J. VAN OOSTERHOUT

The Quest
for Legitimacy:

On Authority and
Responsibility in Governance

The Quest for Legitimacy:
On Authority and Responsibility
in Governance

The Quest for Legitimacy:
On Authority and Responsibility
in Governance

Zoeken naar legitimiteit: over autoriteit en verantwoordelijkheid in bestuur.

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr.ir. J.H. van Bemmel
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
donderdag 2 mei 2002 om 16.00 uur

door

Johannes van Oosterhout

geboren te Kampen

Promotiecommissie**:**

Promotoren:
Prof. mr. H.R. van Gunsteren
Prof. dr. T. van Willigenburg

Overige leden:
Prof. mr. M.A.P. Bovens
Prof. dr. B. Nooteboom
Dr. J.J. Vromen

For Sem, Mink and Irene, who paid for this study in valuable time lost.

Contents

I

Background, Stake and Issues

Background

The nature of authority in pre-modern society

Authority is a more or less resilient pattern of actions and mutual expectations that is characterized by an asymmetrical structure of command and obedience. Until quite recently in the history of mankind authority was a central feature of human association. For by far the greater part of human history it was a generally accepted fact of social reality that social relationships involved the ascendancy of some men over others, with the former being in the position to issue commands and the latter expected to obey. For reasons involving history's undeniable bias towards the public realm and the role of significant individuals therein, the relationship between kings and their subjects has become more or less paradigmatic for authoritarian governance.[1] This is certainly understandable as monarchy – literally meaning rule by one – was by far the dominant mode of political organization in the world prior to 1500 AD and for most of the world well beyond that date.[2] In practice, however, relationships of command and obedience neither restricted themselves to the public sphere nor to the summit of the social hierarchy. Via different forms of aristocracy they stretched well into the capillaries of social life.[3] In both economic and family life for example – and it must be noted that historically the two often coincided – authority was a central feature of everyday life with the head of the household being something of a sovereign of the private domain.[4] Because of its ubiquity for the greater part of human history, authoritarian governance was considered a natural phenomenon.[5] That some were in a position to issue commands while others were

---

[1] Hereafter I will use the terms 'authority', 'authoritative governance' and 'authoritarian governance' interchangeably.

[2] For an elaborate historical study on royal authority, see: Bendix (1978).

[3] Bendix (1978:218 ff.)

[4] Even where authority was largely absent from the public sphere – as was the case during the heydays of classic Greek democracy for example – it was still the dominant mode of social organization in the 'private' sphere. See: Arendt (1953:84 ff.) but also Bendix (1978:247)

[5] Authority may be natural in a much deeper sense as authority as authority relations are apparently prevalent among many non-human species, the most interesting of which are

expected to obey was simply the way things were. Until a mere couple of centuries ago it was not the time to question the way things were, if only because it was very difficult to see how things could be otherwise.

It was hard to see how things could be otherwise because authority was considered natural in another sense too. It was not just that it was a dominant feature of social life but authoritarian governance was also believed to correspond to the natural order of things. One has to keep in mind that until only a few centuries ago transcendental metaphysics ruled the human mind. The image that transcendental metaphysics depicted of reality – and of society in particular – was that of an unchangeable and hierarchical 'whole' with ordinary man at the bottom, God or some idea of the Good at the top and earthly authorities such as kings, princes and priests in an intermediary position. Because transcendental metaphysics did not discriminate what is good from what is true, the picture it portrayed of the natural order of things provided both an explanation of and a justification for the institution of authority. The way in which religion – and Christian theism in particular – was translated into the political architecture of pre-modern society is a telling example of how transcendental metaphysics played this explanatory and justificatory role simultaneously. The theistic worldview not only constituted a cognitive scheme explaining the 'logic' behind the de facto order that characterized pre-modern society, but it also provided a powerful role model expressing the mutual normative expectations between rulers and the ruled. Thus in the same way that order in the universe was thought to derive from divine purpose and intervention, order in society was believed to be the product of both the royal will and capacity to intervene in human affairs. And in roughly the same way the theistic God could perform miracles in blatant contradiction with the laws of nature, a king could intervene in the social status quo and revoke even the most firmly established rights of his subjects. The institution of kingdom exemplified a strong parallel between the divine and the political in general. According to a particularly radical view the worldly political order was a mere reflection of reality's deeper metaphysical nature.[6]

Transcendental metaphysics provided fertile soil for the institution of authority in an indirect way too. The world behind appearances was not thought to be equally accessible to all. It was commonly held, for example, that those holding positions of authority had some kind of privileged access either to God or to the deeper truths of nature. It is telling in this respect that the origin of (hereditary) kingship has often been explained in terms of an extraordinary charisma of certain clans or clan-leaders which was commonly attributed to their being held in favour by the gods.[7] It is a highly significant fact in this regard that no royal or princely authority became valid or

---

obviously primates. See on this: Willhoite (1976). Although I do not deny that even amongst humans authoritarian governance may indeed be in some respects *biologically* natural, I will presume cultural factors to trump biological predispositions for both phenomenological and normative reasons.

[6] Schmitt (1922/1985:36-52)

[7] Weber (1968:1141-1143)

effective until some form of religious consecration had taken place first.[8] Charisma constituted such an important foundation for (royal) authority that next to traditional and legal authority, charismatic authority became one of the three "pure types" of authority that Max Weber distinguished in his famous study of regimes of "legitimate domination".[9] It is not just the historical connection between kingdom and religion, however, that exemplified the link between positions of authority and a privileged epistemological status. The great metaphysician Plato also defended such a link in more philosophical terms. The central thesis of his *Republic* is that in a just society – due to his privileged access to the world behind the appearances – the philosopher should also be king.[10] Ironically then – and although the notion of authority itself was not yet available to him[11] – in claiming that being *an* authority on the deeper truths of reality warrants being *in* (a position of) authority in society, Plato was the first to make a systematic distinction between *theoretical* and *practical* authority: i.e. between authority with respect to matters of belief and authority over what actions to take or refrain from taking.[12] As we shall see later on, this distinction has become rather central to more contemporary understandings of authority.

One inevitably wonders how the abstract and intangible image that transcendental metaphysics painted of reality and human society in particular could grab hold on society as a whole, given the fact that only so few were thought capable of understanding this image in the first place. In particular it is hard to conceive how the epistemologically privileged position that authorities claimed for themselves could be sustained in practice given that the supply of true belief was dependent on the extraordinary position of but a few, while the demand for it was exercised by so many. There had to be some kind of intermediary social mechanism by means of which supply and demand could adequately be matched. As a matter of fact there was such a mechanism: *tradition*. It was again Max Weber who pointed out the significance of tradition as a foundation of authority, and once more he thought it to be such an important foundation that he took it as characteristic for one of his three 'pure types' of authority; not surprisingly that of *traditional* authority.[13] But tradition was arguably more than just another foundation of authority. The most interesting light on the role of tradition in buttressing authoritarian governance has been shed by Hannah Arendt, who revealed how tradition complemented transcendental metaphysics in this regard. In her view, the value of tradition in supporting authority was primarily a Roman 'invention' that was subsequently brought to perfection by the Roman Catholic church.[14] Tradition's main achievement was that it proved capable of connecting the rather abstract and intangible metaphysical worldview – that Roman civilization and

---

[8] Bendix (1978:21-60)
[9] Weber (1968:212 ff.)
[10] Plato (1974)
[11] Arendt (1958:84).
[12] Friedman (1973:122-124).
[13] Weber (1968:226-241)
[14] Arendt (1953:98-106)

subsequently Christianity embraced from the Greeks – with the practical necessities of everyday life. The way tradition achieved this was by 'sanctifying' certain common experiences. In particular, it sanctified those common experiences that were perceived of as foundational for the status quo in some way or another.[15] Thus even very common human actions – such as eating a piece of bread – could become normatively highly significant because they stood for a deeper meaning that was itself unavailable to most.

From a more down to earth view tradition functioned in practice because the "fact that something has been done in a certain way strikes most men as a perfectly adequate reason for doing it in that way again".[16] There is more to this rather unsophisticated observation than appears at first sight. In the context of a society in which knowledge and intellectual capabilities were distributed highly unevenly over the population, tradition provided a very efficient means of social ordering because it relieves individual actors to a significant extent of the demanding need to elaborate on what to believe or do themselves. This point also makes clear why tradition fitted rather well with transcendental metaphysics. In the metaphysical worldview, deliberation by individuals who were not in an epistemologically privileged position was simply not thought of as an adequate source of true belief or valid reasons for action to begin with.

It is perhaps difficult to conceive of tradition – something that we now commonly associate with folk-phenomena – playing such an important role in sustaining authoritarian governance in the context of pre-modern society. Yet it must be noted that even at present the social force of tradition cannot be overestimated. It is telling in this respect that a contemporary notion that bears a strong resemblance with that of tradition – i.e. that of *institution* – has been jumped at in quite a number of different intellectual disciplines to come to grips with large 'chunks' of contemporary social reality that systematically appear to escape the explanatory reach of intellectual traditions that proceed from more 'enlightened' behaviourist and individualist assumptions.[17] These explanatory efforts generally testify to the fact that even in modern society ongoing social practices constrain the beliefs and actions of individual actors by filtering out possibilities for belief and options for action through either cognitive or normative screens.[18] I will return to this later on in this study. At this stage I merely want to point out that there is no reason to belittle the historical significance of tradition in understanding authoritarian governance in its pre-modern societal context.

It has been argued that a sufficient understanding of the nature of authority in pre-modern society is already given once the historical "trinity" of transcendental

[15] Arendt (1953:98 ff.)
[16] Wolff (1970/1990:22)
[17] I elaborate on this in chapter V
[18] Cf. Pettit (1996a) The notions of cognitive and normative screening will be explained in chapter V.

metaphysics, tradition and authority has been revealed.[19] Although I appreciate the centrality of this trinity in understanding the nature of authority in pre-modern society, I do not fully agree with this view. The reason for my disagreement is that this view ignores a feature so central to the *resilient* functioning of authoritarian governance that it is simply indefensible to omit it here. The feature I have in mind involves the power that authorities generally had – and used – to sanction deviance and non-compliance. A serious complication of taking this additional feature into account is that we thereby necessarily touch upon the conceptual relationship between the notion of authority and that of power. The problem is that this conceptual relationship has proven rather controversial among modern theorists of authority. There roughly exist two camps on this issue. On the one hand there is the straightforward and rather unsophisticated view according to which authority is but "a manifestation of power" working through the "dispensation of rewards or punishments".[20] On the other hand there is the radical, yet quite sophisticated position that authority and power are more or less mutually exclusive. In the often-cited words of Hannah Arendt: "where force is used, authority itself has failed".[21]

As will become clear later on, I think this conceptual issue warrants a thorough discussion. Yet I also believe that we must not let this issue stand in the way of a proper understanding of the nature of authority in pre-modern society. Both the view that authority and power are mutually exclusive and the position that authority consists of but the power to sanction and coerce are simply not consistent with the (historical) facts as we know them to be. One does not have to be a distinguished historian to know that throughout recorded history authority and power have gone hand in hand with those claiming the right to be obeyed generally not hesitating to resort to coercion in cases they were not. Moreover, it is highly unlikely from an evolutionary perspective that authoritarian governance would have been so ubiquitously around for so long if it did not somehow involve a capacity to sanction or coerce those into obedience who purposefully decline to obey for all the opportunistic and self-regarding reasons we can imagine being relevant. No conjunction of metaphysics and tradition could have prevented such motivational problems from arising. Thus even if we concede that natural compliance was extraordinarily high in pre-modern society, it remains that the persistent and intentional disobedience of some would have eventually undermined the (natural) compliance of many if it remained without consequence.[22]

---

[19] Cf. Arendt (1953:98)

[20] Michels (1937)

[21] Yet at the same time Arendt held that "where arguments are used, authority is left in abeyance". In her view therefore " If authority is to be defined at all, then, it must be in contradistinction to both coercion by force and persuasion by arguments." Arendt (1953:82). Connolly (1974) subscribes to this view for roughly similar reasons.

[22] Cf. Weber (1968:903 ff.) The argument I present here comes down to what Pettit (1993/1996) has called a "resiliency explanation" of social phenomena.

Reducing the notion of authority to the power to sanction and coerce, on the other hand, is historically equally unsatisfactory for similar reasons. For what this view ignores is the fact that authorities in pre-modern society commonly succeeded in gaining acknowledgement for the *claim* to legitimacy that, as we shall see later, is *inherent* to the practice of giving commands. It is self-evident that commands are given with at least the expectation of obedience. But there is simply no way this expectation could have been resilient enough to turn into a sustainable form of authoritarian governance if it did not combine both positive and normative features, that is, if *de facto* and *de jure* authority were consistently separated.[23] More precisely: if threat and coercion were the only means by which compliance to commands could be assured, authoritarian governance would hardly have been viable due to the excessive monitoring and sanctioning cost that would have had to be incurred to make it work.[24] The de facto ubiquity of authoritarian governance in pre-modern society therefore flies in the face of attempts to reduce authority to the power to sanction and coerce.

Allowing power to have played a non-trivial role in sustaining authoritarian governance throughout pre-modern society need not imply a reduction of authority to such a narrowly conceived understanding of power as the capacity to sanction and coerce. If there is one concept in the history of political theory the meaning of which has continuously been broadened it is that of power.[25] I therefore see no reason why authority should be conceptually separated from that of power broadly conceived, but as I have already indicated, I will come back to that issue later on. For now it seems justified to conclude that power has been part of authoritarian governance throughout pre-modern society albeit perhaps mainly in a "virtual" capacity.[26] What seems to have been the case in pre-modern society is that power constituted something of a reserve circuit in authoritarian governance; a factor that became active and manifest only under certain conditions.[27] Typically such conditions obtained when the combination of transcendental metaphysics and tradition failed to secure obedience on their own.

The image of authority that emerges in pre-modern society is one in which it was part of a conjunction of *interdependent* historical conditions – both of an empirical and theoretical nature – that made authority more or less the natural mode of governance in human association. Transcendental metaphysics made authority natural in a direct sense by placing it within the dominant holistic and hierarchical worldview, and indirectly by attributing to those *in* authority some kind of privileged access to the world behind appearances. Secondly, tradition not only constituted a (direct) foundation for authority but also played a more complementary role in connecting the abstract and intangible world of transcendental metaphysics with the day to day

---

[23] Friedman (1973/1990)

[24] Pettit (1993/1996, 1996)

[25] Cf. Lukes (1974, 1986)

[26] Pettit (1995a)

[27] The question, of course, is whether power isn't a *virtual* factor in this sense by definition?

reality of the masses. Moreover, tradition provided an efficient and sustainable social coordination mechanism in a society where knowledge and intellectual capacities were very unevenly distributed throughout the population. Non-trivial differences in power between those in command and those expected to obey, finally, gave authorities the capacity to sanction and coerce into obedience those who lacked the motivation to comply naturally. This more or less closed the circle as it rendered the institution of authority not only natural but also robust enough survive the inevitable occurrence of motivational divergence within human association.

Modernization and the decline of authority
The fact that I have spoken of an *historical* conjunction of interdependent theoretical and empirical conditions in understanding authoritarian governance in pre-modern society already implies the ultimate dissolution of this conjunction. The process that led to this was comprehensive and complex and is generally referred to as modernization. It started somewhere during the Renaissance and took several centuries to complete. Without claiming to give an historically or theoretically adequate account of this process here, I will sketchily point out a few salient contributing factors while focussing in particular on the way in which they undermined the natural status of authoritarian governance. Because the overall purpose of this study is to offer a *systematic* understanding of authority rather than an empirically or historically adequate account of it, the factors I identify are looked at through a lens that enlarges their theoretical rather than their causal relevance.

A first factor responsible for the dissolution of the conjunction of metaphysics, tradition, power and authority involves the historical decline of the role of metaphysics in human life. From the Renaissance onwards, metaphysics – and religion more in particular – were gradually yet surely displaced by a perspective on reality that focussed on what was openly en directly available to all instead of what lay behind the appearances and was accessible to but a few. A renewed interest for the observable and concrete was inspired mainly by the pragmatic successes such an outlook facilitated in science and technology, and thereby indirectly in warfare, navigation, communication, economic development and so on. Society could ignore these successes only at its peril. Yet together with the background assumptions against which they were realised these successes ultimately undermined the hierarchical and static picture metaphysics painted of reality. Looking back it appears that at some stage in human history man simply had to become aware that it was possible to acquire reliable and useful knowledge through rational means without either transcendental metaphysics or its earthly intermediaries taking part in it. Because man increasingly learned to trust this capability, the hierarchical metaphysics that had ruled the human mind for so long ultimately gave way to the egalitarian epistemology of the Enlightenment we currently take for granted.[28]

---

[28] Only in a practical sense I must add. Frankly I do not believe there to be anything close to a theoretically undisputed modern epistemology.

The notion of method came to take a central place in this epistemology. For method – i.e. the conscious and deliberate imposition of rational constraints on the acquisition of knowledge – liberated the human capacity to acquire knowledge from the apparent and inevitable shortcomings of actual individual men.[29] In other words: method made the human capacity to acquire knowledge transcend the *de facto* human reality without, however, resorting to the metaphysical assumptions that had denied (ordinary) man these capacities before. The methodical approach to reality turned out to be outright hostile towards tradition and authority. It was hostile to tradition in that it both challenged and facilitated men to demonstrate that what actually was, was not so by *necessity*,[30] and was certainly not always how it ought to be. It was hostile to authority because it did not tolerate any intermediary between the knowing subject and the object to be known. Method's most important achievement is that it cleared the way to accepting the individual subject as the seat of reason. Thus method became more or less for the egalitarian epistemology of the Enlightenment what tradition had been for metaphysics; i.e. the concrete social vehicle through which the theoretical perspective on the world was connected with the reality of every day life.

It was not just in an epistemological sense that modernisation freed man from the theoretical and empirical chains that had kept him at bay throughout pre-modern society. The disenchantment of the worldview that characterised the process of modernisation also made the holistic and hierarchical *normative* universe of pre-modern society lose ground in favour of a more egalitarian normative 'space' that was more in line with man's newly acquired epistemological status. Within that space reason, autonomy and individual responsibility became the three central dimensions. The Reformation had already dispensed with the intermediary position of the church with respect to the responsibility of individuals vis á vis their God. It thereby also indirectly gave man the key to his autonomy, for no man can be responsible if he is not autonomous at the same time. The autonomy and responsibility of the Reformation, however, were mere preliminaries to the autonomy and responsibility that came in sight with the incumbent rule of reason in modern society. Man's liberation from the hierarchical and traditional normative chains of pre-modern society did not imply a regression to a life fit for animals. In such a life the human animal is ruled by passions and appetites and is constrained only by his capacities to realise them. Instead, *the* central thesis of modernity has become that man – because of his unique capacity to reason – is capable of living by *reasonable* next to de facto constraints to human action. Moreover, because reason is a universal human capacity these constraints are potentially *normatively* binding. They are normatively binding when it is unreasonable for anyone to reject them. Thus instead of being free from normative constraint as such, the main thrust of modernity has become that man is bound by reason and by reason *alone*.

For man to be bound by reason at all two additional conditions had to be satisfied. The first condition was already referred to above and involves the assumption of

---

[29] Cf. Wolin (1960:382-388)

[30] I speak of necessity *de re* here.

individual responsibility. Responsibility is a necessary condition for there being normative constraints to human action because without there being some primitive obligation to answer, that is, to *justify* one's beliefs and actions in the light of reason, there could be no such thing as normativity. It is even hard to conceive of there being anything close to reason if there were not at least some primitive susceptibility to arguments. But being responsible, in turn, is only reasonable if one is also autonomous, that is, if one is the *ultimate* judge of the reasonableness of any action. Without autonomy, there would be no reason to call an action truly ones own. And if there is no reason to call an action one's own, how can one reasonably be (held) responsible for that action? It can be made clear now why there is no place for authoritarian governance in modern society. For modern society is characterised by the rule of reason and reason simply cannot rule in the presence of authority. The main problem is that authority involves a "surrender of judgement"[31] – i.e. an a priori deference to the judgement of the authority – while reason necessarily requires that each man judge on his own account. It goes without further argument that reason exists by the grace of a capacity for judgement. For man to give up judgement is to dispense with reason altogether. But it is not just reason that one gets rid of with the surrender of judgement. The same holds true for autonomy and responsibility. The former is true because if one surrenders ones judgement with respect to a particular domain of decision there is really no reason to ascribe the decisions within that domain to oneself. But if an action cannot be said to be my action why should I nevertheless be responsible for that action? Authority, in sum, is incompatible with the normative universe of modernity because it is in conflict with each of the central dimensions of that universe. This is true if only because these dimensions are connected so closely that to be in conflict with one is to be at war with all. The only authority that is consistent with the rule of reason is authority over *one self,* that is, the authority of self-legislation, so to speak.[32] But that kind of authority is commonly referred to as autonomy.

