives the *criterion* of happiness. On this consolidated reading a requirement for adopting ends, which can then be harmonized, is the expectation of pleasure from their achievement.[43] Thus, although the rational activity of combining one's ends remains an important element of happiness, an equally vital aspect of it is "rooted in the sensible order".[44] Paton suppresses this sensible aspect of happiness because he assumes that it would reduce desiring to pleasure-seeking. However, as my earlier discussion of Reath has shown, this assumption is incorrect. It is possible to maintain that all desires are ultimately based on some affection of the sensibility without compromising the thought that desires are rationally conceived.

The sensible aspect of happiness is central to Kant's claim that we cannot form a conception of happiness. He writes that "the concept of happiness is so indeterminate a concept that although every man wants to attain happiness, he can never say definitely and in unison with himself what it really is that he wants and wills".[45] Such a strong claim would not be warranted if happiness were to be understood merely as the systematic satisfaction of one's ends. On this view, forming a conception of one's happiness would be *difficult*: the trouble would be to imagine how precisely one's various ends can be combined without clashing. But Kant says that "it is *impossible* for the most intelligent, and at the same time

[41] Paton, *The Categorical Imperative*, p. 85.

[42] Ibid., p. 87.

[43] Daniel O'Connor, "Kant's Conception of Happiness," *Journal of Value Inquiry* 16, no. 3 (1982): pp. 189-91.

[44] Ibid.: p. 190.

[45] Kant, *Groundwork of the Metaphysic of Morals*, II: 418, p. 85.

most powerful, but nevertheless finite, being to form here a determinate concept of what he really wills."[46] The key to his idea is that we are finite beings, which means that we have empirical needs that implicate us in the world of sense.[47] We are bound to seek satisfaction from the world outside us, and do so through desiring.

However, as I argued earlier, our sensible basis for this interaction with the world, our identity, must remain obscure to us. This inability to access one's identity explains why happiness is an indeterminate concept. Happiness requires the satisfaction of all desires, and yet we cannot know anything about the way in which our receptive faculty is constituted. Hence, we can strive to attain what our desires demand, but can have no guarantee that this will bring us satisfaction. Since we are at the most fundamental level opaque to ourselves, we cannot be sure that attending to one desire will not inconvenience us in ways that we cannot comprehend, and thereby produce *new* needs and desires. It is therefore an illusion to think that we can form a conception of our happiness because we have no way of knowing what makes us happy. We are caught in a cycle of second-guessing ourselves — a cycle in which we must be frustrated because our identities, the sensible condition of the possibility of desires, cannot itself be intuited.

---

[46] Ibid., II: 418, p. 85, emphasis added.

[47] Cf.: "The human being is a being with needs, insofar as he belongs to the sensible world, and to this extent his reason certainly has a commission from the side of his sensibility which it cannot refuse, to attend to its interest and to form practical maxims with a view to happiness in this life and, where possible, in a future life as well." Kant, *Critique of Practical Reason*, 5: 61, p. 53.)

8. Conclusion: Transcendental identity and legitimacy

8.1. The normative significance of transcendental identity

The problem with which this thesis started is that current concepts of identity are unclear and undertheorized. Taking into account this shortcoming of available treatments of identity, I have examined the possibilities for developing a concept on identity from the perspective of reasons-based liberalism. Specifically I have evaluated those approaches to identity that regard it as a determinant of individuals' reasons for action. An argument that is inspired by John Rawls states that identities generate reasons insofar as individuals care to maintain their self-respect. However I showed that Rawls's "social bases of self-respect" cannot be understood in the way that is required for the argument to succeed. I then considered Harry Frankfurt's claim that identities give rise to constraints on individuals' power of willing. The claim fails, I suggested, because Frankfurt cannot show that it is truly impossible for us to disregard our identities. I also examined Christine Korsgaard's view of identities as the subjective ground of maxim-adoption, and argued that it vacillates between an unacceptably naturalist understanding of identity and one that cannot account for persons' commitment to particular identities. Following these criticisms, I suggested that the link between identities and reasons needs to be investigated by reconsidering the notion of a reason for action. I argued that for a conception of reasons to be acceptable to liberals, it must present reasons as universal in scope. This condition is met by Kant's conception, according to which reasons consist in the conformity of maxims with certain objective principles. Finally, this account allowed me to present identity as a sensible condition for the possibility of universal reasons.

The substantive argument of this dissertation ends with the elaboration of the transcendental concept of identity. In this final chapter I will offer some speculative remarks about the implications of that concept. These remarks should be taken as pointers towards the wider relevance of my findings, rather than as elements in an argumentative case. In this spirit, I would like to address the following worry about my concept of transcendental identity: that it seems normatively insignificant, even though it claims connection with the normative theory of reasons-based justification.

Consider, first, one way in which my claims about identity are weaker than those made in the ordinary identity-discourse. I have proposed thinking of a person's identity as a certain arrangement of the receptivity belonging to his "inner sense". It is a sensible condition of desiring, which ultimately provides the material for practical reasoning. However, my claims about transcendental identity do not show that each person has a *unique* identity that is different from everyone else's. To make sense of desiring it is necessary to assume that the receptive apparatus of each person is arranged in a certain way – so that he is susceptible to affection by the world outside his mind. But this claim does not warrant the conclusion that basic receptive arrangements differ from person to person. Against this claim one might argue as follows: different individuals have different desires, which can be taken as indicative of differences in these individuals' basic conative constitutions. However this line of argument is unpersuasive. The reason for this is that even though desires depend for their existence on initial sensible affection, desires themselves are reasoned. They are conceptual representations of external stimuli. Being reasoned, desires are agents' creations rather than externally given attributes of the person. Therefore, the fact that different people desire different things is not a sure sign of differences in their receptive apparatus.[1]

---

[1] It is interesting to note that Williams also denies that his argument about practical character establishes the distinctiveness of characters across persons. The idea of character is necessary, he argues, for a person to think that he has a reason to live his life rather than resign himself to passivity. And yet "that is compatible with these drives, and his life, being much like others'". (Williams, "Persons, Character and Morality," p. 15.) Williams's claim is that it must be

Although it is widely, albeit implicitly, assumed that identities are unique, they might ground normative claims even if they are not. Even if we are all pretty much the same with regard to our conative constitutions, it might still be the case that those constitutions warrant public protection. However, I have advanced a further claim about identities that rules out the possibility of making this case. A person's identity, as the sensible condition of desire-based reasons, is *inscrutable* to himself and others. Even if our reasons bottom out at our identities, the configuration of our receptivity is inaccessible to us because there is no standpoint from which it could be experienced. This precludes the possibility of any normative claims that are directly based on identities: one cannot maintain that his identity depends on preserving a certain body of language or customs, or complain that an obscene work of art offends against his identity, or petition against lax immigration laws on the basis that they endanger his national identity, or that impartial norms are too demanding because they run roughshod over his identity. None of these claims can be made because they rely on the assumption that a person is in a position to discern, directly and especially intimately, his identity. I have argued against this assumption: at the most basic level, which concerns our sensuous engagement with the world, we cannot know ourselves.

Given that identity cannot directly ground any normative claims what, if anything, can be normatively significant about it? In this chapter I would like to suggest that its normative relevance lies in the fact that is structures our thinking about political norms. In this respect, the role of transcendental identity can be usefully compared to the one played by self-respect in Rawls' theory of justice. As I noted in chapter 3, although there is great temptation to interpret Rawls as espousing a psychologistic notion of self-respect, his account of self-respect is restricted to the context of the original position. It is a property of the parties choosing principles of justice behind the veil of ignorance, not of citizens in real-life circumstances. Therefore, there is no way of justifying public demands by appealing to the contribution of some policy or measure to persons' self-respect.

---

supposed that each person must have some grounds for going on but this does not necessarily mean that these grounds are singular or exceptional.