A third and final factor responsible for dissolving the historical conjunction of metaphysics, tradition, power and authority that I wish to point out here concerns the manner in which modernisation changed the way in which social order came to be perceived. Both the radically changing worldview and rapidly changing economic and social circumstances that accompanied the modernisation of society came to shed a different light on the social and political order that had characterised pre-modern society. While metaphysics had presented this order as necessarily holistic, hierarchical, and unchangeable, these developments demonstrated the relative autonomy of the concrete and the particular and the non-hierarchical spontaneous order that apparently structured nature's unlimited variety of appearances nevertheless. Moreover, man increasingly became aware that instead of being a passive and insignificant part of that order, he could intentionally change and

---

[31] Friedman (1973/1990)
[32] Cf. Korsgaard (1996)

manipulate conditions to his benefit. And so he did; and with significant consequences too.

A combination of technology-driven specialisation and the development of novel means to communicate and disperse knowledge throughout society contributed to both the "intellectual mobilisation" of the masses and an increasing division of labour in society.[33] In turn these developments strained the institutions that had dominated pre-modern society while at the same time facilitating the expansion of others. As a rule, the institutions that involved unreasonable extraordinary advantageous or even exclusive positions of particular groups in society ultimately proved unable to accommodate these developments. Examples of institutions that perished include the guilds, serfdom and the economic, judicial and administrative privileges held by the aristocracy. The institution that profited most from the decline of others was the market. In the same way that tradition had been a highly efficient mechanism of social co-ordination in a society characterised by large inequalities in knowledge, intellectual capabilities and power, the market proved a particularly robust and efficient social co-ordination mechanism in a society in which these inequalities were gradually becoming smaller. In turn, the market also contributed to levelling capabilities and opportunities in society. It did so because of its immanent reward of productive merit in processes of economic exchange.

But the market was more than just a social co-ordination mechanism that fitted well with changing social and economic circumstances. What is far more interesting in the context of this study is that it also became a powerful symbol of the practical emancipation of man. In roughly the same way that method provided an adequate cure for the inevitable epistemological shortcomings of individual and actual men, the market provided a potent social medicine for the practical defects of human nature that remained even after man had been declared as the seat of reason. What the market proved capable of in particular, was to absorb the selfish and often outright knavish behaviour that emerged despite the call of reason and turn this into what is socially desirable; namely wealth and social order, or, to say the same in more technical terms: welfare and equilibrium.[34] The market, then, constituted extant proof that wealth and social order could exist in the absence of hierarchy and authoritarian governance. It needs to be noted that even the staunchest supporters of the free market as a social system of coordination have always insisted that markets – and invisible hand mechanisms in general – are not capable of providing all man holds to be desirable. Markets in particular are widely admitted to be unable to provide the conditions necessary for them to function properly in the first place. Amongst others such conditions involve the existence of property rights and a judiciary to settle inevitably arising disputes.[35] But for markets to question and undermine the natural

---

[33] Bendix (1978:247 ff.)

[34] The notion that markets and invisible hand mechanisms more generally can turn selfish behaviour of individuals into public benefits is central to an long and powerful intellectual tradition that traces back to Mandeville (1714/1989) and Adam Smith (1776/1976)

[35] Cf. North (1990, 1991)

status of authority and hierarchy in pre-modern society it was simply not necessary to completely replace authoritarian governance. It was sufficient to demonstrate that there were alternatives for authoritarian governance that are both viable and in many respects even superior to it.

It is time to turn to the overall effect that the process of modernisation had on authoritarian governance. Perhaps the best way to concisely describe this effect is to say that the process of modernisation resulted in the 'de-naturalisation' of authority in human association. The decline of transcendental metaphysics and tradition in society and the simultaneous emergence of the rational, more egalitarian worldview of modernity increasingly undermined the natural and unquestioning obedience on which authority thrives. Instead, modernity made authority something to be wary of; something that *ought* to be avoided if possible. The reasons for modernity's rejection of authority have already been made clear. Relationships of command and obedience and the surrender of judgement that such relationships imply clash with some fundamental assumptions at the heart of the modern worldview. These assumptions involve the unhindered rule of reason, the position that the individual is ultimately the best judge of what is reasonable, and the corresponding individual responsibility of man for actions or non-actions that can be ascribed to his judgement.

Next to these theoretical changes, the de facto advance of the market as a mechanism of social co-ordination in society constituted the living proof that wealth and social order were both possible and viable in the absence of authority. As the mode of governance in human association turned out to be variable rather than natural, no authority could make an a priori claim to obedience anymore. Instead, since authority was held to be incompatible with the core assumptions of modernity, the burden of proof in the case involving the duty to obey authority shifted from those on the receiving end of commands to those giving them. To put it more precisely, the conjunction of the rejection of authority as incompatible with modernity's core assumptions of autonomy and rationality on the one hand, and the demonstration that there were viable alternatives to authoritarian governance on the other, led to the birth of *the question of legitimacy*; the position that obedience to authority is a matter of debate rather than self-evidence.[36] It was the birth and progressive persistence of this very issue in modern society that led Hannah Arendt to conclude that authority had become something of the past; something the real meaning of which is lost in modern society.[37] For the blind and unquestioning obedience on which authoritarian governance thrives – and which in her view constituted the very nature of authority – had been replaced by the presumption that authority ought to be questioned before it is obeyed.

---

[36] Connolly (1984:2-4)
[37] Arendt (1953)

Stake: the Question of Legitimacy

A modern paradox

At the beginning of the new millennium it is clear that history has proven wrong Hannah Arendt's claim that authority was something of the past. Even casual observation will quickly learn that authoritarian governance is ubiquitous in modern society. Authoritarian governance is arguably even more omnipresent today than it ever was in pre-modern society. The state – to take a particularly salient example – has proven an important vehicle of the process of modernization. It has played a central role in establishing many of our modern rights and achievements and is widely considered to be of no less importance in safeguarding those rights and consolidating these achievements. The state has even become something of a paradigmatic case for authoritarian governance in modern society. It is often seen as the embodiment of Weber's legal-rational type of authority, that is, the only type of authority *alleged* to be consistent with modernity's core assumptions of rationality and autonomy.[38] Recurrent reports of the state's untimely death have therefore proven greatly exaggerated.[39] For instead of being dead and buried, both the number and scope of (nation) states have consistently grown ever since their emergence at the world stage after the West-Phalian peace of 1648.[40] From what we can see now it does not look like it is going to disappear from this stage anywhere in the near future.

It is not just at the level of the (nation) state, however, that authoritarian governance has proven persistent and flourishing. In the last two centuries or so we have also witnessed the emergence and growth of both intra- and international layers of public authority. Within the confines of the modern state public authorities have emerged or evolved in both geographic and functional dimensions. Examples of very common contemporary public authorities involve the courts, municipalities, provinces, metropolitan and regional authorities, central banks, anti-trust authorities, aviation and harbor authorities, waste and water-management authorities, and so on. Ironically the birth of quite a few novel – mostly functional – public authorities has coincided with the gulf of liberalization and privatisations that has characterized the last two decades of the twentieth century.[41] Thus a decline of governmental *dominion* does not necessarily imply a decline in authoritarian governance as new authorities may be instituted to regulate privatised agencies and industries.[42]

Perhaps more interesting than the persistence and development of intra-national forms of public authority is the development of forms of public authority at the international level. The international arena has characteristically been thought of in terms of anarchy and is widely acknowledged as the only place where Hobbes' "state

---

[38] Weber (1968: 212-301;941ff.)

[39] Cf. Mann (1997:472-496)

[40] A quick glance at the UN-membership directory will learn that the number of states has consistently grown until the end of the 20th century. UN (2001).

[41] For a comprehensive and current review see: Megginson & Netter (2001)

[42] Raz (1994)

of nature" has proven to be more than a theoretically convenient metaphor.[43] Seen in this light it is a highly remarkable fact that in the almost two centuries separating the end of the Napoleontic wars from the fourth quarter of the twentieth century the number of International Governmental Organizations (IGO's) has grown even faster than the number of states.[44] Now it must be admitted that most of these IGO's do not come close to rivalling the *sovereign* authority of the nation state. The alleged impotence of the United Nations is a case in point. The latter, however, constitute only one example of what I am talking about here and arguably not a very representative one. The authority of the European Union – to take another example – does come close to the sovereignty of the nation state and has even surpassed it in economic affairs. Regardless of the particulars, then, it is clear at least that the emergence and development of forms of authoritarian governance in the international arena is a highly significant fact in its own right.

Developments in the private sphere show more or less the same general picture. But since authoritarian governance is by definition instituted in a voluntary manner within this domain, the facts – all other things being equal – are more telling here. Significant developments have taken place, first, in what is now commonly referred to as the non-profit sector. The emergence and development of non-profit organizations has been characterized as a "global associational revolution"[45] involving an "upsurge of organized private, voluntary activity in literally every corner of the world."[46] Thus the institutional universe of modern society is inhabited by a broad variety of non-profit organizations ranging from widely dispersed and highly prevalent organizations such as hospitals, welfare organizations and universities on the one hand, to the relatively rare and local chess club or university debating society on the other. Although these organizations differ widely in size, scope, purpose and capabilities, they have in common that they (a) have some formal organizational authority structure with some people in positions from which they can issue directives and bind organizational members through their decisions, (b) have not only been brought to life by their constituents in a voluntary manner, but (c) are commonly also dependent on sustained voluntary support for their continued existence.[47] Their heterogeneity and voluntary nature does not prevent these institutions from being a highly significant factor in modern society, however. They are politically significant, first, because they position contemporary public authority not in opposition to the inert and easily divided masses but rather to potentially powerful and often professionally

---

[43] Hobbes (1651/1968). The characterization of the international arena in terms of anarchy is widespread. An interesting example is Bull (1977). The presumption of international anarchy is an article of faith of the so called 'realist' school in international relations. Cf. Morgenthau (1948)

[44] Walace & Singer (1970)

[45] Salamon (1994)

[46] Salamon, Anheier & associates (1999)

[47] These are just a few criteria underlying the research project of Salamon, Anheier & associates (1999, appendix A).

managed organizations that are capable of forming alliances that in some cases rival the power of the state.[48] They are socially significant, secondly, because much of modern life takes place within and is shaped by these institutions. Think of the schools and universities, for example. Such institutions clearly do more than merely educate youth in modern society. In an important sense they also raise and enculturate modern youth as traditional family arrangements have decreased in significance with the ascent of modernity. Given the centrality and scope of voluntary non-profit institutions in modern society it does not seem off the mark to say that the organizational nature of contemporary civil society is in many respects for modern society what tradition was for pre-modern society. Both have proven to be powerful mechanisms of social reproduction and social integration in their respective contexts.

Arguably the most interesting facts about authoritarian governance in the private sector involve developments in the market. In the previous paragraph it was argued that the market was not only a particularly attractive alternative for authoritarian governance but that it also constituted a powerful symbol for the practical emancipation of man, expressing that autonomy and individual responsibility were compatible with wealth and social order. In this light it is all the more significant that *within* modern markets authoritarian governance is rampant.[49] In the words of a particularly expressive observer modern markets show there to be "islands of conscious power in this ocean of unconscious cooperation like lumps of butter coagulating in a pail of buttermilk"[50] Throughout the twentieth century the "visible hand" of management has increasingly manifested itself within the domain modernity had reserved for governance by the invisible hand.[51] By the turn of the millennium "the ubiquity of organizations" in economic life "have prompted the Nobel laureate Herbert Simon to question the use of the term 'market economy' to describe the structure of economic interactions, suggesting that 'organizational economy' would be the more appropriate term."[52] Thus even in the marketplace which is both a symbol and a paradigm of spontaneous governance in modern society, authoritarian governance has proven persistent and widespread. The fact of the matter, then, is that authoritarian governance is rampant in virtually every corner of modern society. From his education and enculturation in the schools and universities, through his laborious life within the firm and his subsequent retirement and leisure in the golf-club and the old peoples home, until his death inside a hospital, modern man is confronted and constrained by a whole variety of authorities the existence of which he typically takes for granted. The upshot of these remarkable facts is an interesting paradox. Despite modernity's presumption against authority and its doctrine that authority should be questioned before it is obeyed, the facts as we know them show authoritarian

---

[48] Wolin (1960) more or less argues that the organizational nature of modern social reality warrants an equivalent shift in political thinking.

[49] Cf. Bendix (1956); Guillen (1994); Robertson (1999)

[50] Robertson (1923)

[51] Cf. Chandler (1977)

[52] The quote is from Baum & Rowly (2001). The reference made is to Simon (1991)

governance to be ubiquitous and often taken for granted in modern society. It is clear that this paradox is in need of explanation. It is also clear that it needs to be resolved in some way if the modern ideals of rationality, autonomy and individual responsibility are taken seriously.

Explaining authority

There are three interesting and powerful explanations for the persistence and ubiquity of authority in modern society. Authority, first, is a familiar and effective means to cope with the high degree of specialization that is characteristic of modern society. It is familiar because the recipe of relying on authorities for knowledge that is not available to oneself had already proven its worth in pre-modern society. It is effective because the enormous growth of knowledge that distinguishes modern society from its predecessors could have been achieved only through a high degree of specialization and a concomitant reliance on specialists. Seen in this light, there is nothing irrational or outrageous in modern man's reliance on doctors for knowledge of his physical health or his dependence on mechanics for an understanding the functioning of his car. But even if we grant the reason behind relying on (specialist) authorities in matters of knowledge, such an explanation could at best give us only half of the answer we need. For it explains only why there is *theoretical* authority in modern society, that is, authority over matters of *belief*, while it leaves us in the dark about *practical* authority, that is, authority involving matters of *action* independent of questions of beliefs.[53] Now I do not wish to deny that explanations pertaining exclusively to theoretical authority can carry us a great way in understanding much of the modern institutional landscape as governments and organizations are clearly far better endowed with knowledge and knowledge creation and application capabilities than any individual possibly could be.[54] Such an explanation, however, could not inform us on the question as to why it is that both governments and organizations typically have authority over their constituents *in excess* of what would be needed from a cognitive perspective alone. It could not explain, for example, why an authority should care whether or not its directives are *actually* followed, while it is an obvious fact that both governments and organizations are typically not indifferent to the issue of compliance. On closer inspection the issue of theoretical authority does not even constitute much of a challenge to the ideals of modern society. The reason for this is that the 'invention' of method has opened up the possibility for reality to be addressed *directly* in issues involving the jurisdiction of theoretical authorities in matters of belief. To put it in other words: method makes it possible that what is 'out there' functions as an ultimate court of appeal in theoretical matters. Precisely because there is nothing 'out there' that can fulfil the same function for practical matters I shall exclusively address the issue of *practical* authority in the remainder of this study.

---

[53] Friedman (1973:122-123)

[54] This is one of central propositions of so the called knowledge-based theory of the firm. Cf. Connor (1991); Connor & Prahalad (1996); Grant (1996).

A second explanation of authority stays within the constraints given by this focus. It involves the thesis that authority is a comparatively efficient solution to coordination problems.[55] Although coordination problems take a great variety of forms in everyday social reality, a paradigm case for coordination problems exists when there are two or more courses of action that are equally valuable to the actors involved but where the realization of this value is dependent on those actors acting similarly.[56] Thus driving on either the left or the right side of the road, to take a famous example, may be equally valuable to all provided that everybody does the same thing. Authorities now, can help solve coordination problems by authoritatively making a certain option for action *salient,*[57] or by otherwise directing behaviour in complementary paths. In doing so they are able to circumvent *potentially* expensive trail and error processes involved with *spontaneous* coordination. It is clear that efficient coordination mechanisms are not a luxury item in modern society with its high degree of specialization and its complex division of labour. It is therefore not surprising that one of the most powerful and telling explanations of authoritarian governance pertains to coordination problems in the context of a complex division of labour.

This explanation of why authoritarian governance has evolved even within the decentralized decision making context of the market was given by Coase.[58] In his seminal essay on the nature of the firm, Coase addresses the rather fundamental question why there exist organizations within markets in the first place. His answer to this question is as significant as the question itself. For certain transactions entrepreneurial authority is simply more *efficient* than the market. This answer presumes another telling insight. Contrary to what economists assumed for a long time there "is a cost of using the price mechanism".[59] In the view that Coase expounds such 'transaction costs' obtain in particular with respect to coordination problems in conducting transactions. These problems include – among others – price discovery, negotiating and drawing separate contracts for individual market transactions and problems that arise from the conjunction of a lack of foresight on the one hand, and discrete and clear obligations to deliver on the other. The core of

---

[55] Cf. Raz (1975/1999, 1986)

[56] Although I describe a more or less standard game theoretical understanding of coordination problems here – as has been developed and applied by Schelling (1963), Lewis (1969), and Ullmann-Margalit (1977) for example – I mean to refer to a more widely defined set of everyday problems involving issues of coordination. I would like to include the difficulties and costs involved with gathering information relevant to coordination under the heading of coordination problems, for example, even though these are no coordination problems in the more strict game-theoretical sense. I nevertheless present the outlines of a game theoretical understanding as paradigmatic for coordination problems here, as this understanding most clearly demarcates coordination problems from problems of motivation and ultimate valuation to which I will attend later.

[57] Cf. Green (1988)

[58] Coase (1937/1996)

[59] Ibid. p.43

Coase's argument is that the transaction costs that result form such problems can be circumvented by replacing the individual contracts for which such costs obtain by one or just a few open-ended contracts whereby owners of production factors "agree to obey the directions of the entrepreneur" in exchange for "a certain remuneration". Such substitution of authority for the market will continue "until the costs of organizing an extra transaction within the firm become equal to the cost of carrying out the same transaction by means of an exchange on the open market or the costs of organizing in an other firm."[60] If we may disregard the rich theoretical detail of this view for a moment, the kernel for us here is that for certain transactions authoritarian governance is a more efficient alternative than the market. Thus the modern view that the market constitutes a viable and attractive alternative for authoritarian governance expounded earlier is complemented here by the provoking thesis that it can be the other way around too. The gist of this view, then, is that authority and the spontaneous governance of the market are alternative modes of governance and that the choice between the two is a matter of comparative efficiency.

Although the thesis that authoritarian governance is a comparatively efficient solution to coordination problems is interesting, it provides an incomplete explanation at best. This is, first, because it is far from clear under which conditions authoritarian governance is more efficient than spontaneous governance in 'solving' coordination problems. Although we already know quite a lot about the comparative strengths and weaknesses of spontaneous and intentional governance in economic interaction, we are only beginning to understand the basic mechanisms of spontaneous coordination outside the economic sphere such as the emergence of conventions[61] or the spontaneous co-evolution of norms and institutions for example.[62] An interesting but inadequately answered question, for example, pertains to the choice between alternative modes of governance in highly dynamic environments. Although spontaneous coordination is often heralded for its greater adaptive capacities, authoritarian governance may very well be in the advantage where immediate adaptation is required.[63] Even where theory is relatively well developed and empirically robust – as is the case in the economic transaction-cost theorizing for example[64] – the explanatory picture is much more complicated than the one painted above. Coordination problems tend to become much more complicated, second, when they concur with interest divergence and problems of motivation. Efficiency, finally, is not the only criterion relevant to deciding between alternative modes of governance.

A third explanation of authoritarian governance therefore additionally focuses on interest divergence and concomitant problems of motivation while at the same time stretching the explanatory scope beyond the confines of comparative efficiency. In

---

[60] Ibid. p.46-47
[61] A classic work is Lewis (1969)
[62] Sugden (1998)
[63] Green (1988:111 ff.)
[64] See below

this view the purpose of authoritarian governance is mainly to combat problems of collective action.[65] Such problems arise when actors agree to collectively produce some good that is beneficial to all, but where it pays for each individual actor to renege on the agreement later.[66] When actors anticipate such *ex post* opportunism they will simply not enter into agreements anymore with the result that no collective good will be produced. The problem is therefore not so much a matter of comparative efficiency but one of comparative *effectiveness*. For spontaneity may lead to *nothing* being agreed upon and nothing being produced while authoritarian governance may secure both agreement and its fulfilment. Effectiveness is the extreme case of efficiency in that questions of efficiency become issues of effectiveness when they are dichotomised.

A theoretically well-developed and empirically robust example of a deeper insight into the matter pertains to the field of transaction-cost theorizing again.[67] Oliver Williamson in particular has complemented Coase's emphasis on coordination problems and focus on comparative efficiency with explanations figuring problems of *motivation* and attention for the issue of comparative effectiveness.[68] If we let go of the ambition to represent Williamson's arguments in their full theoretical detail the core of his transaction-cost thesis can be stated with great parsimony. Assuming that economic actors are cognitively constrained by "bounded rationality"[69] – i.e. limited fore- and oversight, limited information-processing capacities, and so forth – and motivated by opportunism – i.e. "self interest-seeking with guile"[70] – they will not hesitate to defect on agreements made with parties that are already committed to these agreements beyond the point of no return. In the context of *free market exchange* economic actors will prey on the weaknesses of their contract partners more in particular when: (a) these partners have made investments that are of value only for

---

[65] Standard references here are Olson (1965) and Hardin (1982)

[66] Again, I mean to refer to everyday problems of collective action rather than to a strict game theoretical understanding of collective action.