However the concept of self-respect fulfils an indispensable constructive function in Rawls. It is part of the procedure that derives and justifies his principles of justice, which are meant to apply in actual political conditions. Specifically, the notion of self-respect answers the question of why contractors engage in any choosing: they do so, argues Rawls, because they care to preserve their sense of the worth of their deliberation and choice. Without settling this question, Rawls's theory could not as much as get off the ground. Thus, although the self-respect of particular, flesh-and-blood persons cannot be invoked to justify laws and policies, the *general idea* of self-respect is nonetheless normatively significant.

I would like to suggest that the general idea of transcendental identity is normatively significant in a similar sense, only not in relation to conceptualizing justice but *legitimacy*. This should not be surprising as it follows the guiding idea of the reasons-based approach to justification. As I noted in chapter 1, the guiding thought behind that approach is that only political norms that can be supported by reasons that hold for all can be legitimately enforced upon all. On this liberal view, political rule involves the exercise of power, an important part of which is the subordination of the citizens' wills to public decisions. Once laws and various other instruments of political rule are in place, subjects are expected to comply with them regardless of, and sometimes even contrary to, their choosing. The search for legitimacy is the search to answer what, if anything, makes this subordination of wills acceptable. It should be noted that what stands in need of justification is not the exercise of political power *per se*, understood as the enforcement of political rules. The liberal question about legitimacy does not come from the anarchist corner, which regards all power as intrinsically suspect. It is only a particular feature of political rule that makes it normatively suspect: the fact that its commands sometimes enjoin persons to act against their own judgement – to refrain from avenging wrongs done to them by others, to tolerate public criticisms of their religion, and so on. As Nagel says, "The real problem is how to justify making people do things against their will."[2]

---

[2] Thomas Nagel, "Moral Conflict and Political Legitimacy," *Philosophy & Public Affairs* 16, no. 3 (1987): p. 224.

The liberal answer to the problem of legitimacy is to insist on the justifiability of political measures (or at least, of the basic norms that ground these measures) to all citizens. If one is subjected to commands supported by reasons that are valid for him, so the argument goes, he has no right to complain about his subjection. The rule over him is legitimate, and he has an obligation to obey it. Various aspects of this argument demand closer inspection, but I am here principally interested in whether the concept of transcendental identity has any place in it. My claim is that it does, in the sense that it structures our thinking about legitimacy. To be more specific: the transcendental concept of identity is part of the explanation of why individuals have a claim to a legitimate political order. The connection between legitimacy and identity can be presented through a sequence of three Kantian claims:

(1)   Individuals' claim to a legitimate political order is entailed by their innate right to external freedom.

(2)   Individuals' external freedom is their ability to act on their sensibly conditioned judgements unimpeded by the deliberate interference of others.

(3)   Transcendental identity is a condition of the possibility of sensible affection.

Identities thus make demands for legitimacy possible in an indirect fashion: they ground the possibility of desire-based external freedom, the concept of which entails the requirement for establishing a legitimate political order.

While claim (3) was discussed sufficiently in the previous chapter, the other two claims that compose this argument need further elucidation. However, engaging with them extensively would go beyond the scope of this chapter as the issues they raise are too diverse and complex to be dealt with quickly. It does not help that Kant's *Metaphysics of Morals*, from which (1) and (2) are drawn, is a book that is notoriously difficult to interpret. Its understanding is hindered not only by Kant's characteristically dense writing style but also by its fragmentary nature. In what follows I will therefore present only a very brief sketch of the concepts of

"external freedom" and a "claim to legitimacy", as much as is necessary for appreciating the normative relevance of transcendental identity. I will take up this topic first. Then I will turn to an aspect of the above sequence that I can afford to examine in greater detail as I have already foreshadowed it in the previous two chapters. This is the idea that practical judgement is sensibility-driven. I will explain why this point is important for Kant's approach to legitimacy and how my interpretation of Kant accommodates it.

8.2. External freedom and legitimacy

Kant's view of what occasions the demand for legitimacy can be usefully contrasted with a more popular approach to legitimacy. According to this liberal approach, individuals have a claim to a justification of state laws and policies because their enforcement encroaches upon their ability to design their lives in accordance with their choices. The "political subjugation" of citizens, as Jean Hampton calls it, is that they are subjected to state commands, backed by the threat of coercion, that limit their ability to lead their lives as they see fit.[3] These constraints on individuals' rights and liberties are often, even if only implicitly, conceptualized in terms of natural rights – to self-determination, to the free enactment of one's judgement, or something like it. A conception that is the unstated background of many present-day treatments of legitimacy is John Locke's theory of the natural right to self-government.[4] In Locke, individuals possess an unalienable prerogative to freely conduct themselves, the grounds for which are found in their pre-political obligation to maintain themselves as God's creatures. Stripped of its original religious grounding, the Lockean presumption of individuals' right to non-interference by others and the state is the basis of much work on legitimacy today.

---

[3] Jean Hampton, *Political Philosophy* (Boulder: Westview Press, 1997), pp. 3-4.

[4] John Locke, *The Second Treatise of Government*, ed. J. W. Gough, 3rd ed. (Oxford: Blackwell, 1966).

The claim about natural rights as the basis for the demand for legitimation is not uncontentious. Not only is it not clear whether it can be sustained without recourse to Locke's, or some other, thick metaphysical outlook, the plausibility of thinking that pre-civil rights apply in the civil condition is also not obvious. Kant thus thinks, like Hobbes, that there is a sharp discontinuity between the pre-civil and the civil condition, such that rights that obtain in the state of nature do not carry over into the political context. Hence both thinkers, controversially, reject the idea of a moral right to revolution. Putting this point aside, it is here more important to explain how Kant diverges from thinking that the claim to legitimacy arises in virtue of the invasiveness of political rule. Instead, he maintains that the demand for legitimate political relations follows from the persons' right to exercise external freedom. For Kant a person's external freedom consists in "other agents' deliberate non-interference with the agent's recognized capacity for rationally self-directed choice and action."[5] External freedom needs to be distinguished from several other terms that Kant uses when talking about freedom. It needs to be kept distinct from the concept of *internal* freedom, which in the *Metaphysics of Morals* Kant uses in the same way as the term *positive* freedom elsewhere. In the *Groundwork* and *First Critique* positive freedom refers to the will's capacity to start a causal chain in nature by means of a representation of a pure practical law. What makes this type of freedom positive is that its operation is not arbitrary: it is constrained by the lawlike form and the various other stipulations that Kant specifies in the formulae of the categorical imperative. By contrast, *negative* freedom in Kant is just the capacity of the will to reach decisions interference from outside, which is achieved by acting on maxims. This is what we saw, in chapter 5, Kant calling the "freedom of the turnspit" since it does not rule out the possibility that the will is in conditioned from the outside. All it requires is that choice does not feel unconditioned from

[5] Katrin Flikschuh, *Freedom: Contemporary Liberal Perspectives* (Cambridge: Polity, 2007), pp. 93-94. Kant's definition of external freedom is that it involves the "following authorizations, which are not really distinct from it...: innate *equality*, that is, independence from being bound to others to more than he can in turn bind them; hence a man's quality of being *his own master* (*sui iuris*), as well as being a man *beyond reproach* (*iusti*)...; and finally, his being authorized to do to others anything that does not in itself diminish what is theirs, so long as they do not want to accept it..." (Immanuel Kant, *The Metaphysics of Morals*, 6: 238, p. 30.)

the first-person perspective. In sum, it can be said that a person is negatively free whenever he acts on a maxim, whatever that maxim is. He is positively or internally free when his maxim conforms with the rule of pure reason, or the categorical imperative.