[67] For a review of empirical research on transaction cost theory see:

[68] Williamson (1975, 1985, 1991a, 1991b, 1991c, 2000). Williamson (1991a) himself explicitly claims to conduct his analysis within an overall comparative efficiency framework. What follows is therefore an interpretation of his work rather than an adequate representation. If we grant that efficiency becomes effectiveness when it is dichotomised, the result is a somewhat generous reading of Williamson that remains consistent with the basic structure of his theory. At certain points Williamson himself seems to endorse an interpretation beyond the comparative efficiency framework. This becomes manifest when he states that one of the central aims of his theoretical framework is "that transactions which differ in their attributes, are aligned with governance structures which differ in their costs and competencies, in a discriminating (*mainly* transaction-cost-economizing) way." (Williamson, 1991b:277, emphasis mine). A bit further down in the same paper (p.278) he says that "adaptability is the central problem of economic organization". Such remarks seem to support a generous interpretation of his work.

[69] Simon (1961)

[70] Williamson (1985:47)

the transaction agreed upon – i.e. have a high degree of asset specificity – (b) transactions with these partners are infrequent such that the continuation of the relationship is not highly valued and reputation mechanisms do not (yet) function adequately,[71] and (c) the transaction is conducted under uncertainty which prevents relevant contingencies to be unambiguously "presentiated" in the contract governing the transaction.[72] The stronger these conditions apply, the more inferior market governance is to authoritarian governance because the latter provides more effective safeguards assuring the integrity and completion of the transactions agreed upon. The safeguards of authoritative *private* ordering are more effective than formal legal adjudication of market contracts because the latter is more expensive, time consuming and uncertain than the former. Williamson's understanding of authoritarian governance therefore involves more than the right to direct actions of relevant parties in complementary paths. As it excludes recourse to formal legal adjudication to resolve contractual disputes it additionally involves a doctrine of *forbearance*.[73] It is therefore not just that Williamson points at additional explanatory variables in explaining authoritarian governance, but the kind of authoritarian governance these variables purport to explain also involve a much stronger form of authority than would be needed to cope with problems of (pure) coordination alone.

Although Williamson's views have met serious criticism,[74] the basic structure of his theory constitutes an interesting "paradigm" for explaining the prevalence of authoritarian governance in modern society.[75] It shows in great detail how problems of motivation add to problems of coordination in explaining the persistence and development of authoritarian governance in modern economic interaction. Although it is clear that this theory cannot be extrapolated outside the sphere of economic interaction without complications, the gist of this view that certain collective values and projects will not survive or will not even be undertaken to be realized in the absence of authoritative safeguards, constitutes a powerful "exemplar" that inspires more specific explanations of authoritarian governance in a variety of different contexts. Yet at the same time it is also clear that there are limits to this kind of explanation of authoritarian governance. The unrealistic assumption of *general* opportunism, for example, constrains this theory to look almost exclusively at problems of motivation. It thereby not only omits other potentially relevant explanatory factors – such as trust, for example – but it also unnecessarily reduces

---

[71] Cf. Williamson (1991c/1996:155ff)

[72] The term "presentiation" traces to McNeil (1978:862-863). It refers to the effort to incorporate all possible future contingencies into a contract as if they were present at the moment of drawing the contract. Williamson (1985:69; 1991b) associates efforts of presentiation with classical contract law and free market exchange.

[73] Williamson (1991b, 1991c/1996:150).

[74] See for example Noorderhaven (1995) and Goshal & Moran (1996)

[75] As is well known, the notion of paradigm derives from Kuhn (1962/1970). I do not refer here to the popularized understanding of this notion as an all embracing theoretical perspective but rather to the original meaning Kuhn had reserved for the notion of paradigm as a successful and hence authoritative "examplar" of how a theory should be applied.

authority to a form of rational control designed specifically to cope with such problems. We already know that there are limits to what rational control can achieve.[76] We also know from the historical roots of authoritarian governance that there is more to authority than a perspective of rational control can possibly capture. Thus even if this kind of explanation adds something to our understanding of the *nature* of authority that cannot be seen from a perspective that exclusively focuses on problems of (pure) coordination, its one-sided concentration on the rational control of motivational shortcomings still seems to omit something that is essential to our understanding of it. This has to do with the normative context in which authority relationships are embedded. More in particular it involves the presumption of obedience that is implicit in the practice of giving and receiving commands. To make sense of this presumption we must open up the 'black box' that authoritarian governance has been for us thus far.

Understanding Authority

The foregoing explanations of the prevalence of authority in modern society have in common that they all proceed from an outside, *third-person's*-perspective. Seen from such a perspective, the notion of command is merely a potentially effective *cause* of relevant behaviour. It is clear from everyday experience, however, that a command is both more and less than the cause of the behaviour that it requires. It is less than a cause because there is an indeterminacy between command and compliance. People often purposefully decide not to obey. A command is more than a cause because a causal analysis alone cannot make sense of the *meaning* that commands have to those actually engaged in giving or receiving commands. Consequentially, *explanations* of authority that proceed exclusively from this perspective fail to capture some important features that relationships of command and obedience typically have in authoritarian governance. In order to identify and understand these features we must exchange the third-person's-perspective we have taken up thus far for the standpoint of those who are actually involved in the practice of giving or receiving commands. Only by taking this *first-person's*-perspective can we reveal the micro social mechanisms – the "nuts and bolts, cogs and wheels"[77] so to speak – that constitute authoritarian governance at its most basic level.

The phenomenological meaning of command is revealed by answering the question what it is that *we* are doing when we are either giving or following orders. I will focus on understanding the latter here as that is the most revealing at this stage of our inquiry.[78] We can already see how an exclusively *causal* account of what is happening when commands are obeyed falls short of what would be considered an appropriate answer to this question. I take it as a matter of everyday experience that to explain the behaviour that the following of an order requires by *causally* reducing it

---

[76] See on this chapter V.

[77] Elster (1989:3)

[78] Lukes (1987), however, has warned that an understanding of authority may differ between those giving commands and those receiving them.

to the utterance of the sounds that communicate the relevant order is actually quite ludicrous. Not only would we think that someone giving such an account had completely failed to grasp what is going on, but this explanation would also be empirically incorrect as there clearly is no straightforward causal relationship between the sounds made and the behaviour it purports to explain. There is therefore more at stake here than what a causal reduction is capable of capturing. What is additionally involved becomes manifest when we ask ourselves the question what would make the behaviour understandable from the *actor's* point of view. Seen from this perspective to follow an order is to presume that there are *reasons* for doing what the order requires. The presumption that following an order must be explained in terms of *reasons* rather than causes is by no means a trivial one. It already helps us to make sense of half of Hannah Arendt's claim that authority should be defined "in contradistinction to both coercion by force, and persuasion by arguments."[79] For although an *exclusively* causal account may be perfectly appropriate in explaining *coerced* behaviour, it is clear that such an account does not suffice in *understanding* what it is that we are doing when we follow an order. The question remains, then, in what sense it is that the reasons that count in favour of following an order are not subject "to persuasion by arguments".[80]

Joseph Raz has given an interesting answer to this question.[81] Presuming that the nature of authority must be understood in terms of practical reasons, his answer involves a *characterization* of commands as a species of practical reasons.[82] The heart of this characterization consists of the thesis that there exist different levels of reasons for action and that these different levels warrant differential treatment in practical reasoning. In Raz' view, the basic 'stuff' of practical reasoning consists of so called "first-order reasons for action". Although Raz is rather precise in defining what first-order reasons for action are, I believe we may conceive of them in terms of what we would ordinarily refer to as beliefs and desires.[83] They are exemplified by the common or garden reasons we give "in explaining, in evaluating, and in guiding people's behaviour."[84] Examples of first order reasons for action are at stake when a person must decide whether or not to accept an overseas job. In all likelihood the reasoning that would guide deciding the matter would include issues of remuneration, secondary benefits, perspective of promotion, as well as questions pertaining to

---

[79] Arendt (1953:82)

[80] Ibid.

[81] Raz (1975/1999, 1986, 1989, 1994)

[82] The idea of characterization involves a definition *per genus ad differantiam*. I shall have more to say about the conceptual strategy of characterization in chapter III.

[83] I am aware that I depart from Raz' position on the nature of reasons for action here, in particular with respect to the (objective) status of beliefs as reasons for action (Raz, 1973/1999:17-20). Since my purpose here is merely to depict in fairly general terms an understanding of the nature of authority in terms of his analysis of practical reasoning, I believe this departure to be quite harmless. Moreover, Raz' thesis that only objective belief qualifies as a reason for action is not uncontested. See in particular: Den Hartogh (forthcoming, chapter VII)

[84] Raz (1975/1999:16)

moving to another country, sending the children to an other school, and so forth. What is characteristic for this kind of reasons for action, now, is that we conceive them as being mutually competitive. They are competitive in the sense that different first-order reasons for action compete with each other for "weight" attributed to them "on the balance" of all relevant reasons.[85] It is easy to conceive of the issue above being resolved in this way, for it is roughly how we commonly proceed in practical reasoning. We standardly assume different reasons for action to have different 'weights' and credit the weights of all relevant reasons on the balance of reasons. The decision is then taken by choosing for the option that has the most weight behind it.

Saying that reasoning on the balance of reasons is familiar to us, however, must not be confused with saying that it is the only kind of practical reasoning we undertake in everyday life. It may very well be the case, for example, that I have promised my wife not to accept a position abroad for a certain period of time. Such a promise is also a reason for action as it bears on the decision at hand. Promises, after all, are made to be kept. The problem, however, is that it is not a reason for action *on par* with the other reasons I have given examples of above. I may very well have made this promise precisely because there are very good reasons for me to accept a position abroad. The purpose of my promise, then, would have been to *exclude* these reasons from being acted upon. A promise is therefore not a first-order reason for action but rather a "second-order reason for action", that is, "a reason to act for a reason or to refrain from acting for some reason".[86] To be more exact: my promise constitutes a second-order "*exclusionary* reason for action" because it aims to exclude whatever relevant (first-order) reasons for action there are from counting on the balance of reasons.[87] There would be no point in making promises if we could always rely on whatever relevant first order reasons for action there are when making a decision.

We can now see the point of the second half of Arendt's claim that authority is to be distinguished form persuasion by arguments. For even though a command must be made sense of in terms of reasons for action, the kind of reasons that are relevant here are second order exclusionary reasons for action whose function it is to override whatever first order reasons there are. Someone who took a command as advice and weighed it against his own views on what the command requires would clearly misunderstand what was expected of him. What this person would fail to capture, however, would not be the propositional *content* of the command. For even when it is perfectly clear what acts or omissions a command requires, the misunderstanding about what is expected *with* giving a command would persist. What is misunderstood here instead is the pragmatic *context* and the performative meaning of command which hold that it requires obedience regardless of one's own views on the matter at

---

[85] For the sake of brevity, the first-order reasons I refer to here are all "operative" and "complete reasons", thereby ignoring all kinds of "auxiliary reasons" for action. See: Raz (1975/1999:22-35)
[86] Raz (1975/1999:39)
[87] Raz (1975/1999:35-48), emphasis mine.

hand.[88] Because it is context and not the content of a command that explains its exclusionary and obligatory features, a command has been named a "content-independent reason for action".[89] The content-independent nature of command was already understood by Thomas Hobbes when he characterized command as "where a man saith, Doe this or Doe not this, without expecting other reason than the Will of him that sayes it".[90] It is also a feature that practical and theoretical authority have in common. For to believe *on* authority that $x = y$ is to have a content independent reason for believing that $x = y$ is true, that is, a reason that has nothing to do with the truth of $x = y$ itself. What is involved here instead is a reason for believing that the person relied up on is indeed *an* authority on questions pertaining to x and y. Likewise, to obey a command is not to act on relevant first order reasons for doing what the command requires. It is rather to have a reason for accepting the orders of a certain person as reasons that override ones own first order reasons for action. The *normative* pragmatics that characterize authority when it is understood from the perspective of one who is subject to it, then, unveils another issue that is indissolubly tied up with understanding what authority is about. This issue pertains to the apparent acknowledgement that "the person to whom one defers is entitled to this sort of submission".[91]

The recognition of authority, to be sure, involves a normative issue. It is the mirror image, so to speak, of the *obligation* to obedience that characterizes authority when it is understood from a first person's perspective. That it is a normative issue indeed becomes evident when one tries to *explain* the authority of X over Y by the *fact* that Y generally obeys the commands given by X. Such an explanation cannot be distinguished from one according to which Y's obedience is the result of coercion. Since we already know that authority must be made sense of in terms of the reasons for, rather than the causes of action, this explanation does not suffice here. What this explanation leaves out is the *normative* question of why it is that X may issue binding directives to Y. This is an inherently normative question, first, in a pragmatic sense in that it involves the practical reasonableness of taking certain speech acts as exclusionary reasons for action. It is a normative question, second, because authority seems incompatible with autonomy and individual responsibility which are important modern values by their own right.[92] Understanding the nature of authority therefore includes issues of *justification* next to questions of explanation.

Because issues of justification are indissolubly tied up with the question what authority must be taken to mean, the notion of authority is not an ordinary concept. Instead it appears to have all the properties of what political theorists have named an "essentially contested concept".[93] An essentially contested concept is a concept of

---

[88] Austin (1962/1975); Searle (1969)
[89] Hart (1982)
[90] Hobbes (1651/1968 part II, chapter 25).
[91] Friedman (1973/1999:68)
[92] Wolf (191970/1990)
[93] Gallie (1956); Conolly (1974)

which the meaning can never be adequately determined by relating it to the facts as they are. The reason for this is that what is at stake with such a concept is not just the world as it is but also the world as it *ought* to be. A paradigmatic example of an essentially contested concept involves the notion of freedom. In defining freedom there clearly is more at stake than merely specifying a descriptively adequate reference for the word freedom as different definitions of freedom have different normative consequences for us. An important issue, for example, is whether freedom must be understood negatively as the absence of external interventions in our lives, or whether there is a more positive ring to freedom such that it also involves certain capacities for action without which one can never truly be free.[94] A related issue involves the question whether "negative freedom" requires the *actual* absence of intervention or whether it extends to even the *possibility* of intervention. Moreover, it is not clear whether the possibility of intervention as such conflicts with freedom or that it is merely the possibility of *arbitrary* intervention.[95] It is easy to see how these different understandings of freedom require different institutional structures in a society that takes freedom seriously. My point, now, is that this is also true for the concept of authority. This may already be intuitively clear as there is a straightforward negative relationship between the notion of freedom and that of authority. What concerns us here, however, is that since our *understanding* of authority implies the presumption of legitimacy, the question what authority is and what it is not is not just a matter of relating it to the facts as they are. Questions of justification also bear on the matter. It are these questions that are at stake in this study. More in particular this study involves an exploration of how a number of justificatory issues impact on our understanding of the nature of authority.

The Issues

The justification of authority

There are roughly two *general* justifications of authority. The first one is rather straightforward and holds that authority can be justified by the consent of the governed. Although this view connects well with our modern intuitions, it must be rejected nevertheless. I shall have little to say about it besides explaining why it should be rejected. The second justification is more complicated. The gist of this view is that authority is justified if, and only if, it makes those subject to it comply better to reasons that apply to them already. This view involves a rather restrictive theory of authority and has appropriately been named the "service conception of authority".[96] It elaborates on Raz's characterization of authority in terms of practical reasoning discussed in the previous paragraph. In contrast to the first view I consider this second view at some length. In fact, this study will develop an argument against this

---

[94] Cf. Berlin (1958)
[95] Pettit (1996b, 1997)
[96] Raz (1986)

view. Before I can elaborate on of this second view, however, I must first show why it is that the first view must be rejected without much further consideration.

Justification by consent

The view that authoritarian governance is justified by the consent of the governed appeals at face value. This is because there seems to be a straightforward relationship between consent and authorization, on the one hand, and authorization and authority, on the other. Consent is implied when I *authorize* a person to do this or that on my behalf. When, for example, I authorize my attorney to sign a contract for me in my absence I am bound to that contract *as if* I had signed it myself. Is not consenting to authoritarian governance like this as Hobbes already suggested?[97] The answer to this question is negative. As I see it there are roughly three reasons why the analogy between authorization and consenting to authority does not hold. The first one is empirical. While an authorization must always be actual, consent to authoritarian governance is often not actually given. The paradigm case in support of this argument involves the authority of the state. As a matter of fact, few, if any, have ever actually consented to being subject to the state's authority. Typically only immigrants pledge obedience to it in the process of naturalization. It is clear, however, that this is an exception. But even if it would be no exception it is still strange to hold that only immigrants who formally apply for citizenship of a country are bound by the authority of its government. The default presumption is that all inhabitants of a country are subject to the state's authority and that even illegal aliens have an obligation to obey its laws.

To make sense of this presumption it has been suggested that consent need not be actual and that *implicit* consent may suffice. Consent may be implied in certain kinds of action – such as voting – or in benefiting from certain collective arrangements – such as safety or social security. The problem with this view, however, is that it may not be clear what kind of action – or lack of it – signals consent. It seems fairly obvious, for example that when I travel to another country I undertake a prima facie obligation to obey its laws. But does undertaking such an obligation mean that I consent to the authority of its government? What if this government has seized power through violent means and oppresses its people in a horrible way? Are we endorsing such a government with consent merely by travelling to this country? There are many good reasons to hold that this is not so. That there is room for debate on this issue highlights a second reason why consent cannot authorize authority. This is that it is often not clear what constitutes consent and what not.[98] Whether one has consented or not may even be problematic when consent has been explicitly and intentionally given. When two persons exchange wedding vows, for example, this does not yet constitute a marriage. The exchange of vows has the intended normative effect only in the presence of a person that is entitled to perform a marriage and when certain other background conditions are fulfilled that apply independently from

---

[97] Hobbes (1651/1968)
[98] Cf. Green (1988:158 ff)

the act of consenting itself. In more technical terms: consent is an "institutional fact", that is, an act is an act of consent only by virtue of a pre-existing institutional background that makes it an act of consent.[99] Because consent is constituted by the conjunction of there being such an institutional background and the performance of the right kind of signalling act, it is often not very clear whether or not consent has been given. Like the act of consenting itself the relevant institutions need not be formal and explicit. They may even be contested in some way. When the latter is the case the validity of consent is questioned.

Questions of validity constitute a third reason why consent cannot authorize authoritarian governance. The validity of consenting to authority is threatened in two separate ways. First, the conditions under which consent is given may invalidate it.[100] Such is the case, for example, when consent is given under duress or threat of coercion. Now it needs to be noted that consenting to authority is something that people typically do only with great reluctance. In fact, the extent to which we value autonomy and self-control in modern society suggests that consenting to authority is actually something we do only when it is practically necessary.[101] This invites the question whether consenting to authority can ever be free of duress or coercion. For the fact that consent to authority is typically given with great reluctance suggests that the context in which consent is given may never be completely free of invalidating conditions. That this is not just a theoretical issue becomes manifest when two important cases of consenting to authority are inspected in more detail. Since there is no real alternative to being a citizen of some state, first, consenting to the authority of the state *in general* does not seem to qualify as a matter of free and unconstrained choice. Whatever flicker of freedom remains in the general case is quickly extinguished once one realizes that in today's world it is typically not even up to the individual to decide to which state's authority in particular consent can be given. Although less dramatic, second, a roughly similar logic applies to corporate authority. In securing our livelihood it is difficult for most of us to avoid the highly prevalent authoritative mode of governance in modern economic interaction I have spoken of earlier. Only those with highly marketable skills and capacities can afford to opt out of the corporate labour market.

The validity of consenting to authority is threatened, second, by a much more fundamental objection. This is the so called anarchist objection that authority is simply incompatible with the modern demands of autonomy and individual responsibility.[102] The anarchist case against authority is clear. Authority involves the surrender of judgement by taking a command as a second order exclusionary reason for action. Autonomy and individual responsibility, however, require that each man judges on his own account on the balance of whatever reasons there are.[103] There is

---

[99] Searle (1995)

[100] Cf. Green (1988;173 ff.); Gilbert (1999/2000:99-101)

[101] The notion of practical necessity is explained in chapter III

[102] Cf. Wolff (1970/1990)

[103] Raz (1990:4)

no way the two claims can be reconciled. It can be objected, however, that the anarchist objection is too demanding. To insist that each man always judges for himself on the balance of reasons would not just rule out surrendering ones judgement to authority but also promising, acting on the authority of doctors, lawyers, mechanics, and so forth.[104] Even if we concede to this objection and accept that it may sometimes be reasonable to surrender one's judgement to such *limited* forms of authority, it remains that no *unlimited* authority can ever be valid. What a valid justification of authority must minimally provide for, then, is a limit to the scope of authority.[105] We can now see why consent alone can never constitute such a justification. For consent – like command – is a form of content-independent commitment, while a content-independent commitment cannot limit to the scope of authority by itself. As was already hinted at above, it can only do so against a background of pre-existing content-full constraints on authority. What a valid justification of authority must consist of, then, is a *general* content-full doctrine that limits the scope of authority.[106] Only against the background of such a doctrine can consent provide a *secondary* justification of authority at best.

The normal justification of authority

Joseph Raz has developed a content-full justificatory doctrine of authority. It proceeds from his characterization of authority as a species of practical reasoning. According to this characterization a command is a content-independent exclusionary reason for action for the actor(s) to whom the command is addressed. This content-independent exclusionary reason for action results in an obligation to obedience if, and only if, the authority who gave the demand is entitled to do so, that is, if it is a *legitimate* authority. The content-full justificatory doctrine proposed by Raz consists of three related theses: (a) "the dependence thesis", (b) "the pre-emption thesis" and (c) "the normal justification thesis".[107] What these three theses involve can be made clear with the aid of an example involving the practice of arbitration.

Arbitration exists when two or more actors refer a dispute to an arbitrator and accept his decision on the matter as binding. Let us assume that the disputants consist of A and B and that C is the arbitrator. They refer their dispute to C because A wants a and B wants b which are presumed incompatible. For both A and B the reasons α and β bear on the issue at hand but α prevails for A while B holds that β has more weight behind it. We already know that the bindingness of C's decision does not follow from the act of consenting to arbitration alone. The reason for this was explained above but becomes manifest when C decides c – which we assume is not some reasonable compromise – for a reason γ that has nothing to do with α or β. We would not consider such a decision valid because it is based on neither α nor β and thus exceeds the *scope* for which authority was given in deferring the case to the

---

[104] Green (1988:29-36)
[105] Raz (1990:12)
[106] Ibid.
[107] Raz (1986:38ff)

arbitrator. The "dependence thesis", now, limits the scope of legitimate authority by requiring that any decision *c* should be based on α or β. More in particular this thesis holds that "all authoritative directives should be based, among other factors, on reasons which apply to the subjects of those directives and which bear on the circumstances covered by the directives". [108]

The pre-emption thesis holds that the authoritative command *c* replaces the reasons α and β on which it is based and thus thereafter pre-empts those reasons from being brought forward again by those seeking arbitration. Thus *C*'s decision *c* – presuming now that it is based, amongst others, on α and β – replaces α and β such that *A* and *B* cannot rely on them anymore after *C* has made his decision *c*. The logic behind this is clear. It does not make much sense to commit yourself to binding arbitration if you can later come back on the arbiter's decision by restating the reasons you already forwarded while committing yourself to arbitration. In seeking arbitration one presumes that the reasons brought before the arbitrator will be considered and weighed in deciding the issue. To resort to those reasons again after the decision has been taken would be like a request to double count these reasons: once in making the decision and once again in overturning it such that one gets ones way whatever was decided initially.

Although theoretical and practical authority are two different things, the dependency thesis and the pre-emption thesis apply to both.[109] The logic behind their application is even more apparent in the case of theoretical authority. The dependency thesis applies because in relying on a theoretical authority we presume that the judgement of authority is based on reasons for belief that apply to us as well and that can – in principle – be inspected by us *if we so desire*. We rely on authority, nevertheless, because we can thereby save ourselves the effort of finding out if and how these reasons apply by our own lights. Typically, theoretical authorities are in a superior position to acquire relevant knowledge because of advantages of specialization. The pre-emption thesis applies to theoretical authority because it does not make much sense to selectively try to improve on the judgement of the authority relied upon if we presume that the authority has superior knowledge on the issue at stake. Such selective intervention can be successful only if one has more or better knowledge than the authority. But if that is the case there is really no reason to rely on authority in matters of belief in the first place.[110]

The "normal justification thesis" can now be formulated against the background of these two theses. It holds that "the *normal* way to establish that a person has authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which

---

[108] Raz (1985/1994:198)
[109] Raz (1986:53)
[110] Ibid. p.67-69

apply to him directly".[111] The normal justification thesis thus asserts that the purpose of authority is to help those subject to it to comply better with reasons that apply to them already. Unlike theories of consent, the normal justification thesis thereby proposes restrictions on the scope of legitimate authority. It limits the scope of legitimate authority, first, to the sphere of reasons that already apply to those subject to it. This does not mean that these are the only reasons an authority may take into consideration. There may very well exist relevant reasons that apply exclusively to authorities, such as administrative considerations, for example.[112] What it does mean is that an answer to the question whether an authority is legitimate has nothing to do with such reasons. What ultimately counts in justifying authoritative commands is the extent to which they help subjects to comply better with their own reasons for action. The normal justification thesis thereby limits legitimate authority to those cases in which relying on authoritative commands is more efficient or more effective than relying on one's own reasons for action directly. There are numerous ways in which this can be true. One particularly interesting case involves the coordination problems I have spoken of earlier. In such cases complying with authoritative directives is often the most efficient way of coming to do what one wants to do *because* everybody else is doing it. Not surprisingly, solving coordination problems is a central function of the kind of authority brought to life by the normal justification thesis.[113]

The normal justification thesis is not the only way in which authority can be justified. There may be many (content-full) reasons why we should take the commands of a person as authoritative. It may be the case, for example, that a valuable cooperative scheme will deteriorate if one does not accept a certain person in a position of authority. The scope of authority is then restricted to realizing that particular cause. What distinguishes the normal justification thesis from such particular justifications is its *general* claim to legitimacy. It does not apply to any particular authority but rather to authority in general. As such it not only constitutes a mould that demarcates the scope and validity of particular justifications, but it also provides a content-full background doctrine against which secondary justifications of authority – such as consent – can do their work. The normal justification thesis constitutes a general justification of authority by virtue of spelling out how authority is *normally* justified under *normal* conditions. Because the justification of authority involves an inherently normative issue, these normal conditions should be understood in normative rather than empirical terms. Strictly speaking, they are therefore ideal rather than normal conditions. Since the ideal function of authority is to serve those subject to it by helping them to comply better with reasons that apply to them already, this view has been named the "service conception of authority".[114]

It is important to note that the service conception proposes a very restrictive understanding of authority. Again, it holds that *only* those authorities are justified that

---

[111] Ibid. p.53
[112] Ibid. p.48 and 72 ff.
[113] Ibid.
[114] Ibid. p.55-56

make their subjects comply better with the reasons that apply to them already. As Raz himself has elegantly acknowledged, projecting this requirement on extant forms of authority in modern society will quickly reveal that not many of them are legitimate authorities and hence may not be taken as authorities at all.[115] The state's exclusive right to apply violence, for example, cannot be based on its authority if we conceive of it in terms of the service conception. For violence is clearly beyond the scope of reason. Even its capacity to issue generally binding laws is seriously hampered by the demand that the legitimate scope of authority extends only to helping citizens comply better with reasons they have already. Thus even a law pertaining to the protection of basic civil liberties such as freedom of speech and the principle of habeas corpus cannot bind those who claim that there are overriding reasons not to support such basic rights. It are not just the ordinary functions and powers of the state, however, that fall outside the scope of the service conception. The typical organization also claims more authority than it can legitimately have. Thus while corporate authority, for example, may be legitimate when management restricts itself to solving coordination problems by guiding peoples behaviour in complementary paths, it certainly ceases to be so when the purpose of organization is rather to combat problems of motivation in economic interaction. Consequentially, the doctrine of forbearance that forms the heart of a widely accepted and empirically robust theory of the firm is largely incompatible with the requirements of legitimate authority as conceived of by the service conception.

The conclusion emerges that although the service conception is theoretically rigorous, it is practically inadequate nevertheless. It is descriptively inadequate because it cannot account for a large part of extant forms of authoritarian governance in modern society. It is normatively inadequate because it cannot justify common and instrumentally valuable forms of authoritarian governance. The major reason for its practical inadequacy appears to be that the service conception is geared primarily to justifying the authority of *one* actor over *one* other.[116] Its ability to cope with authority over groups of actors in which we are most interested here is severely constrained by the requirement that it may not exceed the sphere of reasons that apply to them *all*. Seen in this light it seems doubtful even whether the service conception can account for the practice of arbitration that I took as an example to explain its three constitutive theses. For in arbitration it is often the case that different reasons apply to actors A and B for wanting *a* and *b* respectively. Thus A may want *a*, for the reasons α and δ, while B may want *b,* for reasons β and ε. As there is no way for the arbitrator to decide without deciding against reasons that apply to one of those seeking arbitration, no verdict can be *legitimate* in this case. The same holds true for problems of motivation for roughly similar reasons. Suppose that A agrees with B to do *x* for the shared reason α. When A has done his part, things may have changed for B – typically *because A* has done so – such that B now wants to do *y* for the reason β which did not exist before A did his part. Again, there is no *legitimate* way to make B

---

[115] Ibid. p.70 ff; see also Raz (1994)
[116] Contrary to what Raz himself asserts (1986:71)

do his part without acting against β which I shall presume is a reason for action in exactly the same way that α is.[117] A serious complication is that these problems are highly prevalent in groups. A perfectly ordinary case exists when A and B want 'a and b' for the reasons α and β, while D wants d for the reason δ which conflicts with α and β. In this case C cannot legitimately impose 'a and b' on D even if C believes that 'a and b' is the right choice for the right reasons. Because such problems are ubiquitous, and since, as we have seen, it is a major function of authoritarian governance in modern society precisely to cope with such problems, it seems awkward to reject extant forms of authority because they do not meet the requirements set by the service conception. It seems more appropriate instead to revisit the question of legitimacy that the service conception provided an answer to in order to investigate how the gap between the service conception, on the one hand, and highly prevalent forms of authoritarian governance in modern society, on the other, can be narrowed. This study aims to explore this gap. More in particular it elaborates on three justificatory issues that are central to understanding what authority is about in modern society.

Issues and plan of this study

The first issue involves an ongoing conceptual controversy. It pertains to the conceptual distinction between authority and power I have touched upon earlier. In characterizing authority I have argued that it must be made sense of in terms of the reasons for rather than the causes of action. More in particular I have explained why authoritative directives are exclusionary reasons for action that result in an obligation to obedience to those to whom the directive is addressed. The *normative* nature of this characterization of authority constitutes the most important reason why it is held that authority must be understood in contradistinction to power. For power concerns merely a (potential) cause of action, which in itself can have no normative consequences, or so at least is the presumption widely adhered to.[118] It has even been claimed that power and authority are mutually exclusive. In the words of Hannah Arendt: "where force is used, authority itself has failed."[119] Although it is clear that an understanding of authority in terms of reasons for action and power conceived of as blunt force are uneasy bedfellows, it is doubtful whether authoritarian governance can *actually* do without power. Power is a necessary ingredient, first, even of authority as conceived by the service conception. For only those authorities can be legitimate that have the *capacity* – i.e. the power – to make those subject to it comply *better* to reasons that apply to them already.[120] It is a misunderstanding that solving coordination problems does not require any power. A second reason why it is difficult to conceive of authority without power was explained while tracing the historical roots of

---

[117] We could of course deny that β is a reason for action. Imposing such a restriction, however, would require too much of the concept of a reason for action.

[118] Cf. Arendt (1953); Conolly (1973); Friedman (1973); Raz (1986, 1994)

[119] Arendt (1953:82)

[120] Raz (1986:76)

authoritarian governance in pre-modern society. It was argued that it is unlikely that any structure of practical authority would have survived for very long if it did not have the capacity to cope with inevitably emerging problems of differential motivation. As there is no reason to presume that time has invalidated this insight it remains that resilient forms of practical authority require a solid power base. A third reason relates to the point of the whole study. If, as I have argued, it is practically unavoidable to extend the scope of authoritarian governance beyond the confines of the service conception, we will at some stage inevitably encounter the non-compliance of those whose reasons are acted against. Coercion seems unavoidable in coping with such cases.

As I see it, much of the conflict between power and authority derives from the presumption against power that I have briefly touched upon above. According to this presumption power is blind and can therefore have no normative consequences. This presumption is deeply rooted. It is reflected in the popular saying that might cannot make right. But is this presumption really valid? In the next chapter I shall argue against this presumption. My point will be that because of a general feature of practical norms, power *can* under certain conditions even be constitutive of practical normativity. If I am right about this claim then the conceptual distinction between authority and legitimate power loses its air-tightness. The result is that no conceptual issue stands in the way anymore of conceiving of authority in the classic Weberian sense as legitimate power. Such an understanding would even be welcome as it could narrow the gap between the service conception and extant forms of authoritarian governance in modern society.

That there is no sound conceptual demarcation between authority and legitimate power makes the question what it is that makes power *legitimate* rather urgent. For the next boundary at which the question of legitimacy is at stake is that between power and legitimate power. The standard answer to this question is that legitimate power is power constrained by norms.[121] This answer is widely accepted in modern society and lies at the heart of central modern institutions like the rule of law and the Rechtsstaat.[122] In chapter III of this study I investigate the validity of this answer. I do so by considering an interesting anomaly. This anomaly involves the so-called problem of "dirty hands".[123] The problem of dirty hands is one of the most persistent and problematic paradoxes that political theory has produced. It holds that sometimes persons in authority must act bad in order to do good. After having characterized this problem in more systematic terms than has been done thus far, and after having ascertained that it involves a real problem rather than some conceptual confusion, I

---

[121] I acknowledge that there may be a circularity here. For if power can be constitutive of norms, how can norms constrain power? An answer is that it still can be if power is constrained to realizing a norm. As we shall see later on, this answer is rather problematic however.

[122] It finds its theoretical foundation in Weber's (1968) legal-rational type of authority.

[123] The terminology traces to Sartre (1947). One of the first systematic accounts of this problem is given by Walzer (1973)

conclude that it is inconsistent with the account of legitimate power as power constrained by norms. I even submit the somewhat radical thesis that the problem of dirty hands is in an important sense what authoritarian governance is all about in modern society. It is one of the central purposes of the modern state, for example, to cope with normative conflict such that one man's religion does not become the graveyard of another man's freedom. Thus it is not just that there is no clear conceptual boundary between authority and legitimate power. Neither is there a clear-cut distinction between legitimate power and power as such.

In chapters IV and V of this study, I abandon the ambition to find a conceptual answer to the question of legitimacy that authoritarian governance confronts us with. In these chapters the possibility is explored of a *pragmatic* solution to what is an ongoing quest for, rather than a question of legitimacy. Thus, instead of finding out if and how authority can be legitimate in any definite theoretical sense, I now investigate the possibility of learning to cope with authority by *making* it *responsible*. The promise implicit in this suggestion is that perhaps responsibility can relate to authority within hierarchy as demand relates to supply within a market, that is, that the practical problems pertaining to the legitimacy of authoritarian governance will be cleared once authority and responsibility are in some sort of equilibrium. In chapter IV the possibility of a *conceptual* understanding of responsibility for authority is explored by mapping the requirements that such an understanding would entail. As it turns out, it is highly unlikely these requirements can be met simultaneously. This is not just true for responsibility for authority, however, but rather for responsibility in general. An interesting paradox results from the analysis in this chapter. While it is not very likely that the requirements pertaining to a conceptual understanding of responsibility will be met in everyday reality, the practice of responsibility is actually ubiquitous in modern society.

In chapter V the air of paradox is cleared by abandoning the requirement that our understanding of responsibility be conceptual. This chapter explores in broad strokes an alternative general *institutional* understanding of responsibility. It is argued that an institutional understanding of responsibility is both theoretically more powerful and practically more viable than any conceptual understanding possibly could be. It is theoretically more powerful because it can incorporate both responsibility for authority and individual responsibility within the confines of a single conceptual framework, i.e. that of a general institutional theory. It is practically superior not only because of its scope of application, but also because it sheds interesting light on the question how the *institution* of responsibility for authority can be made more *resilient* in practice. A broad outline is subsequently given of what an institutional understanding of responsibility for authority could possibly look like by contrasting it with the currently dominant agency theoretical approach.

I conclude this study by formulating the central theses that run through the five chapters of which this study consists. These theses not only recapitulate the major findings of each chapter, but taken in conjunction they also constitute the skeleton of

an argument against any unambiguous and unified theory of authority.[124] More precisely, the major line of argument of this study rules out any understanding of the nature of authority that *theoretically* incorporates an answer to the question of legitimacy within that understanding. If there is one overall conclusion to this study, it is that like responsibility, authority is best conceived of in institutional terms. It is claimed that an institutional perspective on authoritarian governance offers the best hope of coping adequately with the quest for legitimacy that authoritarian governance confronts us with.

---

[124] In contrast with MacMahon's (1994, 1995) views, for example. Although I do agree with him that there is no principle conceptual difference between authority in the private sphere and authority in the public sphere, I do not think that there is an underlying *concept* of authority that is primitive to both understandings.

II

Authority and Intervention

Introduction

I have argued that *understanding* authority requires opening up the black box of authoritative practices by taking up the internal perspective of those involved. The understanding of authority that emerges from this first-person's-perspective is one in which authority is characterized as a species of practical reasoning. In this view, commands are seen as content-independent exclusionary reasons for action that aim to exclude whatever relevant content-dependent reasons there are on the balance of reasons of those addressed by it. It was explained that taking a command as a content-independent exclusionary reason for action presumes that it is legitimately given. This presumption thereby makes the question of legitimacy part and parcel of what authority must be taken to mean. According to the "service conception" of authority proposed by Joseph Raz, a command is legitimately given *if, and only if,* it helps those addressed by it better to comply to reasons that apply to them already.[125] It is important to note that the resulting understanding of authority is a *normative* one in three different yet clearly related ways. It is normative, first, because commands are *normally* understood in terms of their normative consequences, which involve a correlate obligation to obedience.[126] It is normative, second, because the service conception involves a doctrine of *legitimate* authority, that is, it is a doctrine that defines authority in terms of the conditions under which it is *justified*. It is normative, third, because authority is made sense of in terms of the reasons rather than causes. It is of the essence of reasons that they have normative force.[127]

I have explained, however, that in spite of its theoretical rigor the service conception is practically inadequate. It is descriptively inadequate, first, because it cannot account for highly prevalent forms of authoritarian governance in modern society. It would not be off the mark to say that the service conception can account only for a small portion of extant authorities. It is normatively inadequate, second, because it cannot normatively ground those existing forms of authoritarian governance that we consider valuable for utilitarian reasons. It is highly significant in

---

[125] Raz (1986)

[126] Ibid.

[127] Cullity & Gaut (1997)

this regard that, because of their comprehensive scope, nor the firm neither the modern state are legitimate authorities in the eyes of the service conception, and hence are not really authorities at all. I have also expressed some conceptual worries regarding the distinction between authority and power that is presumed by the service conception. For in spite of authority's conceptualisation "in contra-distinction to power", it seems doubtful whether authority can actually ever do without it.[128] I have argued, first, that even legitimate authorities must have the power to do what they may legitimately do and that it is a misunderstanding to think that *efficiently* solving coordination problems does not require any power at all.[129] I have pointed out, secondly, that no authority structure can be resilient and therefore lasting if authorities do not have a way to cope with occasional yet inevitable non-compliance. In line with this requirement, thirdly, I have argued that any attempt to expand the scope of authoritarian governance beyond confines of the service conception will necessarily invite a reliance on power because natural compliance may be expected to halt precisely at these confines. Together these issues warrant revisiting the conceptual distinction between authority and (legitimate) power.

There are, to be sure, different ways in which the conceptual distinction between authority and power may be scrutinized. One way would be to reconsider our understanding of the notion of power. It could be said, for example, that the account of power I have (implicitly) given thus far overemphasizes its coercive features while at the same time neglecting its more 'benign' qualities that become manifest when men purposefully act together, and that often come with an air of legitimacy.[130] I do not pursue this strategy for two reasons. First, the kind of power I believe is (at least partly) co-extensive with, and potentially complementary to legitimate authority does not merely involve this 'benign' collective ability but also its more base coercive features.[131] I have argued that coercion is unavoidable in coping with (occasional) non-compliance. Focussing on the less problematic features of power, secondly, does not do away with its more troublesome aspects in general. Ones understanding of what it is like to be an executioner is not greatly enhanced by focusing on what executioners do in the time off from their unholy profession. The perspective on power that I have in mind is therefore a comprehensive one. It involves both the capacity that men exclusively have when they act in collusion, as well as the more asymmetrical capacity that one particular actor may have to make another do something against his will.[132] What I think is a common denominator to the broad extension to which this understanding applies is that power involves a capacity to *intervene* in the *social* status quo. Two features stand out in this description. First, power

---

[128] Arendt (1953:82)

[129] One may think of the kind of collective ability that men create when they act together as conceived by Arendt (1969) for example.

[130] Something like this is suggested by Green (1988:59-62). That structural accounts of power often come with a claim to legitimacy is argued by Lukes (1974) and Connolly (1974)

[131] Arendt (1969)

[132] Cf. Lukes (1974,1979,1986)

involves a capacity to intervene, that is, a capacity to change the 'natural' course of things. More in particular, second, power involves a capacity to change things that are *socially* significant. Thus the capacity to break a metal bar does not constitute power unless it has *socially* significant consequences. It is important to note that such intervention may be (a) benign or malignant, (a) the result of all acting in unity or of one actor acting against all others, and (c) actual or merely possible. Because I want to avoid confusion over latent capacities, however, I will from hereon refer to power *as if* it were actualised, that is, I will refer to power as *intervention* rather than the *capacity* to intervene.