Kant regards the right to external freedom as innate, or natural. However it serves to ground a conception of political legitimacy that is fully political, by which I mean that it does not appeal to normative criteria that are independent of the relations within the political society. The transition from a natural right to a civil conception of legitimacy relies on two features of external freedom: its relational, or social, character, and its basis in materially conditioned desires. Unlike the concepts of positive (or internal) and negative freedom, external freedom requires regarding an agent in relation to other reasoners. More specifically, it concerns the relationship between the deliberate actions of individuals and the deliberate actions of others. As Kant says, external freedom is a "reciprocal relation of choice", that is, "a practical relation of one person to another, in so far as their actions, as facts, can have (direct or indirect) influence on each other".[6] This definition excludes the relations between persons and natural factors: it cannot be said, for example, that handcuffs or a debilitating illness can limit a person's external freedom. The effect of others' unintentional behaviour is also irrelevant to a person external freedom. A person that comes to block my exit from a bus as a result of the vehicle's sudden movement does not present an obstacle to my external freedom.

The interdependence of individuals' choices that is at the heart of external freedom is a function of the desirative basis of human choosing. I should note that the suggestion I am making here relies on a contentious interpretive claim. My assumption is that the choosing that is relevant to Kant's concept of rights is of a kind with the deliberation that he discusses in his moral philosophy. However, this interpretative claim appears suspect in light of the difference in the status of practical principles that govern these two kinds of choosing. Principles of public Right, which have to do with the proper relation between persons, are

---

[6] Kant, *The Metaphysics of Morals*, 6: 230, pp. 88-89.

enforceable, whereas principles of morality, which concern the proper disposition of the human will, are essentially unenforceable. This division, it seems, warrants thinking that politically relevant agency falls outside the remit of moral agency. Strict right, Kant says, "is not mingled with anything ethical."[7] My suggestion, however, is that that the two types of agency share the same structure: both consist in forming rational representations on the basis of sensibly conditioned material provided by desires. An argument for the idea that moral agency displays this structure was in the background of my discussions in the previous two chapters. I will make it more explicit, with reference to the issue of practical judgement, below. The argument that politically relevant agency is a desire-driven but nonetheless rational activity cannot be discussed here. Examining it would require a sustained exegesis of obscure and highly technical passages in Kant, which would be distracting in the present context. The interested reader can consult the work of Flikschuh for a thorough interpretation of Kant that brings economic desires under the umbrella of morality.[8]

The innate right to freedom is exercised by acting on the basis of one's sensibly conditioned desires. However, this activity is subject to an inevitable constraint. Rawls has popularized, drawing on Hume, one view of limitations that apply to individuals' reasoning outside the civil condition.[9] Under the "circumstances of justice" no person is guaranteed to succeed in his life-plans because of the moderate scarcity of goods and the disinterest of others in his well-being. For Kant, however, the limitations on persons' external freedom are not empirical but given prior to experience. His core claim is that desires always aim at the attainment of external objects, which exist within a spatio-temporally bounded domain, the spherical Earth. The fact that humans must seek the satisfaction of their desires in a domain that does not stretch infinitely in space and time is an

---

[7] Ibid., 6: 233, p. 25.

[8] See Katrin Flikschuh, *Kant and Modern Political Philosophy* (Cambridge: Cambridge University Press, 2000), pp. 80-112.; and Flikschuh, "Kantian Desires: Freedom of Choice and Action in the *Rechtslehre*."

[9] Rawls, *A Theory of Justice*, pp. 126-130.; Hume, *An Enquiry Concerning the Principles of Morals*, III.1.; David Hume, *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (Oxford: Oxford University Press, 2000), III.ii.2.

important determinant of their quest. It means that in looking to realize their desire-based projects persons run up against others, looking to realize theirs. Furthermore, they must necessarily do so: the choice of others is an inevitable limitation on persons' practical pursuits. It is an objective constraint, which means that it does not arise from the properties or the functioning of the human will, in which case it would classify as subjective. Rather, it is a condition that restricts desire-based reasoning "from the outside" – independently of the first-person perspective of reasoning agents.

Kant's account of the claim to legitimacy arises from his characterization of the constraints on agents' external freedom. The formation of a legitimate political order is the resolution of the tension that is implicit in the concept of external freedom. While agents have an innate right to freedom, they cannot realize that freedom because of the *a priori* constraints issued by the presence of others. The first right of individuals in the state of nature thus, once taken to its logical conclusions, calls for the establishment of a system of *political* rights, guided by the following Universal Principle of Right: "Any action is right if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law."[10] Kant's thought is that while the claim to legitimacy arises out of the natural right to external freedom, its realization requires overcoming natural rights and stepping into the civil condition. The natural right anticipates its own abolishment.

8.3. Desires and judgement

I have suggested that the normative significance of transcendental identity lies in the fact that it (indirectly) grounds the possibility of a claim to legitimacy within Kant's normative system. This claim relies on the idea in Kant that moral and political agency share the same structure, in the sense that they are both, at bottom, sensibility-driven. Although it has not been possible to discuss this claim

---

[10] Kant, *The Metaphysics of Morals*, 6: 230, p. 24.

in relation to political agency, I am in a position to elaborate and defend it with regard to moral, or more generally practical, agency. I will present my reading of Kant as a friendly amendment to Onora O'Neill's account of practical judgement.

My revision of O'Neill can be approached by first considering her opposition to an interpretation of Kant that understands his views on judgement on the model of reflective judgement. A good representative of that view is Alessandro Ferrara, who draws on Hannah Arendt's reading of the *Critique of the Power of Judgement*.[11] This reading gives special weight to the fact that politics, understood as an essentially social activity that often revolves around struggle for political power, was not of much interest for Kant. The works in which Kant writes about politics are in fact centrally concerned with issues in the metaphysics of law (or Right), philosophy of history, and anthropology. This led Arendt to remark that Kant "never wrote a political philosophy",[12] and to reconstruct from his views on judgement a political philosophy that he *should* have written. His ethical treatises, writes Arendt, cannot serve as a guide in this enterprise because they concern "the conduct of the self in its independence of others".[13] By this Arendt means that human plurality plays no part in determining the content of Kant's moral principles. For Kant, Arendt suggests, human interaction is a necessary prerequisite for the existence of morality insofar as it gives rise to individuals' motives and incentives that moral standards must suppress. However, what those standards enjoin appears to have nothing to do with the fact of human sociability.[14]

---

[11] See Alessandro Ferrara, *The Force of the Example: Explorations in the Paradigm of Judgment* (New York: Columbia University Press, 2008).; Alessandro Ferrara, *Justice and Judgment: The Rise and the Prospect of the Judgement Model in Contemporary Political Philosophy* (London: SAGE, 1999).; Hannah Arendt, *Lectures on Kant's Political Philosophy*, ed. Ronald Beiner (Brighton: Harvester, 1982).

[12] Arendt, *Lectures on Kant's Political Philosophy*, p. 7.

[13] Ibid., p. 19.

[14] A similar line of argument, which charges Kant with espousing a "monological" account of morality, is pursued by Jürgen Habermas. See for example Jürgen Habermas, *Moral*

One Kant's distinction that is especially emphasised in Arendt's revisionist reading of Kant, and which also figures centrally in Ferrara's account of judgement, is that between reflective and determinant judgement. After calling judgement the faculty of thinking "the particular as contained under the general",[15] Kant writes:

> The power of judgment can be regarded either as a mere faculty for reflecting on a given representation, in accordance with a certain principle, for the sake of a concept that is thereby made possible, or as a faculty for determining an underlying concept through a given empirical representation. In the first case it is the reflecting, in the second case the determining power of judgment.[16]

In the case of determinant judgement the universal, a rule of correct action, is given and the task of judgement is to find particulars, specific acts that fall under it. In reflective judgement, by contrast, only the particular is given and a rule must be found in terms of which that particular can be understood and assessed. In the first case judgement is a matter of algorithmic subsumption of an act under an existing principle, as when one know what conditions make a good chair and then asks whether some specific chair satisfies them. In the second case judgement is creative: when a person contemplates an action for which no determinate norm exists, he is compelled to make one for himself in order to be able to evaluate that action. For example, there are no hard and fast rules specifying what counts as bravery in battle. When assessing the conduct of a soldier, therefore, one has to reconstruct from his actions a principle that will

Consciousness and Communicative Action, trans. Christian Lenhardt and Shierry Weber Nicholsen (Cambridge: Polity Press, 1990).