From the perspective of power as intervention we can now focus more sharply on what conceptually distinguishes it from authority as conceived by the service conception. What is highlighted in comparing these two perspectives are the distinctively *normative* features of authority. Although these features can be understood from different angles, they all boil down to the normative *consequences* that authority has for those subject to it. As explained, the *normal* way to conceive of an authoritative command is to connect it with a correlate obligation to obedience. In contrast, the normal way to conceive of the exercise of power does not involve any normative consequences for those subject to it. To be threatened with sanctions in order to do this or refrain from doing that does not entail any obligation to submit to such threats. The best we can normatively make of coercion is that it may sometimes be justified in the light of the circumstances, but even then there is no *obligation* to submit to it. Joel Feinberg has conceived of the conceptual apparatus to make sense of this asymmetry between authority and (justified) power.[133] In this view authoritative commands involve so called "claim rights" which imply a correlate obligation for others to respect that right. Thus my claim-right to property, for example, entails your obligation to respect it. In the same way my claim-right to command entails your obligation to obedience. The exercise of power, on the other hand, does not entail any correlate obligation to obedience. Even when the exercise of power is justified does it result in so called "justification-rights" at best. Contrary to claim-rights, justification-rights do not have any normative consequences for others than those to which they apply. Thus my justification-right to self-defence, for example, does not entail any obligation for others to respect that right. It merely implies that acting in self-defence should not be considered wrong if it can be justified in the light of the prevailing circumstances. This conceptual distinction between authority and (legitimate) power is visualized in figure 1. This image presumes that authority over groups implies at least some power to do what one is justified to do. Hence there is no authority without there being at least the power to support it. The image depicted further shows that authority and legitimate power have differential normative consequences and that extending the scope of authority with that of legitimate power requires a *qualitative* reduction in normative claims – from claim-rights to mere justification-rights – over those over which it is exercised.

---

[133] The distinction originates from Feinberg (1970). The interpretation I use is that of Ladenson (1970, 1980/1990).

This reduction *a priori* weakens any attempt to come to a *complementary* understanding of authority and legitimate power.
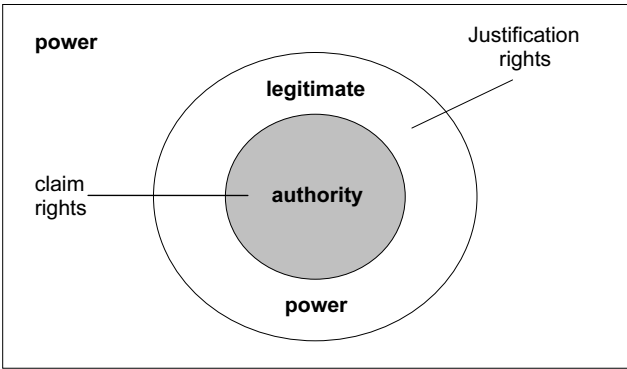

Figure 2.1 Authority and legitimate power

In this chapter, however, I shall argue against this *asymmetrical* understanding of authority and (legitimate) power. In my view this understanding hinges on a critical presumption that I believe to be invalid. This presumption involves what from hereon I will call the presumption of justificational asymmetry. While, according to this presumption, norms can justify intervention, the reverse is not possible. This presumption is widely adhered to. It is reflected in the saying that might cannot make right.[134] My argument against this presumption is relatively straightforward and is gradually developed in the following paragraphs. The major premise is that for practical norms to (a) work properly in guiding actions and (b) be binding, they must be *effective*, that is, they must *generally* be applied and complied with. There are two reasons supporting this premise. The first one is of a theoretical nature and involves a rule-following argument. The gist of this argument is that the proper functioning of a norm presumes a practice in which this norm is *actually* applied and conformed to. The second reason is of a practical nature. The gist of this reason is that one is bound to comply to a norm *if and only if* it can reasonably be assumed that others will comply too. In effect, this requires that the norm is actually applied and complied with. The minor premise now, is that not all norms we consider valid are *spontaneously* effective, that is, are effective without some external intervention *making* them effective. Since for some norms it is true that they can function and be binding only when they are supported by intervention, it follows that intervention is (partly) constitutive of such norms. Hence there are cases in which might does make right and the distinction between our *normative* understanding of authority and (legitimate) power loses its air-tightness, as well as the distinction between claim-rights and justification-rights that is

---

[134] An interesting exception to this saying has been developed by Hardin (1987)

founded on it. I conclude this chapter by surveying some of the implications that follow from its major argument.

Effectiveness and the Action-guidingness of Norms

I have pointed out three different – yet related – ways in which the service conception is a normative understanding of authority. Normativity is involved in (a) the *obligation* to obedience, (b) the *justification* of authority and (c) practical *reasoning* in general. The simplest way to conceive of the normative nature of authority is to conceive of command as an authoritative *norm*.[135] The case exemplifying understanding authoritative commands in this way is when the normativity of *A*'s command *a* – say: 'do not commit cruelty to animals'– is based on α which is a non-authoritative norm constituting a reason for action for *B,C,* and *D* – say, 'cruelty to animals is wrong' – such that *B,C,* and *D* are bound to the authority of *A* in all cases in which it is more efficient or effective to comply with *a* rather than to consider the application of α directly.[136] I have also indicated that I understand power as the capacity to intervene in the social status quo and that, because I want to avoid confusing talk of latent capacities, I refer to power as *intervention*, that is, as an *act* of intervention that demonstrates the (latent) capacity to perform that act. The presumption of justificational asymmetry, now, holds that while norms can justify intervention the reverse is not possible. If this presumption is somehow invalid, than the concurrent asymmetry between authority and power is ill founded. I will first examine what appears to be the least problematic part of this presumption. This is the part that says that norms can justify interventions and (particular) actions in general. It is clear that any justification of action by a norm presumes that the norm somehow applies to that action. The justification of an action by a norm therefore involves rule following: i.e. knowing when the norm (which is a rule) applies or not. A problem, however, is that following a rule is not something that may be taken for granted. Ever since Wittgenstein's first exposition of the problem we know that rule following is a philosophically puzzling activity.[137] The problem has been well documented in the literature.[138] In essence the problem arises because there is no way that the proper application of a rule or norm can be derived from any *finite* presentation of that

---

[135] Cf. Kelsen (1945)

[136] I depart from Raz's (1973/1999) conception of (mandatory) norms here. In my view norms are not authoritative if there is no authority *making* them authoritative by giving it the status of command. Authority is therefore a social phenomenon rather than something that figures in the practical reasoning of individuals.

[137] Wittgenstein (1953)

[138] Sellars (1963); Van Gunsteren (1976); Kripke (1982); Boghossian (1989); Minar (1990); Pettit (1990:1993/1996)

rule.[139] Since it is not obvious why or how this is so from merely stating the problem let me explain in broad strokes what it involves.

In order for a norm to apply to any action it must somehow have meaning to the subjects it addresses. In everyday life, it is presumed that a norm derives this meaning from doing two things. First, it depicts a certain kind of action. Second, it appreciates this action in an evaluative sense. Legal norms are perhaps paradigmatic here as the latter is done by specifying the sanctions to be put to use under the conditions depicted by the former.[140] What is wrongly presumed by this view, however, is that this depiction of the conditions to which a norm applies can be understood as being *exhaustively* and *unambiguously* referred to by that norm. To demonstrate that this presumption is problematic consider the example of a relatively simple and straightforward norm against speeding.[141] This norm says that it is forbidden to drive more than 50 kilometres per hour inside the city limits and threatens non-compliance with a fine of 100 Euro. It is already quite clear that this norm does not apply to emergency situations even if the norm itself does not make provision for such exceptions.[142] There are always exceptions to a norm and it is rather obvious that not all of them can be explicitly incorporated within that norm; hence the term: *exception*.[143] But even the non-exceptional application of this norm is less straightforward than it may seem. For this apparent simple and straightforward speeding norm requires it being embedded in a whole system of norms in order for it to be applied unproblematically. First, there must be (conventional) norms defining the concepts of 'kilometre' and 'hour' which in turn invite other norms that stipulate how these conventional norms should be operationalized under differential conditions. Unproblematic application of this speeding-norm, second, also requires that it be clear where the city-limits are and how they are indicated. The latter is possible only by virtue of more general norms stipulating the use of traffic signs, which in turn are dependent on a whole system of general traffic-regulating norms from which these signs derive at least part of their meaning. Moreover, there must be constitutional norms that prescribe how city limits can be instituted and changed. Third, there are also the norms stipulating what constitutes sufficient evidence of a violation of this norm, which in turn refer to a set of norms again that define how observation results in valid and reliable indications of the behaviour forbidden by the norm. Because the norm threatens with punishment, fourth, principles of punishment

---

[139] Be that the presentation of a "rule in intension" such as: 'for all x it is true that y' or the presentation of a 'rule in extension' which could be something like: {1, 2, 3, 4,…}. See: Pettit (1990, 1993/1996)

[140] This is how the German legal theorist Hans Kelsen defines a legal norm. See: Kelsen, (1945).

[141] My example constitutes a "rule in intension", that is "an abstract function that delivers a correct response as output for every situation presented as input". See: Pettit (1993/1996:81)

[142] Van Gunsteren (1976:111ff.)

[143] As we shall see later on, Carl Schmitt (1922/1985) even bases a theory of sovereignty on the notion of exception.

are applicable that define under which conditions punishment is appropriate, how punishment should be applied, by whom it should be administered, and so on. The general point I want to make here is that no norm *exhaustively* and *unambiguously* defines its conditions of application. The proper application of a norm always entails referring to other norms, which in turn refer to still other norms, and so on. This is not just true for practical norms but also for the theoretical norms that govern the application of concepts in language use.[144] The use of concepts is a normative activity because concepts can be applied either correctly or wrongly. The problem of rule following is therefore a problematic all the way down to our most taken for granted conceptual understanding in everyday life. It is because of this that it is typically conceived in terms of an infinite regress of rules upon other rules.[145]

The rule-following problem remains intractable, however, even if we disregard the threat of infinite regress for a moment and focus on the embeddedness of norms alone. We have seen that without a norm being embedded in a whole system of other norms there really is no way that that norm can be applied unproblematically. Yet this embeddedness of norms creates problems of its own. What kind of problems are involved can be explained best with the aid of an example again. Consider the case of a man testing how fast his car can accelerate at a traffic light on a road inside the city-limits. As soon as the light turns green he steps on the gas until he reaches the speed of 100 kilometres per hour. Having reached this speed he relaxes his right foot again until the car reached the legal speed limit of 50 kilometres per hour. Let us also assume that the police has put several radar-operated cameras along this road to enforce the legal speed-limit and that this man's car was photographed exceeding this speed-limit by two of them. Let us assume furthermore that all of this has happened in a country where the law rules supreme and the principle of *ne bis in idem* or 'double jeopardy' – i.e. a principle prohibiting someone to be prosecuted and/or punished more than once for the same offence – is a fundamental norm in criminal law. The question arises whether this man has broken the speeding law once or twice here and whether he should be punished once or twice accordingly.

At least two scenarios are conceivable. In the first scenario the principle of double jeopardy is taken to be such a cornerstone of the penal code that it is considered relevant prior to referentially interpreting the norm stipulating the speed limit. Consequently, the man is considered to be punishable only once, for he is taken to have acted in contradiction with the speeding law only once – even if he has been caught for it twice. It could be argued here that: (a) the man intended to commit only one act (determining how fast his car accelerates), (b) he performed only one action ('stepping on the gas'), and (c) the consequences of his actions can analytically be understood as a whole (a car exceeding the speed-limit). It is important to note, however, that all these reasons can be given both for determining the *internal* relationship between the principle of double jeopardy and the speeding norm and for

---

[144] This is, in fact, the point that Wittgenstein purports to make with his analysis of rule following. Wittgenstein (1953/1989) Cf. Sellars (1963)

[145] Sellars (1963)

relating the latter to *external* states of affairs in reality. The second scenario has a different outcome albeit for the same basic reason. In this scenario the norm prohibiting speeding is referentially interpreted without a priori considering the principle of double jeopardy relevant. What is primarily considered relevant to interpreting this norm are its methods of verification; one of which involves the radar-operated cameras on the roadside. Only once it is verified that a certain action contradicts the speeding norm is the principle of double jeopardy taken into consideration with the result that in this case the speeding norm is interpreted in such a way that each observation of 'punishable fact' constitutes a separate offence. This interpretation is considered reasonable because it seems inherent to the norms methods of verification that someone can be caught twice for stepping on the gas only once. It could be argued here that (1) not accepting this could seriously frustrate connecting this norm with any actions, and that (2) since a person can easily determine the duration of the action prohibited by the speeding norm it is not unreasonable to see every observation of facts contradicting the norm as a separate offence. Again, it is important to note that both arguments can serve the double purpose of establishing the meaning of a norm by connecting it with other norms in the whole system of norms and establishing its meaning by connecting it to the empirical conditions to which it applies

This example demonstrates, first, what we have seen to be true already. This is that no norm faces reality in isolation. The reason for this is that norms – and meaningful statements in general – derive at least part of their meaning from their connections with other norms in the (meaning) system in which they are necessarily embedded. Such systems may have many layers with norms governing highly specialized activity at one extreme, and the norms governing our ordinary language use at the other. Similarly, there may be norms that are very close to the activity that they aim to regulate and norms that are highly abstract and do not seem to regulate any activity at all. Van Roermund has given an example of such abstract norms in different areas of positive law. He refers to such norms as "legal keystones" because they *seem* to be mere tautologies that serve the purpose of integrating a variety of different legal norms into a systematic whole.[146] These norms say, for example, that one should behave as one is supposed to, or that no act is legally punishable unless there is a law to that effect. The second point, now, is that the connection between any particular norm within the whole system, on the one hand, and the empirical conditions to which it applies, on the other, is one of "underdetermination".[147] In other words: there is no *fixed* relationship between that norm and the part of reality to which it applies. The reason for this, again, is that at least part of any norm's meaning derives from its connection with other norms in the in which it is embedded. The consequences of this *holistic* feature of normativity is that the relationship between any part of that system – i.e. any particular norm – with the 'outside world' can be manipulated by changing the internal relationships between norms within that system.

---

[146] Van Roermund (1997), my translation.
[147] Quine (1963)

We have seen that empirically interpreting the speeding-norm may have different results depending on how one conceives of the (internal) relationship between the speeding norm and the principle of double jeopardy. Wittgenstein has radicalised this indeterminacy in his famous 'paradox of rules': "This was our paradox: no course of action could be determined by a rule, because every course of action can be made out to accord with the rule. The answer was: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here."[148] It is clear, then, that rule-following is philosophically highly problematic.

That rule following is philosophically problematic does not imply that it is problematic in practice, however. It is important to know that Wittgenstein did not merely state the rule-following problem but that the purpose of his analysis was to replace the at that time received view that concepts themselves exhaustively determine their application with his own *pragmatic* account of conceptual meaning. The gist of Wittgenstein's account of rule-following is that to follow a rule is to be engaged in a "practice"[149] in which that rule is actually "obeyed" – or acted against – rather than "interpreted" that is, scrutinized for purposes of clarification.[150] How a given rule should be applied in this view ultimately follows from the way it is *actually* applied in practice. There is nothing prior to that practice that makes any particular application right or wrong. Instead, the *proper* application of a rule (or concept) is 'fixated', so to speak, by the facts of the matter involved in its application. The problem of infinite regress is thereby simply short-circuited by breaking off the chain of justifications at a point where actions speak louder than words. In Wittgenstein's own words: "If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say: 'This is simply what I *do*'.[151] The general idea of this view, then, is that norms are not just necessarily embedded within a whole system of norms but ultimately also in practices – i.e. "language games"[152] or "forms of life"[153] – in which these norms are *actually* applied and complied with. Since the holistic feature of normativity makes the connection between norms and the empirical world to which they apply one of "underdetermination", it are these practices that fixate the proper application of any particular norm.

There are two questions that remain to be answered, however, before we can accept any pragmatic or naturalist account of rule following with confidence. First, it must be explained how rule-following behaviour actually comes about if rules do not speak for themselves. Second, it must be shown how rule following can be *fallible* – i.e. open to error and mistake – in an account of rule-following that proceeds from the facts *as they are*. It must be clear that no rule can be normative without fallibility,

---

[148] Wittgenstein (1953/1989: §201)
[149] Ibid. §202
[150] Ibid. §201
[151] Ibid. §217; emphasis mine
[152] For a definition: Ibid. §7
[153] Ibid. §241

that is: there could be no way in which applying a rule could be either right or wrong if there were not at least the possibility of it being wrongly applied. Philip Pettit has proposed an understanding of rule following that answers both of these questions.[154] In broad strokes his answer to the first question is that rule following comes about through the *inclination* that actors have or develop to extrapolate what appear to be similar cases in a similar direction.[155] Thus an actor can make sense of a rule (in extension) exemplified by the sample of numbers 1,2,3,4 by acting on the inclination to continue with 5,6,7, and so on. Now it needs to be remembered that any finite representation of a rule – in this case through exemplification – entails an infinite number of different ways in which that rule can be extrapolated. Thus one could extrapolate 1,2,3,4 with 5,10,15, if the rule followed were: "under five add 1 over five add 5" or one could continue with 5,6,7,8 and so on if the rule followed were simply: "add 1".[156] The inclination the actor has to continue in a particular way makes salient *one* of these infinite possibilities and thereby picks out, so to speak, the rule followed. There still remains the question how it is that rule following can be fallible in this account, given that the rule is picked out by the same behaviour that subsequently counts as rule-following. In other words: if picking out and following a rule and come down to the same stretch of behaviour how can rule-following ever go wrong? For isn't whatever one does right *by definition*? Pettit's answer to this question is rather complicated so I can present only a caricature of it here. In his view, rules can be applied in either good or bad conditions. Thus one can apply the concept of 'murder', for example, in conditions where one has a lot of information of what happened or in conditions where information on what has taken place is relatively scarce. Actors who experience sub-optimal conditions will be prepared to "discount" for these conditions *ex post,* and revise their actual application of a norm should they find reason to do so. They could, in principle, do so by comparing *their own* judgements at different points in time, but with practical norms it is typically the case that *different* actors correct *each other* in applying norms to different conditions. It is important to note that such a *social* pragmatic account of rule following connects well with our experience in everyday life. It can be recognized in many institutions of modern society. It is also consistent with the everyday view that applying a rule is something one learns only through experience and that *social* correction and reflective scrutinising of norms is an indispensable part of that experience.[157]

Let me return now to the reason why I have brought up the rule-following issue in the first place. This was to debunk the presumption of justificational asymmetry and the asymmetrical understanding of power and authority that is founded on it. This presumption holds that while norms can justify intervention the reverse is not

---

[154] Pettit (1991, 1993/1996)

[155] Pettit's understanding therefore involves a *dispositionalist* account of rule-following. See: Van Willigenburg (2000)

[156] Cf. Wittgenstein (1953/1989); Sellars (1963)

[157] Van Willigenburg (2000) stresses the reflective component in *ex post* practices of correction of response. I agree with his views insofar reflection itself is understood as a practice.

possible. We have started out by investigating what appeared to be the most unproblematic first part of this presumption, that is, the part that says that norms can justify intervention. On closer inspection it turned out that this part was more problematic than it seemed. The problem was that the justification of action requires rule following while rule following is philosophically highly problematic. In essence the rule-following problem holds that no finite presentation of a norm can exhaustively and unambiguously determine how that norm should be applied. It was explained, however, that our understanding of rule-following can be saved by understanding it in pragmatic rather than philosophical terms. In this view the answer to the question how a norm is to be applied follows from how that norm is *actually* applied in practice. This practice encompasses both individual extrapolative dispositions and social and inter-temporal *ex post* correction of habits of response. What is important for us here, however, is that a norm can therefore only *properly* be applied when it is embedded in a practice in which that norm is actually applied – and on the whole – complied with. It is therefore the effectiveness of a norm on which its action-guidingness ultimately depends.

Effectiveness and the Bindingness of Norms

In line with the social-pragmatic account of rule following proposed in the previous paragraph a norm may be considered to function properly "if, under normal circumstances, it fulfils its purpose".[158] Up till now, however, I have said nothing on how the proper functioning of *practical* norms should be conceived of. Nor have I said anything on how it is that the functioning of practical norms may be threatened. It is time to attend to these questions. A tentative yet robust answer to the first question is that *practical* norms serve to represent, create or uphold *legitimate* mutual expectations in social interaction.[159] Legitimate mutual expectations are mutual expectations that are justifiably held. They entail more than justified belief about future behaviour, however, as they involve *normative* rather than *positive* expectations. At face value normative expectations can be distinguished from positive expectations by their "different direction of fit".[160] When positive expectations do not accord with actual behaviour it is rational to revise them accordingly. The reverse holds true for normative expectations. When inconsistent with the facts, it is the behaviour, and not the expectations that should be adjusted to relieve the conflict. I can now be more precise about the proper functioning of practical norms. Practical norms serve to represent, create, or uphold normative mutual expectations in social interaction.

Although the distinction between normative and positive expectations is certainly useful in coming to understand the proper functioning of practical norms, it should not be taken as absolute. Depicting normative expectations as *unconditional*

---

[158] Wittgenstein (1953/1989:40-41; §87)

[159] An early protagonist of this view is Glastra van Loon (1956, 1958,1980, 1987). Much of what follows is inspired by his views.