[15] Immanuel Kant, *Critique of the Power of Judgment*, trans. Paul Guyer and Eric Matthews (Cambridge: Cambridge University Press, 2000), 20: 202-203, pp. 8-9.

[16] Ibid., 20: 211, p. 15.

allow one to determine whether his behaviour was brave, or foolhardy, or cautious, or cowardly.

Ferrara argues that reflective, rather than determinant, judgement constitutes a proper "model of validity" for conceptualizing justice in the aftermath of the so-called linguistic turn. The principal contribution of the linguistic turn, as Ferrara conceives it, to political philosophy was that it illuminated the dependence of norms on the conceptual schemes within which they originate. This point discredits "foundationalist" normative approaches, which start from facts or principles that are supposedly immutable and self-contained. In one form or another, foundationalist theories are grounded in what Williams called the "absolute conception of reality" that is independent of the cognizing subject. In opposition to foundationalist theories, philosophers such as Wittgenstein, Quine and others argued that all thinking is conceptually laden. In this vein, for example, Wittgenstein rejects the "Augustinian" theory of meaning. As we saw in chapter 1, the disagreement between realists and nominalists is over whether abstract objects exist, or whether abstract terms are just names. Nonetheless, the two positions agree that all words refer to objects. This view of reference is mistaken, argues Wittgenstein, because the claim that words pick out things in the world assumes that the world is already divided up into discrete things for us to name.[17] But the assumption of the basic orderliness of the world, which our language is supposed to map onto, is arbitrary and illegitimate. This unwarranted presumption can be avoided, Wittgenstein suggests, by regarding words in light of the intention behind their use, or the practical point of view of the speaker. This leads to a picture very different from that which presumes a pre-existing orderliness of the world: "How we group words into kinds will depend on the aim of the classification—and on our own inclination."[18] Categorial discriminations constitutively depend on the points of view, plans, and interests

---

[17] Wittgenstein writes: "When we say: 'Every word in language signifies something' we have so far said *nothing whatever*; unless we have explained exactly *what* distinction we wish to make." (Wittgenstein, *Philosophical Investigations*, p. 7, original emphases.)

[18] Ibid., p. 8.

of the speakers.[19] Their meaning is given by the function they play in "language games" – in the activity of speaking, understood as "a form of life."[20]

The point about the conceptual ladenness of language, argues Ferrara, applies also to moral principles. Kant views moral reasoning, Ferrara suggests, as a species of determinant judgement: its hallmark is the subsumption of particular maxims under the constant, *a priori* rule that is the categorical imperative. However Kant's view must be revised. There are no fixed, universally valid rules under which maxims can be brought because all moral standards are expressions of, and hold only within, "holistic frameworks that are situated and do always come in the plural."[21] The solution, according to Ferrara, is to understand moral reasoning on the model of reflective judgement, as the process of transcending the particularity of one's position but without succumbing to the illusion of a "view from nowhere". This is the outlook which he attributes to Rawls, Habermas, Ackerman, Larmore and most other leading figures in contemporary liberalism. The project is one in which justice does not demand absolute impartiality, but only impartiality "with respect to a finite set of actual conceptions of the good".[22] Although this groundedness of the philosophical enquiry in some specific social and political context is inescapable, it does not require surrendering the striving for rules that are in some sense objective. However it is a particular kind of objectivity, in which practical reason constructs, out of the material of empirically situated claims, the principles for assessing those same claims.

While Ferrara's reflective model of normative validity may be useful as a reconstruction of much of contemporary liberal theory, it is unsatisfactory as an

---

[19] For example, writes Wittgenstein, there is more than one way of viewing the array of handles that one finds in the cabin of a locomotive: the engineer will be able to discern and distinguish the handle of a crank, the handle of a switch, the handle of a break-lever, the handle of a pump, and so on. For the layman, there will be nothing to set the various handles apart. (Ibid., p. 7.)

[20] Ibid., p. 11.

[21] Ferrara, *The Force of the Example: Explorations in the Paradigm of Judgment*, p. 19.

[22] Ferrara, *Justice and Judgment: The Rise and the Prospect of the Judgement Model in Contemporary Political Philosophy*, p. 3.

account of practical judgement. In one sense, his reading of Kant on judgement cannot be accused of interpretative inaccuracy because it is explicitly revisionary. However, the revision must not be so drastic that it contradicts the central tenets of Kant's philosophical system. It is in this sense that Ferrara's interpretation is unacceptable, since it runs counter to a distinctive and important aspect of Kant's critical philosophy: his idea that practical reason, unlike its theoretical counterpart, creates its own object.[23] In Kant, theoretical reason is concerned with determining what there is, in which activity it is compelled to rely on the deliverances of the senses. Its access to its objects is not direct but mediated by sensible intuitions, which is to say that reason in its theoretical employment seeks to apprehend, not manufacture external reality. Principles of theoretical reason – categories or pure concepts of the understanding – are therefore consigned to a regulative role. They specify how the understanding must order given intuitions in order for representations of object to be possible. The determinations of practical reason, on the other hand, "do not have to wait for intuitions in order to receive meaning".[24] The objects of practical reason – the concepts of good and evil – refer to disposition of the will, with regard to which reason can be active. The good will is not something that a person cognizes and from which he is therefore necessarily at a remove. It is something that he brings about, using rules of reason as constitutive of practical reality. Principles of practical reason "have objective reality in their practical use, that is, in the moral use."[25]

In reflective judgement, however, the particular is given, not created by the activity of judging. The task of reflective judgement is to ascend to a type from a present token – to extract a practical rule from an act that has already happened. Hence, contrary to Ferrara's claims, it cannot be regarded as an instance of *practical* judgement. For example, the reflective judgement of a soldier's bravery presupposes that he has already committed certain actions in battle that cannot be

---

[23] See Kant, *Critique of Practical Reason*, 5: 65-66, pp. 56-57.; Kant, *Critique of Pure Reason*, A807-808/B835-836, p. 678.

[24] Kant, *Critique of Practical Reason*, 5: 65-66, pp. 56-57.

[25] Kant, *Critique of Pure Reason*, A807-808/B835-836, p. 678.

subsumed under a hard and fast standard. Unlike what Kant thinks of practical reasoning, this kind of reflection is compelled to work within the boundaries of what there is. The orientation of reason is retrospective, concerned with adjudicating past acts, rather than prospective, aimed at producing action in accordance with laws that specify what there ought to be. Onora O'Neill writes:

> Both determinant and reflective judging are forms of theoretical judgement that may be used when a particular is given: but they have no role when no particular is (yet) given. No amount of theoretical judgement will help us when the task is practical, and we have to judge what to do... *For here it is in the nature of the case that the particular is not given.*[26]

Avoiding the assimilation of the practical standpoint to a "retrospective, spectator perspective"[27] is a central aim of O'Neill's interpretation of Kant on practical judgement. This implies abandoning the ambition to understand practical deliberation on the model of either determinant of reflective judgement. One may note a superficial resemblance between O'Neill's construal of practical judgement and how Kant describes determinant judgement. O'Neill thinks that practical judgement is akin to subsumption in that its major premise is given. That premise is either a moral duty or some non-moral principle of action that the person makes for oneself – a principle that expresses the person's long-term commitments, plans, and interests, such as to bring about the communist revolution, or to master chess.[28] The task of practical judgement is then to bring one's particular actions under these universal rules. However the resemblance between this activity and determinant judgement breaks down in two ways. First, for O'Neill, practical deliberation is forward-looking: it looks to bring about actions rather than assess past acts from a detached and impassioned standpoint.

---

[26] Onora O'Neill, "Experts, Practitioners, and Practical Judgement," *Journal of Moral Philosophy* 4, no. 1 (2007): p. 159, original emphasis.

[27] O'Neill, "Rationality as Practical Reason," p. 95.

[28] The task that individuals face when employing practical judgement is "to search for ways of acting which meet the constraints of all the principles (and the varied goals) which they see as important." (O'Neill, *Bounds of Justice* (Cambridge: Cambridge University Press, 2000), p. 60.)

It does not pass verdict on what has already happened but determines what ought to be done. Second, practical judgement is not algorithmic. A determinant judgement that brings some specific act under an exhaustive and fixed norm allows no variance of outcome. Whether the act either falls under the norm or not, the deliberator has no latitude in judging. For example, the question of whether a student deserves to be awarded a degree under University regulations usually leaves no discretionary scope for the awarding body. If the student has fulfilled all the necessary requirements he automatically gains the right to the degree, and vice versa. But practical judgement, argues O'Neill, is different. Rather than being a blind application of principles, it is an activity that leaves scope for deliberators' creativity.

O'Neill's point about the non-algorithmic nature of practical judgement is based on a particular view she takes of practical principles. Appealing to Kant, she presents both moral and non-moral principles as constraints on agents' various pursuits that do not fully determine action. I already discussed and opted against the interpretation of Kant's rules of reasons as side-constraints in chapter 7. Here I would like to focus on the idea that rules of judgement are indeterminate. According to O'Neil, practical principles do not "regiment action"; they "recommend types of action, policy and attitude rather than providing detailed instructions for living. They usually specify no more than an aspect of action, and this often quite vaguely."[29] Norms, whether they are moral or non-moral, specify how we must not act: we must not lie, smoke in bed, move the rook diagonally in a chess game, hurt our loved ones, and so on. Since they can be honoured in a multitude of ways, these negative prescriptions leave open the issue of what precisely we are to do. This is not true, of course, when there are only two available courses of action, one of which is ruled out by the normative prohibition; then the negative requirement fully determines action (albeit indirectly). However, argues O'Neill, such cases are not paradigmatic of practical principles. Even when norms leave few possible courses of action, judgement is required for deciding between them – for bridging the gap between principle and action. For example, even when we are asked not to move the rook

---

[29] Ibid., p. 53.

diagonally, this still leaves us having to decide whether to move it horizontally or orthogonally. Typically, though, the task of judgement is more complex, with more possibilities for enacting any given principle. This is compounded by the fact that individuals usually judge what to do in light of various principles. More often than not, there is more than one requirement that bears on a situation, and making up one's mind requires striking a balance between them, or compromising some in favour of others. Judgement is then "a question of finding ways of meeting multiple constraints and recommendations set by a plurality of principles."[30]

O'Neill's account of practical judgment is attractive on several levels. It presents practical deliberation as a genuinely forward-looking activity, rather than the impassive contemplation of past acts. It also suitably explains how acting on practical principles can leave room for the agent's creativity. Finally, it displays sensitivity to the complexity of practical reflection by stating that appraising a prospective action typically involves consideration of diverse and competing principles. Nonetheless, it suffers from a phenomenological shortcoming: her claim that every act of judgement is done on the basis of the person's pre-existing principle seems untrue as a general description of how persons actually judge. To clarify my point it is necessary to distinguish two possible senses of the claim that judgement requires pre-existing principles. First, it can be taken to mean that an act of judgement requires principles that are *conceptually* prior. This claim draws on Kant's thesis, in the *Groundwork*, that practical reasoning must be regarded as law-governed. Kant writes: "Everything in nature works in accordance with laws. Only a rational being has the power to act in accordance with his idea of laws that is, in accordance with principles and only so has he a will."[31] Therefore, it is necessary for some principle to be prior to judgement because it enables judging as a reasoned activity. This idea is certainly present in O'Neill's account, albeit mostly implicitly, but it is not the one I wish to challenge.

---

[30] Ibid., pp. 62-63.

[31] Kant, *Groundwork of the Metaphysic of Morals*, II: 412, p. 80

My issue is with the claim of the priority of principles in another, *temporal* sense. This claim, which O'Neill' repeats frequently, is that persons enter deliberation with principles of evaluation that they gave themselves at a prior time. From the phenomenological standpoint, there is something odd about the picture in which persons have the most important part of the practical task figured out before they even get down to judging – in which they survey their options with a rulebook in their hand, as it were. This makes practical reasoning into a strangely pre-meditated and impassive activity, even if it is not algorithmic. Even if a judgement must make use of a conceptually prior rule in order to be reasoned, the thesis of the temporal priority of long-standing personal policies or resolutions seems dubious. Many of our decisions are made on the basis of personal principles that are created on the spur of the moment, tweaked and modified at the time of deciding, or without having settled in advance on what sort of behaviour to commit to. A significant part of practical reflection seems to be making one's principles as one goes along. O'Neill view of judging as instantiating indeterminate but nonetheless previously fixed rules appears incapable of accommodating this fact.

The artificial assumption of temporally pre-existing principles can be avoided by maintaining that neither of the two components of judgement – particular actions or universal rules – takes logical precedence in the agent's consciousness. The starting point for practical reflection is not the person's awareness of an act or maxim for which a rule must be constructed. Nor is it his awareness of a rule with which acts must be brought into accord. The particular and the universal element in practical judgement enter the agent's first-person perspective simultaneously as they are dialectically related. As I argued in chapter 6, the particular element, a maxim, can only be conceived of in terms of a universal rule of reason. Since for Kant practical reasoning must be evaluative, and the only evaluative principles that express necessary goodness are rules of reason, a maxim just is a subjective principle that claims conformity with one or more rules of reason. Conversely, a rule of reason, as the universal element in practical judgement, can only be conceptualized with reference to a particular maxim. It does not have an existence that is separate from agent's maxims. Rather, rules of reason are valid as requirements on action only because there are maxims that

utilize the concept of goodness. Even the categorical imperative applies to us only because our subjective principles of volition take the evaluative form.

The acknowledgment of the dialectical relationship between maxims and rules of reason helps in identifying what is creative in practical judgement. Practical creativity consists in the fact that judgement is a problem-solving activity, in which the problem is set by our sensible nature – this is the regard in which our practical judgement is sensibility-driven. The claim I am making here borrows something from Korsgaard's moral psychology, although it also departs importantly from her position. What I inherit from her outlook is the view that as desirative beings we are always and necessarily implicated in the world of sense, which sets us the task of orienting our reasoning. What is implausible about Korsgaard's account is her view that our implication in the world of sense directly delivers desires. She maintains that desires, caused by the world impinging upon our senses, are the primary element in practical judgement – they are what is given prior to judgement, and with which judgement must work. Therefore, on Korsgaard's view, the guiding question of practical reflection is "How should I rationally respond to my various externally given inclinations, cares, and concerns?"

However Korsgaard's view of the context in which maxims are considered is undermined by the implausibility of her interpretation of desiring. It cannot be assumed that sensible affection forces desires upon us. Instead, sensible affection presents persons with material whose conative status they must determine. The starting points for judgement are "raw" deliverances of the "inner sense" that precede, and make possible, all practical experience. These are biddings from the outside in relation to which the person must take a stand. As such, they are the practical equivalent of the sensible manifold in theoretical cognition. Once presented with some impingement of his "inner sense", the person has a problem: to determine, by means of conceptual representation, whether he enjoys it and wants more of it, whether he wants it to stop, or is indifferent. In other words, he must determine whether the sensation is one whose future occurrence he desires or not. It is only on the basis of this determination that both the particular and the universal element of practical judgement – a maxim and a rule of reason – can

jointly come into being. This is not to say that a conative representation of a sensible affection *must* be a prelude to practical judgement. As I noted in the previous chapter, we are sometimes content to be guided by a desire, as when we give in to the pleasant feeling of sleepiness in the afternoon sun, or to a desire for revenge. The point is merely that practical judgement could not occur without the initial impetus provided by the external conditioning of our "inner sense". Once that conditioning is accounted for conatively, it is possible for the person to form a maxim – an evaluative principle enjoining him to pursue or shun the future occurrence of the sensation. And that subjective principle, as noted, must contain within itself a reference to one or more objective evaluative principles.