[160] Vromen (forthcoming)

claims upon reality smacks of a normative metaphysical realism that has rightly been discarded for long in Western thought.[161] Thus while Plato could afford to conceive of his ideal city as immune to considerations of actual everyday life, it is nowadays typically presumed that normative expectations do not lay their claim upon the actual world *come what may*.[162] One reason why normative expectations are generally not conceived as unconditional is that there will always be conditions in which they do not, or do not fully, apply. We have already seen that honouring normative commitment *necessarily* involves the possibility of *ex post* correction for invalidating conditions. It was even claimed that there could be no such thing as normativity in the first place, if it were not for at least the possibility of *ex post* correction of dispositional habits of response. Correcting for invalidating conditions is not just necessary for *theoretical* reasons, however. There are also *practical* considerations why this is so. As we shall see later on, the possibility of normative or value conflict may seriously hamper the application of practical norms to particular situations.[163] Where two or more norms claim acts or omissions that cannot be realized simultaneously, for example, the straightforward application of any one of these norms may result in grave injustice. Moreover, I have insisted that there are always exceptions to a norm and that not all of them can be incorporated in the norms we have and live by at any particular point in time. Even if we could, the very fact that we then most certainly would discount a norm for *known* exceptions demonstrates my point, which is that normative expectations should not be understood as unconditional claims upon the actual world. The distinction between normative and positive expectations is therefore not very hard. The question is therefore not if, but *how* normative expectations and positive expectations are related.

There are, to be sure, quite a number of ways that normative and positive expectations can be related to each other. One way was revealed in the previous paragraph. It was explained there that normative expectations guide the application of the concepts we use in our positive expectations. Another way to conceive of the relationship between positive and normative expectations involves *naturalism*. Naturalism is particularly interesting because it is diametrically opposed to the view that I have just disqualified as inadequate, that is, the view that normative expectations make their claims upon the actual world come what may.[164] Naturalism *explains* normative expectations from the facts as they are. In this view, normative expectations arise or evolve out of the positive expectations we have towards the behaviour of others. Sugden provides an interesting contemporary version of such naturalism.[165] Following David Hume,[166] Sugden explains the existence of normative

---

[161] See: Korsgaard (1996) for a magnificent four-page history of the evolution of Western normative theory.
[162] Plato (1974)
[163] See the next chapter.
[164] Pigden (1991)
[165] Sugden (1998)
[166] Hume (1739/1978 book III)

expectations from the conjoining of human reactive sentiments with conventions. But where Hume appears to give primacy to these sentiments – which become moral only after the convention of language constrains or discounts them to "steady and general points of view"[167] – Sugden starts off from conventions and sees a role for reactive sentiments only after conventions have been established. Because in his view it is both "part of normal human psychology … to feel resentment towards individuals whose acts are both contrary to our expectations and contrary to our interests",[168] and typical for us to be uncomfortable with being the object of the resentment of others, it is only natural that normative expectations grow out of conventions or agreements for which it is true that it is in everybody's interest to comply, provided that others comply too. Thus in a naturalist perspective normative expectations rest completely on whatever positive expectations there are. They either build, or merely *supervene* on these expectations.[169]

Sugden seems to me right and wrong at the same time about the emergence and status of normative expectations, however. He is wrong, in my view, in turning Hume's story upside down. In line with the social-pragmatic account of rule following proposed in the previous paragraph it seems to make more sense to take emotive dispositions rather than conventions as primitive. Such dispositions guide us to whatever response human evolution has 'programmed' us to give. Only subsequently is this response corrected for the "steady and general points of view" typical of convention. Now I do not deny that one can have feelings of resentment towards those acting against already established conventions, as resenting non-compliant behaviour is perhaps exactly what conventions are about.[170] What I do claim, however, is that these feelings result in distinctively *normative* expectations only after they have been discounted and corrected from a more or less resilient general perspective which rests on some *higher order* convention making that perspective possible. I want to emphasize, however, that although convention therefore underlies normative expectations, one should not think too mechanically about how any such convention in particular comes about.[171] I say this because the conventional underpinnings of normative expectations may suggest exclusively behavioural or evolutionary origins given that the concept of convention connects well with behavioural and evolutionary theories of normativity. But this is not what is implied in saying that normative expectations are embedded in convention. It should be noted that with respect to any particular domain of action there is typically more than one

---

[167] Ibid. p. 581-582

[168] Sugden (1998:81)

[169] According to Vromen (forthcoming), Lewis conceives of normative expectations supervening on the positive expectations that constitute a convention.

[170] Lewis (1969)

[171] I do not mean convention in the narrow sense of the word here, but rather a more or less resilient practice in which a norm is actually applied and generally complied with. I will be more precise about convention later.

possible convention.[172] Picking out a particular convention may happen in a number of ways and certainly not exclusively through spontaneous evolution. In modern society we take great pride in *deliberately* and *reflectively* picking out the normative standards we live by. Deliberation and reflection play an important role anyway in the social process of *ex post* correction of whatever our dispositions may lead us to. "Reflective endorsement" is therefore in an important sense what practical normativity is about in modern society.[173]

Sugden is right in my view, however, in claiming that normative expectations *ultimately* rest on their positive counterparts. Although reflective endorsement is unmistakably in an important part constitutive of the norms we have and live by in modern society, it is simply not a *sufficient* condition for any particular norm to give rise to, or uphold mutual normative expectations in social interaction. As we have seen in the previous paragraph, the proper application of any norm rests on a practice in which that norm is actually applied and complied with. It is easy to see how this is true for the theoretical norms guiding our use of concepts. One would be quite puzzled about how to apply the concept of 'fiduciary duty', for example, if there were not a practice in which this concept was actually applied and generally complied with. Similarly, the notion of 'fiduciary duty' would not have any (practical) *normative* force for us if nobody to whom these duties supposedly apply actually behaved in accordance with what they require. Now I do not say that we would not admire anyone who lives up to these duties regardless of what others do. Most of us probably would admire a person displaying such integrity. What I deny, however, is that such a situation has any *normative* force for us, that is, that it results in any *obligation* for anybody to behave in accordance with what these duties require. We have seen that normative expectations 'transcend' the particular context in which relevant reactive responses are elicited, and that habits of response evolve into distinctively normative expectations only after they have sufficiently been corrected for invalidating conditions. I have argued that such correction requires a resilient general perspective from which these habits of response can be corrected by relating them to differential conditions under which they apply. In my view, such a resilient general perspective can be achieved *if, and only if,* one has reasonable assurance that the norm in question is generally applied and complied with. The reason for this is that pervasive non-compliance or even merely the expectation of pervasive non-compliance will critically influence the process of *ex post* correction of response in such a way that there is always *sufficient* reason to hold that the circumstances obtaining constitute invalidating conditions. That this is indeed so may become intuitively clear when holding a person to a norm nobody else complies with. Such a person may always excuse himself by saying that it is *unreasonable* to hold him to that norm in the light of the prevailing circumstances. It would always be valid defence for a person in those conditions to hold that since nobody complies with the norm in question there is really no *normative* ground to hold him to it nevertheless. The upshot of this paragraph and the previous

---

[172] As we shall see later, this is typical for problems of pure coordination.

[173] Cf. Korsgaard (1996)

one, then, is that although reflective endorsement is central to what normativity is about in *modern* society, it is not a *sufficient* condition for normativity. The reason is that a norm must necessarily be *effective* – i.e. it must generally be applied and complied with – in order for it to (a) be able to guide action and (b) be binding, because pervasive non-compliance, or even merely the expectation of it, critically undermines both the action-guidingness and the bindingness of norms. Whatever it is that reflection comes up with, and whatever it may be that actual behaviour results in, then, are therefore *complementary* rather than competing foundations of normativity.[174]

Having made my claim as to how it is that the proper functioning of practical norms must be conceived of, I now turn to the question how it may be threatened in practice. In a nutshell, the problem is that not all norms we reflectively endorse are *spontaneously* effective, that is, are effective without some outside intervention *making* them effective. There are many reasons why this is so, and it is clearly beyond the scope of this study to go in them in great detail. Some basic game theoretical insights, however, may help us to make sense of what is involved here by providing some "ideal-types" that characterize in broad exploratory strokes the paths that may lead to the effectiveness of norms being undermined.[175] In the remainder of this paragraph I will briefly discuss five of these ideal-types in order of the increasing difficulties they exemplify for the effectiveness of practical norms.[176] I pay attention in particular to what kind of intervention would be required to tackle these problems.

I have argued that the resilient general perspective that is necessary for the *ex post* correction of the application of norms rests on practices in which these norms are actually applied and generally complied with. Not having been very precise earlier, I have in some cases referred to these practices as *conventions*. It must be clear, however, that the concept of convention has received a very specific meaning in game theory. There, a convention is perceived of as a salient solution to a problem of (pure) coordination.[177] Problems of coordination exists when there are two or more Nash-equilibria – i.e. a combination of choices for which none of the players can improve his position by unilaterally changing his choice – that are equally valued by the players involved. When we translate this problem to the context of practical normativity we may think, for example of two competing conventional norms between which one is *reflectively* indifferent. Figure 2.2 formally depicts a simple version of this a predicament. Now it is true of conventions that they are self-policing once they have been established. The reason, in short, is that it pays for everybody to comply with the conventional norm provided that everybody else complies too. In an existing

---

[174] Sen (1998)

[175] Weber (1968)

[176] The following discussion of ideal types inspired by game theory does not have the ambition to be very precise. As said, I merely use different kinds of games to ideal-typically depict structures of mutual positive expectations underlying different kinds of norms. In the following I have made use of Hargreaves Heap, Hollis, Lyons, Sugden, Weale, (1992) and Hollis (1994)

[177] Lewis (1969)

convention this condition is sufficiently fulfilled. Again, driving on either the left or the right side of the road is probably the most famous example of a conventional norm. Once we know where it is that one is expected to drive, there is really no reason not to comply with these expectations.

Pete

	Left	Right
Left	1,1	0,0
Right	0,0	1,1

Pierre

Figure 2.2 A traffic convention

The problematic aspect of conventions therefore does not involve their sustenance but rather their emergence. For our purposes a typical problem exists where two or more norms have equal reflective support, but were only one of these norms can be effective – and hence both action-guiding and binding – at the same time. Now it is by no means excluded that conventions emerge spontaneously. As a matter of fact, many paradigmatic real-life conventions probably have. But as I have already briefly touched upon in the introduction of this study, it may not be prudent to wait until that happens. The stakes of general compliance may be very high from the onset – as is clearly the case in many traffic regulations – or conditions may change frequently such that the spontaneous selection of conventional norms may not be very prudent. In these cases intervention will help to make one of these norms salient and thereby secure the its effectiveness. Typically such intervention merely involves guiding actor's behaviour in complementary paths. Although such interventions clearly require some power, the kind of power involved here concerns the capacity that men have when they act in unison.[178] Once patterns of behaviour and relevant expectations underlie the norm in question, it is subsequently self-policing without any further intervention being required.

Things get more complicated when coordination problems combine with a relatively mild divergence of interests. It may be the case, for example, that all

---

[178] Cf. Arendt (1969)

European countries prefer to have a single norm with respect to issuing academic degrees over several different such norms, while at the same time each individual European country wants its own norms to be adopted as the standard. The explanation may be that there are sunk costs involved, for example, or that there is a resistance to change. The well-known battle of the sexes game illustrates this problem with great clarity (See figure 2.3). Here there are multiple coordination equilibria that are differentially valued by the actors involved. Again, it is the emergence rather than the sustenance of norms that is problematic here. But the emergence of a particular norm is more problematic in this case than in case of a problem of pure coordination because each player wants his own norm to be adopted. To secure the effectiveness of this kind of norms, intervention *may* be required to pick out one norm as salient.[179] But the kind of intervention required here involves more than merely the power to guide the behaviour of actors in complementary paths. As different parties differentially endorse different possible norms, it also involves *coercing* reluctant actors into what they consider to be an inferior equilibrium. Once a norm is selected, however, it does not pay not to comply because compliance to the norm now in place is generally preferred over non-compliance. It is noteworthy that here there is already

France

	French	English
English	2,1	0,0
French	0,0	1,2

England

Figure 2.3 'Battle of the sexes'

a sense in which *might makes right* here that was the object of analysis in this chapter, but I will come back to that later on.[180]

---

[179] Although these conventions may also evolve spontaneously, there will always initially be an element of coercion by dominant players or dominant coalition of players involved here.
[180] Cf. Hardin (1987).

More problematic than the 'battle of the sexes' predicament is when the structure of positive expectations underlying a norm resembles a so-called 'chicken-game' (See figure 2.4).

Pierre

	Swerve	Middle
Swerve	3,3	2,4
Middle	4,2	1,1

Pete (left side label)

Figure 2.4 A Chicken game

In the archetype of this game, two drivers of two cars purposefully follow a collision course. The driver that swerves first is the 'chicken'. A famous real-life norm that comes dangerously close to this structure of mutual positive expectations is the principle of 'no first use' in nuclear deterrence.[181] During the cold war the United States and the Soviet Union both claimed that they would not be the first to make use of nuclear weapons in case of a crisis, as both countries acknowledged that using nuclear weapons would have devastating effects for the world as a whole. Yet at the same time neither of these countries wanted to be 'caught with their pants down' in case of a nuclear strike by the other party, which they disvalued over an all-out nuclear exchange. But since not using nuclear weapons at all was still preferred by both parties over any use of nuclear weapons, the non-use of nuclear weapons was a prudent strategy, provided that the threat of a pre-emptive strike by the other party is sufficiently offset. In contrast to the coordination predicaments discussed above, however, the intervention necessary to create sufficient assurance against a pre-emptive strike in such a predicament involves an enduring rather than a single shot effort, as the need for assurance remains for as long as the predicament lasts. That intervention in a such a predicament need be invasive nor expensive is illustrated by

---

[181] The characterization of this predicament as a chicken game is not undisputed.

the fact that an apparently effective solution to the threat of a pre-emptive nuclear strike consisted of the institution of a direct phone connection between the Kremlin and the White house – the so-called Hotline – that was useful in getting a certain degree of assurance against *accidental* nuclear strikes at least.[182]

A far more problematic situation exists when the 'superstructure' of normative expectations rests on a 'basic structure' of mutual positive expectations that is exemplified by the predicament of the so-called prisoners' dilemma (see figure 2.5).

France

	Clean	Dirty
Clean	3,3	1,4
Dirty	4,1	2,2

England

Figure 2.5 A Prisoner's dilemma

The players in a prisoners' dilemma are locked into an inferior Nash-equilibrium because it pays for each individual player to defect on whatever normative commitment has been made regardless of whether other players are expected to comply or not. Norms protecting the depletion of common natural resources – such as the environment for example – are perhaps the most realistic exemplifications of this predicament. Thus while most countries may be expected to be committed to a norm or treaty limiting the emission of greenhouse gases because the global reduction of these gases is in the interest of all, each individual country will be tempted to defect on this commitment *ex post* because national industry will be in a situation of comparative advantage when it is allowed to produce dirty. A real-life perversion of a normative prisoners' dilemma involves drug-laws. While a society may reflectively decide on a prohibition of psychotropic substances in order to protect a minority of citizens from the harms of drug-use and drug-addiction, the actual implementation of that very same prohibition will create incentives for other citizens to produce and sell the drugs prohibited for a premium price precisely because these drugs are now illegal. The example of drug-laws illustrates nicely that the kind of intervention necessary to uphold norms that rest on a structure of mutual positive expectations

---

[182] The "Hotline agreement" was made in 1961 and modernized in 1971.

resembling a prisoner's dilemma predicament is typically enduring, coercive, expensive and invasive at the same time. It is highly questionable, for example, whether even the reflectively least disputed prohibition of drugs – that of opium-products for example – would survive any straightforward *utilitarian* analysis. For such an analysis is very likely to result in the conclusion that it simply does not pay to incarcerate a substantial part of the population in order to uphold a norm which has virtually no effect on the supply of drugs to the country involved.[183]

Arguably the most problematic structure of positive mutual expectations underlying a practical norm is exemplified by a so-called zero-sum-division-game. The structure of these games is such that one player's gain is another player's loss. It is therefore characteristic of such games that they ruthlessly divide the population of players into winners and losers. The most realistic examples of norms of which the effectiveness is threatened by a structure of positive mutual expectations resembling such games involve laws that aim to contain religious ideas and practices to the private sphere. Thus a law ordaining that evolutionary theory be taught in high school biology lessons may be expected to meet fierce opposition and non-compliance from the members of those religious groups for which evolutionary theory constitutes blasphemy. Now it is quite typical of such norms that they represent a majority position and that they are often found highly oppressive by the minorities whose values they intrude on. This already indicates that upholding such norms may involve *sustained* coercive intervention. Invasive intervention may be necessary when the behaviour forbidden is relatively easy to conceal, which is the case with female circumcision or forced marriage, for example. Often, but not always, such intervention is hard to reconcile with other norms applying, such as those expressing basic human rights and civil liberties.

It is time for me to summarize the argument made thus far. The upshot of the last two paragraphs is that norms can be action-guiding and binding only when they are effective, that is, when they are generally applied and on the whole complied with. The reasons for this were twofold. First, there was the theoretical argument that a norm can only be applied properly if that norm is embedded in a practice in which that norm is *actually* applied and complied with and which fixates, so to speak, the norm's proper application. Second, there is the practical argument that it is not reasonable to hold a person to a norm that is not generally applied and complied with, for such a person may always claim that pervasive non-compliance constitutes invalidating conditions for that norm. As a result, the distinctively normative expectations that the norm represents, creates or upholds rest on a structure of mutual positive expectations that is largely consistent with what is normatively expected. Relevant positive expectations even constitute a necessary condition for the existence of normative expectations in the sense that once these positive mutual expectations break down, the normative expectations upheld by them are bound to founder too. It was subsequently argued that since many norms we reflectively endorse are not spontaneously effective, intervention is required to make these norms

---

[183] Cf. Van Oosterhout & Kribbe (1995)

effective and hence action-guiding and binding. Different (ideal) types of situations were discussed that depict different conditions under which the effectiveness of norms may be threatened. From this discussion it can be concluded that to have and uphold norms for which the underlying structure of mutual positive expectations involve significant conflicts of interests (or values)[184], does not come cheap for the community to which that norm applies.

Normativism or Decisionism?

The previous paragraphs already concluded that under certain conditions intervention makes the difference between effective, and hence action-guiding and binding norms, and ineffective norms which are neither action-guiding nor binding. There is already a sense here, then, in which intervention is *constitutive* of normativity and in which might does indeed make right. This conclusion is nothing new. Thomas Hobbes was already quite explicit about both the requirement that norms be effective in order to be binding, and on the necessity of intervention in safeguarding the effectiveness of norms. In Hobbes' own famous words: "covenants without the sword are but words, and of no strength to secure a man at all. Therefore notwithstanding the laws of nature..(..).. if there be no power erected, or not great enough for our security; every man will, and may lawfully rely on his own strength and art, for caution against all other man".[185] A more contemporary version of the view that, in the end, it is might that makes right, must be attributed to the infamous Carl Schmitt.[186] Like those of Hobbes, Schmitt's views leave little doubt about the foundations of normative order.[187] Since his views involve a substantial *radicalisation* of both Hobbes' views and the findings in the previous paragraphs, discussing them may be helpful in more precisely formulating my own position on the issue at stake. Schmitt's views on the foundations of normative order are exposed most clearly in his dispute with Hans Kelsen on the foundations of legal order. This dispute is of particular interest to us because both Schmitt and Kelsen held that the effectiveness of a legal norm is a necessary condition for its validity and bindingness. Although the dispute between the two legal scholars is about the foundations of *legal* order and not on the grounds of *normative* order in general, the arguments involved can be extrapolated in broad terms from the former to the latter without much difficulty. Before bringing this chapter to its conclusion, then, I will first demonstrate how deep it is that the requirement of effectiveness and the central role of intervention in securing it bite in our understanding of normativity. I begin with giving a rough sketch of Kelsen's

---

[184] I will elaborate on value conflict in the next chapter. For now, it is sufficient to conceive of value conflicts as zero-sum division games.

[185] Hobbes (1651/1968: part I, chapter 13)

[186] Schmitt's bad reputation is mainly due to his position as 'Kron jurist' in the early days of Nazi Germany. See: Bendersky (1983)

[187] Cf. Benderski (1996)

normativism. I subsequently discuss Schmitt's *decisionism* as both a radicalisation and critique of this position.

Although the work of the Austrian born legal theorist Hans Kelsen is voluminous and comprehensive,[188] I will focus exclusively on Kelsen's position on the foundation(s) of legal order. Such a narrow focus doesn't do injustice to the comprehensive and elaborate character of Kelsen's views, however. His work is well known and often praised for its highly systematic and coherent nature. It is difficult, therefore, to get things out of context. Moreover, Kelsen's views on the foundation(s) of legal order belong to the core of his work and are primarily addressed in a work that takes a central position in his oeuvre: the *Reine Rechtslehre*.[189] In this work, Kelsen sets out to develop a *normative* science of norms for the purpose of developing a *pure* theory of law, that is, a theory of law rid of all metaphysical, moral and ideological influences.