If my argument is correct moral agency in Kant is sensibility-driven, which is to say that it is at the most fundamental level enabled by an affection of the "inner sense". Various commentators, whose authority I have here assumed, have argued that political agency is also driven by sensibly-rooted desires. Sensible affection therefore serves as a link that connects morally and politically relevant types of reasoning. The notion of transcendental identity that explains its possibility is, then, at the bottom of Kant's claims about legitimacy.

Bibliography

Albritton, Rogers. "Freedom of Will and Freedom of Action." *Proceedings and Addresses of the American Philosophical Association* 59, no. 2 (1985): 239-251.

Allison, Henry E. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1991.

Ameriks, Karl. "Hegel's Critique of Kant's Theoretical Philosophy." *Philosophy and Phenomenological Research* 46, no. 1 (1985): 1-35.

Anderson, Joel, and Axel Honneth. "Autonomy, Vulnerability, Recognition, and Justice." In *Autonomy and the Challenges of Liberalism: New Essays*, edited by John Christman and Joel Anderson, 127-149. Cambridge: Cambridge University Press, 2005.

Anscombe, G. E. M. *Intention*. Cambridge, MA: Harvard University Press, 1957.

————. "Under a Description." *Nous* 13, no. 2 (1979): 219-233.

Appiah, Kwame Anthony. *The Ethics of Identity*. Princeton: Princeton University Press, 2005.

————. "How to Decide If Races Exist." *Proceedings of the Aristotelian Society* 106, no. 3 (2006): 363-380.

————. "Identity, Authenticity, Survival: Multicultural Societies and Social Reproduction." In *Multiculturalism: Examining the Politics of Recognition*, edited by Amy Gutmann. Princeton: Princeton University Press, 1994.

————. "Reply to Gracia, Moody-Adams and Nussbaum." *Journal of Social Philosophy* 37, no. 2 (2006): 314-322.

Arendt, Hannah. *Lectures on Kant's Political Philosophy*. Edited by Ronald Beiner. Brighton: Harvester, 1982.

Aristotle. *Nicomachean Ethics*. Translated by Roger Crisp. Cambridge: Cambridge University Press, 2000.

————. *Politics*. Translated by Stephen Everson. Cambridge: Cambridge University Press, 1996.

Atkins, Kim. *Narrative Identity and Moral Identity*. New York: Routledge, 2008.

Baier, Annette. *A Progress of Sentiments: Reflections on Hume's Treatise*. London: Harvard University Press, 1991.

Barry, Brian. *Culture and Equality*. Cambridge, MA: Harvard University Press, 2001.

————. "Something in the Disputation Not Unpleasant." In *Impartiality, Neutrality and Justice*, edited by Paul Kelly, 186-257. Edinburgh: Edinburgh University Press, 1998.

Bennett, Jonathan. "Truth and Stability in Descartes' Meditations." *Canadian Journal of Philosophy* Supplementary vol. 16 (1990): 75-108.

Birondo, Noell. "Kantian Reasons for Reasons." *Ratio* 20, no. 3 (2007): 264-277.

Bittner, Rüdiger. *Doing Things for Reasons*. Oxford: Oxford University Press, 2001.

Blackburn, Simon. *Ruling Passions*. Oxford: Clarendon Press, 1998.

Brubaker, Rogers, and Frederick Cooper. "Beyond "Identity"." *Theory and Society* 29, no. 1 (2000): 1-47.

Chazan, Pauline. *The Moral Self*. London: Routledge, 1998.

Cohen, Joshua. "For a Democratic Society." In *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 86-138. Cambridge: Cambridge University Press, 2002.

Copp, David. "Social Unity and the Identity of Persons." *Journal of Political Philosophy* 10, no. 4 (2002): 365–391.

Cottingham, John. "Descartes and the Voluntariness of Belief." *Monist* 85, no. 3 (2002): 343-60.

Dancy, Jonathan. *Practical Reality*. Oxford: Oxford University Press, 2000.

Daniels, Norman. "Equal Liberty and Unequal Worth of Liberty." In *Reading Rawls*, edited by Norman Daniels, 253-281. New York: Basic Books, 1975.

Darwall, Stephen "Two Kinds of Respect." *Ethics* 88, no. 1 (1977): 36-49.

DePaul, Michael R. "Intuitions in Moral Inquiry." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 595-623. New York: Oxford University Press, 2005.

Descartes, René. *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff and Dugald Murdoch. Vol. 2. Cambridge: Cambridge University Press, 1985.

———. *The Philosophical Writings of Descartes*. Translated by John Cottingham, Robert Stoothoff and Dugald Murdoch. Vol. 1. Cambridge: Cambridge University Press, 1985.

Dillon, Robin S. "Introduction." In *Dignity, Character, and Self-Respect*, edited by Robin S. Dillon, 1-49. London: Routledge, 1995.

Dworkin, Ronald. "Liberalism." In *A Matter of Principle*, 181-204. Oxford: Clarendon Press, 1986.

———. *Sovereign Virtue*. Cambridge, MA: Harvard University Press, 2000.

Eyal, Nir. "'Perhaps the Most Important Primary Good': Self-Respect and Rawls's Principles of Justice." *Politics, Philosophy & Economics* 4, no. 2 (2005): 195-219.

Feinberg, Joel. *Rights, Justice, and the Bounds of Liberty*. Princeton: Princeton University Press, 1980.

Ferrara, Alessandro. *The Force of the Example: Explorations in the Paradigm of Judgment*. New York: Columbia University Press, 2008.

————. *Justice and Judgment: The Rise and the Prospect of the Judgement Model in Contemporary Political Philosophy*. London: SAGE, 1999.

Finlay, Christopher J. *Hume's Social Philosophy: Human Nature and Commercial Sociability in a Treatise of Human Nature*. London: Continuum, 2007.

Fischer, John Martin. "Responsibility and Self-Expression." *The Journal of Ethics* 3, no. 4 (1999): 277-97.

Fleischacker, Samuel. *Integrity and Moral Relativism*. Leiden; New York: E.J. Brill, 1992.

Flikschuh, Katrin. *Freedom: Contemporary Liberal Perspectives*. Cambridge: Polity, 2007.

————. *Kant and Modern Political Philosophy*. Cambridge: Cambridge University Press, 2000.

————. "Kantian Desires: Freedom of Choice and Action in the *Rechtslehre*." In *Kant's Metaphysics of Morals: Interpretative Essays*, edited by Mark Timmons, 185-207. Oxford: Oxford University Press, 2002.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66, no. 23 (1969): 829-39.

————. "Concerning the Freedom and Limits of the Will." In *Necessity, Volition, and Love*, 71-81. Cambridge: Cambridge University Press, 1999.

————. *Demons, Dreamers, and Madmen: The Defense of Reason in Descartes's Meditations*. 1st ed. Princeton: Princeton University Press, 2008.

————. "The Faintest Passion." In *Necessity, Volition, and Love*, 95-107. Cambridge: Cambridge University Press, 1999.

————. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1 (1971): 5-20.

————. "Preface." In *Necessity, Volition, and Love*, ix-xi. Cambridge: Cambridge University Press, 1999.

————. "Rationality and the Unthinkable." In *The Importance of What We Care About*, 177-90. Cambridge: Cambridge University Press, 1988.

————. *The Reasons of Love*. Princeton: Princeton University Press, 2004.

————. "Reply to Eleonore Stump." In *The Contours of Agency*, edited by Lee Overton and Sarah Buss, 61-64. Cambridge, MA: MIT Press, 2002.

————. "Reply to John Martin Fischer." In *The Contours of Agency*, edited by Lee Overton and Sarah Buss, 27-32. Cambridge, MA: MIT Press, 2002.

————. *Taking Ourselves Seriously & Getting It Right*. Edited by Debra Satz. Stanford: Stanford University Press, 2006.

————. "Three Concepts of Free Action." In *The Importance of What We Care About*, 47-57. Cambridge: Cambridge University Press, 1988.

————. "What We Are Morally Responsible For." In *The Importance of What We Care About*, 95-103. Cambridge: Cambridge University Press, 1988.

Fraser, Nancy. "Rethinking Recognition." *New Left Review* 3 (2000): 107-120.

Freeman, Samuel. *Rawls*. London: Routledge, 2007.

Galeotti, Anna E. *Toleration as Recognition*. Cambridge: Cambridge University Press, 2002.

Gardner, Sebastian. "The Primacy of Practical Reason." In *A Companion to Kant*, edited by Graham Bird, 259-274. Oxford: Blackwell, 2006.

Guignon, Charles. *On Being Authentic*. New York, London: Routledge, 2004.

Habermas, Jürgen. *Moral Consciousness and Communicative Action*. Translated by Christian Lenhardt and Shierry Weber Nicholsen. Cambridge: Polity Press, 1990.

Hacking, Ian. "Making up People." In *Reconstructing Individualism*, edited by Thomas C. Heller, Morton Sosna and David E. Wellbery, 222-236. Stanford: Stanford University Press, 1986.

————. *Rewriting the Soul*. Princeton: Princeton University Press, 1995.

Hampton, Jean. *Political Philosophy*. Boulder: Westview Press, 1997.

Harris, Henry. "Preface." In *Identity*, edited by Henry Harris, v-vi. Oxford: Clarendon Press, 1995.

Hegel, Georg Wilhelm Friedrich. *Elements of the Philosophy of Right*. Translated by Hugh Barr Nisbet. Edited by Allen W. Wood. Cambridge: Cambridge University Press, 1991.

————. *Hegel's Logic: Being Part One of the Encyclopaedia of the Philosophical Sciences*. Translated by William Wallace. 3rd ed. Oxford: Oxford University Press, 1975.

————. *Phenomenology of Spirit*. Translated by Arnold V. Miller. Oxford: Clarendon Press, 1977.

Herman, Barbara. "Leaving Deontology Behind." In *The Practice of Moral Judgment*, 208-240. Cambridge, MA: Harvard University Press, 1993.

Hobbes, Thomas. *Leviathan*. Edited by Richard Tuck. student revised ed. Cambridge: Cambridge University Press, 1996.

Hogg, Michael A. "Social Identity." In *Handbook of Self and Identity*, edited by Mark R. Leary and June Price Tangney, 462-479. New York: Guilford Press, 2002.

Hogg, Michael A., and Dominic Abrams. *Social Identification: A Social Psychology of Intergroup Relations and Group Processes*. London: Routledge, 1988.

————. *Social Identifications*. London: Routledge, 1998.

Honneth, Axel. *The Struggle for Recognition*. Oxford: Polity Press, 1995.

Honneth, Axel, and Avishai Margalit. "Recognition." *Aristotelian Society, Supplementary Volumes*, no. 75 (2001): 111-139.

Hume, David. *An Enquiry Concerning the Principles of Morals*. Edited by L. A. Selby-Bigge and P. H. Nidditch. 3rd ed. Oxford: Clarendon Press, 1997.

————. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press, 2000.

————. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Clarendon Press, 1960.

Johnston, David. *The Idea of a Liberal Theory: A Critique and Reconstruction*. Princeton: Princeton University Press, 1994.

Kant, Immanuel. *Critique of Practical Reason*. Translated by Mary J. Gregor. Cambridge: Cambridge University Press, 1997.

————. *Critique of Pure Reason*. Translated by Allen W. Wood and Paul Guyer. Cambridge: Cambridge University Press, 1998.

————. *Critique of the Power of Judgment*. Translated by Paul Guyer and Eric Matthews. Cambridge: Cambridge University Press, 2000.

————. *Groundwork of the Metaphysic of Morals*. Translated by H. J. Paton. London: Routledge, 2005.

————. *The Metaphysics of Morals*. Edited by Mary Gregor. Cambridge: Cambridge University Press, 1996.

Korsgaard, Christine M. "Acting for a Reason." In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, 207-229. Oxford: Oxford University Press, 2008.

————. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.

————. "Skepticism About Practical Reason." *Journal of Philosophy* 83, no. 1 (1986): 5-25.

————. *The Sources of Normativity*. Edited by Onora O'Neill. Cambridge: Cambridge University Press, 1996.

Kukathas, Chandran. *The Liberal Archipelago*. Oxford: Oxford University Press, 2003.

Kymlicka, Will. *Liberalism, Community and Culture*. Oxford: Clarendon, 1989.

Laegaard, Sune. "On the Prospects for a Liberal Theory of Recognition." *Res Publica* 11, no. 4 (2005): 325-348.

Locke, John. *An Essay Concerning Human Understanding*. Edited by Peter Harold Nidditch. Oxford: Clarendon Press, 1975.

————. *The Second Treatise of Government*. Edited by J. W. Gough. 3rd ed. Oxford: Blackwell, 1966.

Loeb, Louis E. "The Priority of Reason in Descartes." *Philosophical Review* 99, no. 1 (1990): 3-43.

————. "Sextus, Descartes, Hume, and Peirce: On Securing Settled Doxastic States." *Nous* 32, no. 2 (1998): 205-30.

Macdonald, Cynthia. "Tropes and Other Things." In *Contemporary Readings in the Foundations of Metaphysics*, edited by Stephen Laurence and Cynthia Macdonald, 329-350. Oxford: Blackwell, 1998.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin, 1990.

Margalit, Avishai. *The Decent Society*. Cambridge, MA: Harvard University Press, 1996.

Margalit, Avishai, and Joseph Raz. "National Self-Determination." *The Journal of Philosophy* 87, no. 9 (1990): 439-461.

Massey, Stephen J. "Is Self-Respect a Moral or a Psychological Concept?" *Ethics* 93, no. 2 (1983): 246-261.

McCarty, Richard. "Maxims in Kant's Practical Philosophy." *Journal of the History of Philosophy* 44, no. 1 (2006): 65-83.

Mendus, Susan. *Impartiality in Moral and Political Philosophy*. Oxford: Oxford University Press, 2002.

Meyers, Diana T. "Self-Respect and Autonomy." In *Dignity, Character, and Self-Respect*, edited by Robin S. Dillon, 218-248. London: Routledge, 1995.

Middleton, David. "Three Types of Self-Respect." *Res Publica* 12, no. 1 (2006): 59-76.

Mill, John Stuart. *On Liberty*. Edited by David Bromwich and George Kateb. New Haven: Yale University Press, 2003.

————. *Utilitarianism*. Edited by Roger Crisp. Oxford: Oxford University Press, 1998.

Miller, David. *On Nationality*. Oxford: Clarendon Press, 1995.

Moody-Adams, Michele M. "Race, Class, and the Social Construction of Self-Respect." In *Dignity, Character, and Self-Respect*, edited by Robin S. Dillon, 271-289. London: Routledge, 1995.

Morris, Christopher W. "Justice, Reasons, and Moral Standing." In *Rational Commitment and Social Justice: Essays for Gregory Kavka*, edited by

214

Jules L. Coleman and Christopher W. Morris, 186-207. Cambridge: Cambridge University Press, 1998.

Nagel, Thomas. *The Last Word*. Oxford: Oxford University Press, 2003.

———. "Moral Conflict and Political Legitimacy." *Philosophy & public affairs* 16, no. 3 (1987): 215-240.

———. *The Possibility of Altruism*. Oxford: Clarendon Press, 1970.

———. "The Problem of Global Justice." *Philosophy &Public Affairs* 33, no. 2 (2005): 113-47.