Kelsen's position on the foundations of legal order consists of two elements. First, there is his understanding of the object of his science of norms: i.e. the concept of a legal norm.[190] Second, there is his development of a distinctive methodology for a *normative* science of norms. Kelsen's conceptualisation of the concept norm, to start with the former, is rather straightforward and has both a substantial and a formal side to it. The formal side of his conceptualisation consists of the requirement that a norm must consist of at least two things: (a) a description (or depiction) of the circumstances – in terms of human action, its conditions or its consequences – under which (b) a specified sanction ought to be applied.[191] On the substantial side, Kelsen defines a norm as the *objective meaning* of an act of will.[192] It is not equivalent to an act of will since an act of will is merely a fact – which is an 'is' and therefore not a norm which according to Kelsen is always an 'ought'[193] – but it is equal to its *meaning*, that is, the content of what is willed. To say that a norm is the content of an act of will and not the act itself is not sufficient for defining normativity, however. For this criterion alone does not exclude a thief's order to give your money or else pay for it with your life also to be a norm, while it is fairly obvious that it is not. A second criterion, therefore, demands that the meaning of an act of will is *objective* rather than subjective. This entails that this meaning must be robust over the different perspectives one can

---

[188] His personal bibliography counts more then six hundred entries (including translations of his major works in over twenty languages) addressing many different central issues in primarily legal but also political theory. See: Klanderman (1986:14)

[189] Kelsen (1934). Kelsen himself completely revised and extended the first edition in 1960, almost doubling the amount of pages of the original work. Because this study is written in English, I will refer to and cite from the English translation of the latter work by Max Knight (1989).

[190] Contrary to what has been done in this essay as a whole, for the sake of brevity, I will refer to 'legal norm' as 'norm' in the section in which Kelsen's position on the foundation(s) of legal orders is characterized.

[191] Cf. Kelsen (1945)

[192] Kelsen (1960/1989.3 ff.)

[193] Ibid.

take on what is willed. Thus not only must the act of will have a definite meaning to the agent willing – i.e. from a first-person's-perspective – but also to both its addressant(s) – from a second-person's-perspective – and an independent outsider – from a third-person's-perspective.[194] The meaning of an act of will, will be robust over different perspectives and thereby objective – and hence a norm – *if, and only if*, certain criteria are met. This is where a second constitutive element becomes relevant: the methodology of the normative science of norms.

Although Kelsen's normative science of norms is inspired by and modelled after the ideal of positive natural science – that is, science based on (a) facts rather than speculative reason and (b) a single and universal method rather than methodical syncretism – it nevertheless differs from this model in three fundamental respects. First, the object of Kelsen's normative science are norms and not facts as is the case in positive natural science. Although facts are certainly relevant to a normative science of norms, it is primarily their normative meaning and not their mere existence that is of interest to the legal scientist.[195] A second fundamental difference between Kelsen's normative science of norms and positive natural science concerns the ultimate goals at which scientific activity is aimed. While positive science aims at *explaining* positive facts in order to come to an understanding of nature as a whole, normative science is aimed at *validating* or *justifying* norms and actions in order to come to an understanding of a normative order as a whole.[196] These differential ultimate aims of positive natural science and normative legal science assume and require a third fundamental difference, that is, a difference in methodology. While the methodology of positive natural science consists of causal explanation, that is, causally relating facts to other facts in terms of general laws, the methodology of the normative science of norms is given by normative *ascription*, that is: by the validation or justification of norms and actions by relating them to higher norms in terms of so-called "rules of law".[197] These three characteristics provide a rough and ready picture of how the normative science of norms works in rendering objective meaning – and thus normativity – to a*ctual* acts of will.

In Kelsen's normative science of norms, the objectivity of an act of will is dependent on two requirements. First, it must be ascribed to a (higher) valid norm. Second, it should be effective, that is, it should generally be applied and complied with. The first requirement holds that the meaning of an act of will X can be ascribed to a (higher) valid norm *if and only if* the statement that X is a valid norm is the conclusion of a syllogism the major premise of which is the higher valid norm, and the minor premise states that X is a particular case of the (higher) valid norm. The latter can be the case in either of two ways. First, X can be logically deduced from the (higher) valid norm. Second, X can be posited in a manner *formally* required by the (higher) valid norm. Since, according to Kelsen, a legal order is inherently a *dynamic*

---

[194] Ibid.
[195] Ibid. p.70
[196] Ibid. p.71-75
[197] Ibid.

normative order – contrary to a religious order, which is inherently static in his view[198]– he focuses exclusively on the latter mode of ascription, that is, on formal rather than material ascription. The process of ascribing the validity of norms to (higher) valid norms goes on until the *ultimate* constitutional norms of a normative order are reached – i.e. those norms on which the validity of all other positive norms in the normative order are based, but which themselves cannot be further ascribed to any positive norm. Although, by definition, it isn't possible to ascribe the validity of these constitutional norms to any other positive norm, they are not grounded in normative nothingness in Kelsen's view. If the normative scientist is to *normatively* reconstruct a normative order, he can do so only by assuming that one ought to behave according to this order's ultimate constitutional norms. If he doesn't make this assumption, it isn't possible to normatively reconstruct any normative system at all, since one cannot understand a normative order as distinctively normative without it. It is crucial to note that this assumption – in spite of its epistemological nature – constitutes a norm*,* that is, an 'ought statement', and not some speculative judgement on how the world is at any point in time. Because of it being a norm this assumption actually functions as "Grundnorm" or "basic norm" of the whole legal order. Because of its epistemological necessity – without it there could not be any norm – it becomes the ultimate norm of the normative order, that is, a norm from which all other norms in the system derive their normativity. The Grundnorm or basic norm, therefore, is the foundation of legal order in Kelsen's views.

As was said above, however, ascription to a higher valid norm is not a sufficient condition for there to be a norm. According to Kelsen, a second necessary condition for the validity of any norm is its effectiveness. If a norm or even a whole normative order is not effective, that is, if it is not generally applied and complied with by those subject to it, it isn't binding and hence no one is bound to obey. This is the case, for instance, when a revolution has overthrown an existing normative order. Because the order thereby ceases to exist, its basic norm can no longer be assumed and therefore cannot serve as a foundation on which a new normative order can be built. In Kelsen's view, it is precisely the other way around. Only after a new normative order has been established by the revolutionary regime can one understand it as a normative whole again and can, or rather must, a new basic norm be assumed as its foundation. It is essential to note here, that according to Kelsen, the foundation of normative order is not some *substantial* first norm or principle whose validity is independent of the legal order that is built on it. Kelsen's ground for a normative order is rather a necessary ultimate assumption one must make in justifying any norm or action. It is, therefore, a transcendental and formal ground of a legal order. Because in this view the normativity of any legal norm is founded on the presumption of a Grundnorm it may justifiably called 'normativism'. It is precisely this foundational quality of this presumption that is questioned by Carl Schmitt.

Although Schmitt's work is about as comprehensive as that of Kelsen, it does not, at a first glance, appear quite as systematic and coherent. Where Kelsen's theoretical

---

[198] Ibid.

edifice seems meticulously put together from analytically clearly constructed components according to some pre-existing master plan, Schmitt's work appears more passionate than patient, more aphoristic than analytical, more reactive than guided by any plan. This impression, however, is quite deceptive. Often the systematic and coherent nature of an oeuvre as comprehensive and complex as that of Schmitt is 'created' only after it has been written, once the exegetics have gotten a hold of it. But although Kelsen and Schmitt were contemporaries, Schmitt's work for long has been something of a taboo due to his association with the Nazi-regime. Consequently, the exegesis of his work has really begun only relatively recently.

The essence of Schmitt's critique of Kelsen's normativism is that although Kelsen acknowledges the requirement of the effectiveness of legal norms as important to our understanding of normative legal order, he subsequently does not give this requirement the central position it deserves. The problems begin, in Schmitt's view, with the application of norms. In line with the view exposed in the previous paragraphs, Schmitt thought it to be self-evident that norms cannot apply themselves. He therefore concluded that the application of norms requires some *external* human action. A crucial problem, however, is that this action cannot be fully ascribed to any norm or normative system and is therefore – from the perspective of Kelsen's normativism at least – a mere fact and not a norm. As Schmitt formulates it: the action necessary for the application and realization of norms "emanates from [normative] nothingness".[199] This is, first, because as we have seen in the previous paragraphs, applying a norm is not some mechanical process of subsumption. According to Schmitt, no concrete application can be traced in the last detail to its normative premises.[200] The reason is that not only can no norm or normative system encompass all exceptions to it, but it cannot even determine whether given particular circumstances do or do not constitute an exception.[201] Although the immanence of the exception to normativity is, in Schmitt's view, the reason for the normative indeterminacy of the act of application, we have already seen in the two previous paragraphs that the normative indeterminacy of action in general is by no means an extreme or exceptional case. As explained, norms can only have meaning as part of a whole normative system and never in isolation. Consequentially, it cannot be determined on the basis of norms alone whether or not a particular action conforms to an individual norm. Hence, the normative indeterminacy of action is a central feature of normativity rather than some extreme or rare contingency. Since the

---

[199] Schmitt (1934/1985:32)
[200] Ibid. p.30
[201] The concept exception takes a central position in Schmitt's analysis of normative order. It is at the heart of his methodological strategy that focuses on the extreme rather then on normality. Schmitt adopts this methodology because, in his own words: "The exception (by definition the extreme case of a norm: HvO) is more interesting then the rule. The rule proves nothing, the exception proves everything.(…).In the exception the power of real life breaks through the crust of a mechanism that has become torpid by repetition'" Schmitt (1934/1985:15)

application of norms is not something that can be understood from the propositional content of norms alone, an 'external' perspective is necessary here. A plausible and viable perspective on how it is possible that norms can be applied at all was proposed earlier. In this view, the application of norms is a practice that combines dispositional habits of response with reflective *ex post* social correction. It is clear, however, that this approach to normativity can only work when norms are effective. As we have also seen, however, the effectiveness of a norm is not something that may be taken for granted.

In Schmitt's view, therefore, normativity is not given with there being (positive) norms. It also assumes there to be order. "There exists no norm" says Schmitt, "that is applicable to chaos."[202] Norms, first of all, presuppose a domain to which they apply. In Schmitt's view such a domain involves an *existential* social entity that encompasses and demarcates the subjects addressed by those norms. Such a domain cannot itself be normatively constructed. It is a precondition for normativity and therefore cannot be its effect. The construction of this domain is in Schmitt's view the effect of *political* action. It is the result of a social contradiction that has turned into an existential struggle; a struggle in which extreme violence is always a *real* possibility and in which everything is divided in the categories 'us' and 'them', or, to stick to Schmitt's own terminology: "friend" and "enemy".[203] Only after a political struggle has been won does the domain of normativity come to existence in the form of a political unit. In Schmitt's view, therefore, order and normativity always originate from political struggle and not from some pre-existential norm. But this does not make political action the foundation of normative order. For the mere existence of a political unit does not guarantee there being order in it.

Since according to Schmitt, normative order is the conjunction of there being norms and there being order, there is always the possibility of it becoming disjunctive again.[204] This happens, for instance, when norms become ineffective and normative order deteriorates as a result. As was explained earlier, the ineffectiveness of a norm causes it both to lose its action-guidingness and its bindingness. In Kelsen's view they thereby even cease to be (legal) norms. When such conditions apply, intervention becomes necessary to restore the legal normative order. A pivotal question, however, is how it is determined that such a situation obtains and warrants intervention. It is clear that norms alone cannot determine whether this is the case. As was argued earlier, norms cannot exhaustively and unambiguously determine the conditions under which they apply. But neither can we rely on an existing practice in which the relevant norm is applied and on the whole complied with, because the lack of effectiveness of the norm in question means that there simply is no such practice anymore. The situation in which intervention to restore the effectiveness of norms is warranted is therefore, from a normative perspective, an *exceptional* situation. Again,

---

[202] Ibid. p.13
[203] In Schmitt's view, the distinction between friend and enemy is the basic distinction underlying all politics. See: Schmitt (1927/1996)
[204] Schmitt (1934/1985:12)

this does not mean that this situation is rare or exceptional in an empirical sense. It only means that this situation is normatively undetermined.

This exceptional situation, or as Schmitt calls it, "state of exception",[205] is normatively undetermined for two distinctive reasons. The first reason has already been explained extensively. Norms simply do not unambiguously determine whether or not an exception exists and whether, as a consequence, intervention to restore normativity is warranted. If there is to be a state of exception at all then, it can be declared only from the order-side of the conjunction of normativity and order. A second reason for the normative indeterminacy of the state of exception is that even once it is clear that such a state obtains, it cannot be normatively determined what may or may not be done in order to restore the effectiveness of legal norms and hence, in Schmitt's view, of normative legal order as such. This is because relevant norms have ceased to be in the state of exception. Deciding on the state of exception and on how normative legal order is to be restored is therefore a normatively undetermined activity. The actual and existential actor taking this decision occupies a supreme position in the normative legal order because, in Schmitt's own words: "Sovereign is he who decides on the exception"'.[206] It is, after all, the *decision* of this sovereign whether there exists a state of exception or an effective and binding legal order. It is, says Schmitt, a *decision* in the pure sense of the word. It is undetermined by any possible norm or even a remotely normative argument and can only be ascribed to the existential actor making the decision. This decision becomes the ultimate foundation of the legal normative order, since it can be made at all times and a priori overrides any reflective 'normative' consideration because without effectiveness no normative consideration has any action-guiding or binding force. Because it is ultimately the capacity to make such a decision on which the bindingness and action-guidingness of legal norms rest, Schmitt's position has appropriately been referred to as decisionism.

Conclusions

It goes without saying that both Kelsen's normativism and Schmitt's decisionism are extreme positions in the theory of normativity, and that neither of these positions is completely right about the foundations of practical norms. Yet, at the same time, each of these two extremes emphasizes different elements central to our understanding of what normativity is about and of the kind and scope of the claims norms can make to be applied and complied with. Kelsen's normativism, with its strong reliance on the deductive ascription of action and norms to higher order valid norms, emphasizes the deliberate and reflective elements that play a role, not only in the genesis of norms, but also in their endorsement once they exist. Moreover, we have seen that reflection is an important part of the social practice of *ex post* correction of habits of response, and that if it were not for such practices, there could be no fallibility and hence no

---

[205] Ibid.
[206] Ibid.

normativity in the first place. In modern society we typically take great pride in selecting and upholding the norms we live by in a deliberate and reflective manner. This does not mean that all norms we have and uphold in modern society have been reflectively chosen. Nor does it mean that all norms we live by are deliberately sanctioned in everyday social interaction. There exist many conventions we have neither deliberately selected, nor particularly care about in a reflective sense, but that are nevertheless worth complying with once they are there for the simple reason that it is in our interest to comply provided everybody else complies too. We even may, and typically do, resent non-compliance to such conventions for precisely that reason. What it does mean, however, is that in modern society it has become increasingly important that the norms we have and live by can survive critical scrutinising; that norms that are found to be unreasonable or even outright inhumane, are abandoned and, if necessary, replaced with more reasonable and civilized norms that can and do pass the test of critical reflective analysis.

We have also seen in this chapter, however, that reflective endorsement is not a *sufficient* condition for normativity. It is noteworthy that even Kelsen subscribes to this position.[207] The reasons why reflective endorsement is not a sufficient condition for normativity were twofold. First, there is the *theoretical* argument that no norm be action-guiding if it is not embedded in a practice in which that norm is actually applied and, on the whole, complied with. This does not mean that no new norm can ever develop or be posited, nor that the application of a norm is forever fixed once it is deeply embedded in a practice of application. It does mean that that there are limits to the extent that the application of a norm can vary, and that norms lose their action-guidingness once these boundaries are crossed. This is intuitively already quite clear. Practical norms that are open to many different interpretations guide relevant action poorly if at all. Second, there is the *practical* reason that the ineffectiveness of a norm constitutes an invalidating condition for the application of that norm. This is also intuitively quite evident. It is not reasonable, after all, to hold a person to a norm nobody else cares about complying with. In sum, then, the mutual normative expectations that a norm represents, creates or upholds, ultimately depend on a structure of mutual positive expectations that is, on the whole, consistent with what these normative expectations require. Once the structure of positive expectations breaks down, the normative expectations are bound to founder too.

The central problem identified in this chapter is that not all norms we reflectively endorse and care about upholding are spontaneously effective, that is, are effective without some 'outside' intervention making them effective. Although there are many ways in which the effectiveness of norms can be threatened, I have pointed out five 'ideal type' situations – derived from common or garden game theory – in which the structure of mutual positive expectations may undermine whatever normative expectations there are. Without the ambition to be exhaustive, I have discussed problems of pure coordination, the battle of the sexes, chicken games, the prisoner's dilemma and zero-sum division games. It is safe to assume that many of the practical

---

[207] Kelsen (1960/1989)

norms we actually care about having and upholding in modern society will not be effective without external intervention making them effective. This is quite clear for those norms we have deliberately imposed upon behaviour we think to be in need of regulation. But it may also be true for many taken for granted real-life conventions in modern society. Even the relatively undisputed traffic convention of driving on the right side of the road is occasionally in need of intervention to secure its effectiveness, because the self-policing mechanism typical of conventions does not seem to work on highways with more than one driving lane. Now I do not claim that driving in the left lane on a highway will undermine driving on the right side of the road in general. I do claim, however, that many real-life conventions may not be naturally effective under all possible conditions to which they apply.[208] The upshot of the requirement that norms be effective in order to be action guiding and binding, then, is that intervention, or at least some dormant capacity to intervene, may be inevitable to uphold any norm or normative order.

Decisionism is therefore right, in my view, in emphasizing that some capacity to intervene is inevitable in any sustainable normative order. I have explained that the effectiveness of the norms we reflectively care to have and uphold can be threatened in many different ways, and that the kind of intervention required to select or uphold practical norms depends on the structure of mutual positive expectations underlying the norm. It was argued that, as a rule, the stronger the conflict of interests (or values) underlying a norm, the more intensive and invasive the intervention necessary to secure its effectiveness must be. I have claimed that the kind of power constituting the capacity to intervene varies from the benign collective capacity that men have when they harmoniously act together, on the one hand, to the outright coercive asymmetrical power necessary to police or impose norms based on underlying conflicts of interests (or values) on the other. Decisionism is also right, in my view, in claiming that it is impossible to *fully* anticipate the nature, extent and invasiveness of the intervention necessary to uphold the normative order we reflectively endorse at any point in time, and that it is therefore equally impossible to define an exhaustive and watertight normative mandate that normatively constrains the intervening actor. There are two important reasons for this. First, there is the universally applying condition of "bounded rationality" that puts definite limits on our information-gathering and processing capacities at any point in space and time.[209] Even if everything is known about the state of the world at a certain point in time and the (natural) laws that apply to it, we may still lack the capacity to compute and predict what will happen in the future. Second, there is the more fundamental problem that such a mandate itself constitutes a norm and is therefore subject to the requirement that it be effective in order for it function properly. The reason for this, again, is that the application of a norm does not follow from its propositional content but rather from the ongoing practice in which that norm is *actually* applied and on the whole

---

[208] Strictly speaking, such norms cease to be convention once the need for policing arises. It is doubtful, of course, whether there exist any pure conventions in everyday reality.
[209] Simon (1945/1998)

complied with. Intervention, however, may be too incidental or idiosyncratic for such a practice to develop and settle in, or a relevant practice may not have developed *yet*. As a result, it may very well tread into terrain that is (yet) normatively undetermined. But decisionism makes a much stronger claim than saying that intervention – or at least some capacity to intervene – is unavoidable for any sustainable normative order, and that it may tread into normatively undermined terrain. What decisionism claims is not just that intervention may be normatively undetermined, but rather that it is normatively *unconstrained* because it is the sole and ultimate foundation of normative order. Although the claim that normative indeterminacy is the rule rather than the exception seems excessively radical, I have not dismantled it in this chapter. A thorough examination of it will have to wait until the next chapter. For now, the only question remaining is how the major argument of this chapter bears on the conceptual distinction between authority and power, or to stick to the terminology introduced in this chapter, between authority and intervention.

Authority was normative *if, and only if,* obeying a command helps those subject to it better to comply with reasons which apply to them already, than if these independent reasons themselves are acted upon. The requirement that norms be effective in order to be action-guiding and binding, however, together with the fact that not all norms are effective without some intervention making them effective, already indicates that intervention may very well create reasons for action that actors do not have without such intervention taking place. There is thus a real sense here in which might can make right and in which intervention can *create* authority as conceived of by the service conception. There is therefore no clear-cut boundary between authority, on the one hand, and the classic Weberian conception of authority as legitimate power, on the other (see figure 2.6).
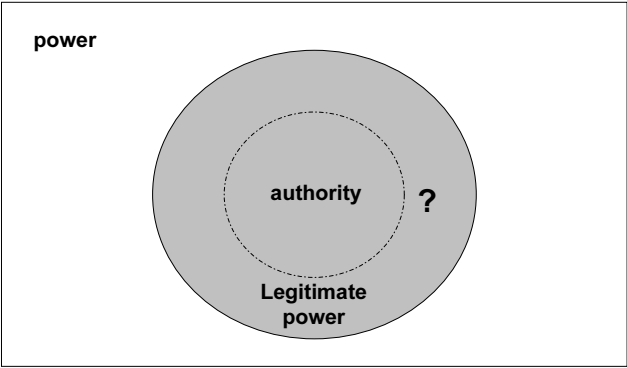

Figure 2.6 Authority and legitimate power revisited.