———. *The View from Nowhere*. New York: Oxford University Press, 1986.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

O'Connor, Daniel. "Kant's Conception of Happiness." *Journal of Value Inquiry* 16, no. 3 (1982): 189-205.

O'Neill, Onora. *Bounds of Justice*. Cambridge: Cambridge University Press, 2000.

———. "Experts, Practitioners, and Practical Judgement." *Journal of Moral Philosophy* 4, no. 1 (2007): 154-166.

———. "Rationality as Practical Reason." In *The Oxford Handbook of Rationality*, edited by Alfred R. Mele and Piers Rawling, 93-109. Oxford: Oxford University Press, 2003.

Parekh, Bhikhu. "Equality in a Multiracial Society." In *Equality*, edited by Jane Franklin, 123-155. London: Institute for Public Policy Research, 1997.

———. *A New Politics of Identity*. Basingstoke: Palgrave Macmillan, 2008.

———. *Rethinking Multiculturalism*. Basingstoke: Macmillan, 2000.

Paton, H. J. *The Categorical Imperative*. Philadelphia: University of Pennsylvania Press, 1971.

Phillips, Anne. *Multiculturalism without Culture*. Princeton: Princeton University Press, 2007.

————. "Recognition and the Struggle for Political Voice." In *Recognition Struggles and Social Movements*, edited by Barbara Hobson, 263-273. Cambridge: Cambridge University Press, 2003.

Plato. *The Republic*. Translated by Desmond Lee. 2nd revised ed. Harmondsworth: Penguin, 1974.

Pogge, Thomas. *John Rawls: His Life and Theory of Justice*. Translated by Michelle Kosch. Oxford: Oxford University Press, 2007.

Rawls, John. "Kantian Constructivism in Moral Theory." *Journal of Philosophy* 77, no. 9 (1980): 515-572.

————. *The Law of Peoples*. Cambridge, MA: Harvard University Press, 1999.

————. *Lectures on the History of Moral Philosophy*. Edited by Barbara Herman. Cambridge, MA: Harvard University Press, 2000.

————. *Political Liberalism*. Expanded ed. New York: Columbia University Press, 2005.

————. "Themes in Kant's Moral Philosophy." In *Kant's Transcendental Deductions*, edited by Eckart Förster, 81-113. Stanford: Stanford University Press, 1989.

————. *A Theory of Justice*. Original ed. Cambridge, MA: Belknap Press, 2005.

Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon, 1986.

————. "When We Are Ourselves: The Active and the Passive." In *Engaging Reason: On the Theory of Value and Action*, 5-21. Oxford: Oxford University Press, 2002.

Reath, Andrews. "Hedonism, Heteronomy, and Kant's Principle of Happiness." In *Agency and Autonomy in Kant's Moral Theory*, 33-66. Oxford: Clarendon Press, 2006.

Roberts, Peri. *Political Constructivism*. London: Routledge, 2007.

Rorty, Amelie Oksenberg. "'Pride Produces the Idea of Self': Hume on Moral Agency." *Australasian Journal of Philosophy* 68, no. 3 (1990): 255 - 269.

Rousseau, Jean-Jacques. *The Social Contract, and Discourses*. Translated by G. D. H. Cole. London: Dent, 1993.

Sachs, David. "How to Distinguish Self-Respect from Self-Esteem." *Philosophy and Public Affairs* 10, no. 4 (1981): 346-360.

Sandel, Michael J. *Liberalism and the Limits of Justice*. 2nd ed. Cambridge: Cambridge University Press, 1996.

Scanlon, Thomas M. "Reasons and Passions." In *The Contours of Agency*, edited by Lee Overton and Sarah Buss, 165-183. Cambridge, MA: MIT Press, 2002.

———. *What We Owe to Each Other*. Cambridge, MA: Belknap Press, 1998.

Smart, J. J. C., and Bernard Williams. *Utilitarianism: For and Against*. London: Cambridge University Press, 1973.

Smyth, Richard. "A Metaphysical Reading of the First Meditation." *Philosophical Quarterly* 36, no. 145 (1986): 483-503.

Stark, Cynthia A. "Hypothetical Consent and Justification." *Journal of Philosophy* 97, no. 6 (2000): 313-334.

Sullivan, Roger J. *Immanuel Kant's Moral Theory*. Cambridge: Cambridge University Press, 1989.

Tamir, Yael. *Liberal Nationalism*. Princeton, NJ: Princeton University Press, 1993.

Taylor, Charles. "Cross-Purposes: The Liberal-Communitarian Debate." In *Philosophical Arguments*, 181-203. Cambridge, MA: Harvard University Press, 1995.

———. "Irreducibly Social Goods." In *Philosophical Arguments*, 127-145. Cambridge, MA: Harvard University Press, 1995.

———. "Overcoming Epistemology." In *Philosophical Arguments*, 1-19. Cambridge, MA: Harvard University Press, 1995.

———. "The Politics of Recognition." In *Philosophical Arguments*, 225-256. Cambridge, MA: Harvard University Press, 1995.

———. "The Politics of Recognition." In *Philosophical Arguments*. Cambridge, MA: Harvard University Press, 1995.

———. *Sources of the Self.* Cambridge: Cambridge University Press, 1989.

Timmerman, Jens. "Kant's Puzzling Ethic of Maxims." *Harvard Philosophy Review* 8 (2000): 39-52.

Van Dyke, Vernon. "Collective Entities and Moral Rights: Problems in Liberal-Democratic Thought." *Journal of Politics* 44, no. 12 (1982): 21-40.

———. "The Cultural Rights of Peoples." *Universal Human Rights* 2, no. 2 (1980): 1-21.

———. "The Individual, the State, and Ethnic Communities in Political Theory." *World Politics* 29, no. 3 (1977): 343 - 369.

van Inwagen, Peter. "Ability and Responsibility." *Philosophical Review* 87, no. 2 (1978): 201-24.

Velleman, J. David. "The Guise of the Good." *Noûs* 26, no. 1 (1992): 3-26.

Waldron, Jeremy. "Theoretical Foundations of Liberalism." *The Philosophical Quarterly* 37, no. 147 (1987): 127-150.

Wallace, R. Jay. "Normativity, Commitment, and Instrumental Reason." In *Normativity and the Will: Selected Papers on Moral Psychology and Practical Reason*, 82-120. Oxford: Clarendon Press, 2006.

Watson, Gary. "Free Action and Free Will." *Mind* 96, no. 382 (1987): 145-72.

———. "Volitional Necessities." In *The Contours of Agency*, edited by Lee Overton and Sarah Buss, 129-159. Cambridge, MA: MIT Press, 2002.

———. "The Work of the Will." In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud and Christine Tappolet, 172-200. Oxford: Clarendon Press, 2003.

Weinstock, Daniel. "Is 'Identity' a Danger to Democracy?" In *Identity, Self-Determination and Secession*, edited by Igor Primoratz and Aleksandar Pavković, 15-26. Aldershot: Ashgate, 2006.

Weinstock, Daniel M. "How Can Collective Rights and Liberalism Be Reconciled?" In *Blurred Boundaries*, edited by Rainer Bauböck and John F. Rundell, 281-304. Aldershot: Ashgate, 1998.

Williams, Bernard. *Descartes: The Project of Pure Enquiry*. London: Routledge, 2005.

———. "Identity and Identities." In *Identity*, edited by Henry Harris, 1-11. Oxford: Clarendon Press, 1995.

———. "Internal and External Reasons." In *Moral Luck*, 101-113. Cambridge: Cambridge University Press, 1981.

———. "Persons, Character and Morality." In *Moral Luck*, 1-19. Cambridge: Cambridge University Press, 1981.

———. "Preface." In *Moral Luck*, ix-xi. Cambridge: Cambridge University Press, 1981.

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G. E. M. Anscombe. 3rd ed. Oxford: Basil Blackwell, 1958.