This has two important consequences. That there is no such distinction, first, seriously undermines our efforts to conceptually come to grips with the nature of authority, for an important way in which the claim to legitimacy could be incorporated in our understanding of authority has now been cut off. It is clear that whatever claim to legitimacy can still be made must extend to power or intervention as well. But if there is no conceptually airtight demarcation between authority and legitimate power, second, then neither can there be a conceptually sound distinction between "claim-rights" and "justification-rights" that is founded on it. If the argument made in this chapter makes any sense, then in the end there is just power that is legitimate and power that is not. But even that distinction is a difficult one to make, or so at least I shall argue in the next chapter.

III

Intervention, Authority and the Problem of Dirty Hands.

Introduction

The major conclusion of the previous chapter was that there is no clear-cut conceptual distinction between authority as conceived of by the service conception, on the one hand, and the classic Weberian account of authority as legitimate power – understood as the capacity to intervene in the social status quo – on the other. The gist of the argument behind this conclusion was that the distinctively normative features that allegedly distinguish authority from legitimate power may themselves only be realized *through* intervention. It was argued that in order for any norm to be (a) action-guiding and (b) binding, it must be effective, that is, it must be generally applied and on the whole complied with. Since not all practical norms are effective without some kind of intervention making them effective, intervention may be constitutive of practical normativity. Consequently, might can indeed sometimes make right. It was even claimed, albeit only implicitly, that reflection holds no privileged position over intervention in the constitution of practical normativity. For in the same way that intervention is not necessary to safeguard the effectiveness and hence the action-guidingness and bindingness of a practical norm, it is not necessary for such a norm to be reflectively endorsed either. There are many action-guiding and binding norms in modern society – typically of a conventional nature – that neither have come into existence in a deliberate manner nor are dependent on reflective endorsement for their sustenance. Now it may be true that in modern society reflection always plays at least a *virtual* role in the sustenance of practical normativity, that is, that reflective endorsement is often only implicit and that reflective capacities only become activated when norms are actually perceived of as unjust or inferior in terms of welfare consequences. More or less the same can be said, however, of intervention. Since even self-policing conventions may occasionally be in need of purposive upholding, it has proven prudent to organize at least some latent capacity to intervene should things go wrong. I have claimed that decisionism is therefore right, both in asking attention for such latent capacities, and in stressing the need to have some capacity to intervene operative in the background of any sustainable normative order.

I have claimed that decisionism is right too in claiming that it is simply impossible to fully anticipate the nature, extent and invasiveness of the intervention necessary to have and uphold a normative order we reflectively endorse at any point in space and time, and that as a result, it is equally impossible to define an airtight normative mandate constraining the intervening actor. This was not just for reasons of bounded rationality, but also because such a mandate is itself a norm and is therefore subject to the requirement that it be effective in order to be binding and action-guiding. I have explained that because intervention may be too incidental or idiosyncratic, relevant normative practices may not have fully developed or settled in yet, and that intervention may therefore tread into normatively undetermined terrain. I now want to address the more radical decisionist claim that intervention may not just be normatively undetermined, but that it is normatively *unconstrained*, in the sense that the norms that apply at any point in space and time cannot determine the outer limits of what intervention may come down to. We have already seen that there is no clear conceptual demarcation between our normative understanding of authority, on the one hand, and legitimate power, on the other. I now want to investigate the question whether there can be made a conceptually sound distinction between legitimate power – i.e. power constrained by binding and action-guiding norms – on the one hand, and power as such, on the other. It is important to note that the distinction between power as such and legitimate power is the final frontier at which the claim to legitimacy can somehow be incorporated in our understanding of authority. If it can be defended, then nothing stands in the way of interpreting authority the classic Weberian sense as legitimate power.[210] If it cannot, then perhaps we should abandon our efforts to conceptually come to an understanding of authority by incorporating the claim to legitimacy within that understanding.

The method by which I want to investigate the question whether or not there is a clear-cut boundary between legitimate power and power as such, is by evaluating the impact of an interesting anomaly to the hypotheses that intervention can be normatively constrained.[211] This anomaly involves one of political theory's most troublesome insights. It concerns the problem that, sometimes, political leaders, managers or other persons in positions of authority must act wrong in order to do good, that is, they must break or set aside valid – i.e. action guiding and binding – norms, for the sake of some overriding social or political concern. In contemporary political theory this insight is often referred to metaphorically as the problem of "dirty hands",[212] after Sartre's play bearing the same name.[213] This play evolves around the

---

[210] Weber (1968)

[211] In analogy with Kuhn's (1962) famous understanding of this notion, I understand an anomaly to be a special case – either actual or merely counterfactual – that is critically inconsistent with the body of theory or (normative) doctrine in respect to which it is a special case, and that has at least the potential of falsifying that theory or doctrine.

[212] Cf. Walzer (1973); Howard (1977); Hollis (1982); Thompson (1987); Stocker (1990); Coady (1991); Cunningham (1992); Winston (1994); De Wijze (1994,1996); Badaracco (1997); Kaptein (1998); Wempe (1998)

dilemma a political party faces that has not come into power yet. On the one hand, there are the moral and ideological principles the party stands for and which provide it with a 'raison d'être'. On the other hand, there are the concrete and practical strategies that must be followed to obtain a position of power which is necessary to realize even the most insignificant objectives the party may have adopted. It is not difficult to imagine where and how the two can clash, and set the stage for someone to dirty his hands, that is, for someone to violate the principles the party stands for in the quest for the power necessary to realize those very same principles.

For the problem of dirty hands to falsify the hypotheses that intervention can be normatively constrained it must be shown that: (a) the problem of dirty hands is a *real* problem indeed, that is, that it denotes some existential and empirical issue and not some conceptual confusion, and (b) it has a systematic rather than merely accidental relationship with intervention and authoritarian governance. If, on the one hand, the problem of dirty hands turns out to be a conceptual confusion in the theory of normativity, then clearly it couldn't tell us anything about either normativity or intervention, let alone about the relation between the two. If, on the other hand, the relationship between the problem of dirty hands and intervention is merely an accidental one, then it does not seem justified to draw any general conclusions from it, regardless of their direction. This chapter is five sections. First, I will trace the problem of dirty hands to its ancient, classic origins and compare these ancient understandings with more contemporary perspectives on the problem. Since, in my view, both contemporary and ancient sources lack sufficient conceptual precision, I subsequently propose a conceptual perspective on the problem of dirty hands that is more appropriate to contemporary analysis. On the basis of this perspective I then probe how deep the problem of dirty hands bites with some of our most fundamental understandings of normativity. Having ascertained that the problem of dirty hands constitutes both a real and a deep conceptual problem for the theory of normativity, I conclude this chapter by investigating the relationship between our conceptual perspective on the problem of dirty hands, on the one hand, and intervention and authoritarian governance, on the other, in order to evaluate our hypothesis that intervention is sufficiently constrained by valid and binding norms.

The Problem of Dirty Hands revisited.

Although the problem of dirty hands is often attributed to the writings of Nicolo Machiavelli,[214] it is not very likely that Machiavelli is it's originator, nor is it arguably the most original feature of Machiavelli's work.[215] It is more likely that the problem dates to a much earlier period of time. It can even be argued that it goes back to the very origins of Western culture. Evidence for this claim is given by the fact that crude variants of the problem are a relatively common theme in some of the earliest texts

---

[213] Sartre (1947)
[214] See *The Prince* (1514/1995) but also his *Dicourses* (1516/1970)
[215] For both claims see: Berlin (1953/1997)

available to us. Next to the bible, Greek tragedy and philosophy prove a particularly rich source of this.[216] In her classic book, Martha Nussbaum has pointed out the particular significance of Greek tragedy in this respect. According to Nussbaum, this ancient literary genre testifies not only of the possibility but also of the inevitability of what she calls "tragic conflicts", that is, situations in which choices have to be made and actions have to be taken that are both wrong and right at the same time. Often such choices involve choosing 'the lesser evil' of a limited set of options an agent is confronted with. Nussbaum's discussion of Aeschylus' play 'Agamemnon' is particularly helpful in understanding what these situations involve.[217] It also raises questions, however. In particular it raises questions as to how the problem of dirty hands may conceptually be distinguished from similar, yet arguably different problems.

In Aeschylus' play, Agamemnon, king of Argos, is commanded by Zeus to head a punitative expedition against Troy. The purpose of the expedition is to avenge the betrayal of hospitality by Paris – son of the Troyan king Priamus – who had escaped to Troy with the wife of his host Menelaos. The expedition, however, is halted at Aulis due to bad winds. These are said by the prophet Calchas, to have been caused by Artemis, who apparently has some unfinished business with Agamemnon.[218] The prophet divines that the expedition will remain stranded unless Agamemnon sacrifices his daughter Iphigenia. Already there are men dying. If nothing happens everyone, including Iphigenia and Agamemnon himself, will die too.

Although the choice facing Agamemnon is indeed a difficult one, it does not appear to be much of a dilemma for him. A dilemma typically involves at least two approximately equally desirable – or undesirable – alternatives of which ultimately only one can or must be chosen. What is problematic, therefore, about a dilemma, is that the often very different (constellations of) reasons for choosing either one or the other of the available alternatives roughly balance each other out, thereby providing little guidance as to which choice is to be made.[219] This is not true for the choice facing Agamemnon. In his case it is quite obvious which alternative is to be chosen. If he does not sacrifice Iphigenia, not only will she die anyway, but also will he himself and all the men gathered with him at Aulis. There are not just consequentialist considerations at stake, however. If Agamemnon does not sacrifice his daughter the divine commands of both Zeus and Artemis will be disobeyed. This would constitute

---

[216] According to Nussbaum (1986) tragic choice and dirty hands (she does not distinguish the two) were already a prominent theme in Greek tragedy, in particular in the work of Aeschylus. Stocker (1986) argues that the problem of dirty hands has an *explicit* place in the work of Aristotle. Weber (1919/1982.) suggests that the theme is not an exclusive feature of Greek pantheism, but that it is a theme with which all major religions have struggled, and continue to do so.

[217] Nussbaum, (1986:25-50)

[218] Nussbaum, (1986)

[219] Note that defined in this way, a dilemma may very well be exclusively the result of our inability to decide the issue. As will become clear later on, I believe the problem of dirty hands to involve something deeper, more disturbing than this.

a serious religious offence. But we do not need Greek pantheism – nor any other reference to religious duty for that matter – to arrive at the conclusion that there are other than consequentialist issues at play here. Even if we interpret the situation as being the haphazard result of the way the world can sometimes work against us, the fact that Agamemnon is heading the Greek alliance against Troy brings with it special responsibilities; at least for the men under his command. Agamemnon seems very well aware of all this. After having briefly moaned his predicament, his mood changes from passive depression to commitment to what is inevitable. Iphigenia is sacrificed and the expeditionary force can be on its way. But that is not where the story ends. The stage is merely set for more significant things to come.

It is telling that Agamemnon – having been victorious at Troy – is killed upon his homecoming by his own wife Clytemnestra. Although it is a well-known fact that most murders are committed by persons close to the victim, it nevertheless strikes us as something difficult to grasp. One wonders how things could have come that far. More significant than the mere fact of this murder, however, is that the play unmistakably suggests that Agamemnon deserved what was coming to him; that although the sacrifice of Iphigenia was inevitable – and therefore the right thing to do under the circumstances – it was also a heinous crime "leaving the man who [did] it guilty of a moral wrong".[220] We must appreciate the significance of guilt and blame carefully here, though, for it is often misunderstood.[221] What is significant is not whether a man in Agamemnon's situation, or in one sufficiently like it, *actually* feels guilt, remorse or regret – even though we may feel that the man who does is somehow deserving of our compassion and perhaps even praise.[222] What is significant in Aeschylus' play is that despite the fact that Agamemnon *himself* appears not to be troubled by these reactive sentiments at all, he is nevertheless judged by *others* (the chorus and Clytemnestra) to be guilty of a (moral) wrong, and therefore at the very least *ought* to feel guilty. If guilt, or better blame, is to be significant in such cases, then, it cannot be suffered in quietness, as Weber's politicians do.[223] These reactive sentiments, in general, can *only* have normative significance if they are socially expressed, as Walzer believes they should be.[224] But it must be emphasized that it *need not* be expressed *by its subject*, as Aeschylus' play most vividly demonstrates. It should rather be expressed by the people surrounding the subject – the moral community if you will – in more or less the same way they would blame a man who had violated the norm yet without the "practical necessity" of Agamemnon's predicament.[225] What the

---

[220] Walzer (1973:161)

[221] For reasons explained later, by Bernard Williams (1981:27) for instance, when he introduces the notion of "agent regret". Perhaps even by Nussbaum (1986), insofar as she interprets Agamemnon's guilt primarily as a demonstration that ethics cannot do without some form of virtue ethical approach as is developed by her in later chapters of her book.

[222] On the lack of normative significance of feelings in general see: Howard (1977:34-35)

[223] Weber (1919/1982)

[224] Walzer (1973)

[225] I will come back on the notion of practical necessity later on.

play ultimately demonstrates, then, is the philosophical insight that even though the world may force us to obey one norm at the cost of breaking another, this does not, by itself, render the norm broken without any claim to obedience – not even in the particular and perhaps tragic situation in which breaking the norm is considered necessary.[226] Although Aeschylus' play is – even at present – convincing in its demonstration of this complex and disturbing philosophical insight, it also has its limitations. What remains hidden in the narrative and the poetic way in which it is expressed are some interesting conceptual nuances that may help us understand what differentiates the problem of dirty hands from other problems.

It is not just in ancient texts, however, that the problem of dirty hands is clouded by conceptual vagueness. If there is one thing immediately apparent from even a limited survey of the more contemporary literature on the problem, it is that many different problematic situations are classified as such without it being either made explicit or implicitly clear what these problems have in common and how they can be distinguished from similar, yet arguably different problems. Insofar as attempts are undertaken to bring some conceptual order to the matter at hand, these generally point in different directions; thereby typically creating more confusion than is being solved.[227] Before going into the matter in some more detail, consider this small sample of very different cases, all of which are claimed to involve dirty hands:

1. Sophie's Choice. A nazi forces Sophie, a mother of two, to choose which of her children is to be saved from being killed. If she does not choose, both her children will be put to death.[228]
2. Torturing a terrorist in order to save the lives of many innocent civilians.[229]
3. In order to save hostages, a deal is made not to prosecute certain terrorists.[230]
4. Nuclear deterrence. The strategic policy to threaten potential aggressors with thermo-nuclear annihilation with the purpose to prevent war.[231]
5. A manager laying off a large number of good and loyal employees in order to secure the firm's continuity.[232]
6. Two parents choosing to amputate a child's limb, in order to avoid an unknown risk of a malignant tumour (that can also be removed by less invasive surgery) spreading throughout the body.

The conceptual ambiguity over dirty hands already manifests itself when the question is asked what it is that these cases are examples of. Some authors believe

---

[226] For more or less comparable arguments see: Williams (1973)

[227] See for example Coady (1991) but also Winston's (1994). Although Winston makes an interesting distinction between three different types of dirty hands, he seems to assume that it is clear what constitutes dirty hands in the first place.

[228] Stocker (1990:19)

[229] Walzer (1973)

[230] De Wijze (1994:31)

[231] Thompson (1987)

[232] Wempe (1998)

that these cases represent what can be called 'right versus right' issues,[233] or, what comes down to the same thing: cases in which the good is the enemy of the best.[234] Others, though, interpret these cases as choices between two evils, with the lesser evil clearly being the preferable alternative.[235] It is also possible, however, to interpret these cases as actions that are *both* right *and* wrong at the same time.[236] Now of course the question whether the problem of dirty hands is to be conceptualised in terms of either right *or* wrong, and perhaps even in terms of right *and* wrong, is a rather trivial one. As we shall see in the next paragraph, it is not very difficult to find some overarching concept in terms of which the seeming confusion over this issue can be resolved. It remains, however, that the dissensus over the answer to this question is representative for the *actual* conceptual debate on the problem to date. Where there still is disagreement at the level of descriptions, conceptual clarification seems a distant ideal.

More serious disagreement, however, arises over the question which cases do, and which cases do not involve dirty hands. Even with respect to the small sample of cases presented above some unsettling questions can be asked. It is questionable, for example, whether cases number 5 and 6 involve dirty hands, for it can be argued that these actions hardly have any public significance while the problem of dirty hands is thought to be situated somewhere in between public and private. In this view, it is understood as a kind of discontinuity between the two spheres.[237] Another example of extensional ambiguity surrounding the problem of dirty hands concerns the case of Sophie's choice. Some authors believe this famous and rather dramatic case to be paradigmatic for dirty hands.[238] In their view, it illustrates how dirty hands cases always somehow involve the 'evil projects' of others than the agent confronted with it. To this view it can be objected, first of all, that it is difficult to see why one should blame Sophie at all. There seems neither action nor attitude warranting legitimate blame in this case. The choice Sophie is forced to make can hardly considered a choice at all. It is a choice only in a nominal sense with precisely those elements missing that could make it a suitable ground for reactions of praise and blame. Furthermore, her subsequent grief and suffering over the matter seem more or less what may be expected form anyone in her predicament. Secondly – and this is more important in the present context – there are other concepts available in terms of which this case can be interpreted that would arguably provide a better fit than interpreting it in terms of dirty hands does. Plausible alternatives would be concepts like 'moral dilemma' and 'tragic choice'. Although these concepts are clearly closely related to the problem of dirty hands, they are not identical with it. For neither moral dilemmas nor tragic choices necessarily result in dirty hands, the latter involving

---

[233] Badaracco (1997)
[234] Hollis (1982)
[235] Nielsen (1996)
[236] Walzer (1973); Stocker (1990)
[237] Nagel (1978)
[238] Stocker (1990) develops this argument, while Nielsen (1996) supports it.

precisely the sort of blame that seems difficult to attribute in cases like Sophie's choice. This points to a more general problem in more recent attempts at conceptualising the problem of dirty hands. They make everything but clear how the problem of dirty hands relates to similar, yet arguably different problems. It seems, therefore, that the contemporary debate on the problem of dirty hands is just as much in need of conceptual clarification as ancient texts on the subject are.

Dirty Hands: a Characterization

A potentially fruitful way of imposing some conceptual refinement on the matter at hand is to ask the question what is *characteristic* for dirty hands, or, to put it more accurately: what would *characterize* the problem of dirty hands against the *background* of all problems that appear, at least *prima facie*, similar to it. Now it needs to be noted that following such a conceptual strategy will not result in a conceptualisation that will neatly determine which particular cases involve dirty hands and which cases do not. The reason for this is that the strategy of characterization can at best result in a *relative* conceptualisation of dirty hands, for what is characteristic for some class of phenomena is dependent upon the background against which it is discriminated from other classes of phenomena. Since there is no fixed (pre-) conceptual background against which the problem of dirty hands can be characterized, the following characterization may not be *exhaustive*. In other words: it cannot be excluded that there exist cases of dirty hands that are not encompassed by the characterization. But neither may it be *exclusive* for it remains possible that phenomena are characterized as cases of dirty hands against one background while these cases do *not* involve dirty hands against another background. An example may illustrate this.[239] Although withdrawable nails are characteristic for cats against the background of all domesticated animals, this is not true against the background of all felines. Against this background it is the feature of domestication that is characteristic for cats, because all felines have withdrawable nails, yet only the cats among them are domesticated. Although, therefore, the conceptual strategy proposed here will not result in a referentially unambiguous conceptualisation of dirty hands, it is nevertheless worthwhile pursuing. For it may very well result in a conceptual *perspective* on the matter that will enable us to distinguish dirty hands from similar problems in a novel and robust way. Such a perspective would not only make us see things differently because of the new conceptual glasses it provides, but it could perhaps also point us to new 'solutions' for the kinds of deep existential problems that are generally associated with the problem of dirty hands. Eventually, it may even lead to a more articulate conceptual framework providing precisely the sort of conceptual refinement both ancient and contemporary texts are lacking.[240] Before we can even begin to think of creating a whole conceptual framework, though, a crucial step must first be taken. This step involves answering the question which general feature can

---

[239] I thank Hans Koeze for this example.
[240] It is not the purpose of this chapter to spell out such a framework, however.

serve as an appropriate background for the perspective to be created. In answering this question, one crucial consideration must be taken into account. Since we do not want to exclude plausible candidates being dirty hands cases beforehand, the general feature that is selected as the background must be prevalent in all cases that appear, at least at face value, candidates for being dirty hands cases.

It is not difficult to give a tentative answer to this crucial first question. The many examples in the literature that are claimed to involve cases of dirty hands appear at least in one crude way similar to each other. They all, in one way or another, involve actions taking place in a context of normative conflict, or, as it is more commonly referred to, a context of value conflict.[241] Such a context is given by any situation in which (at least) two equally ultimate norms or values guide an agent in taking different courses of action, yet with each course of action excluding the other(s). It is essential to note that the notion of normative conflict involves something deeper, and therefore more serious than any epistemological inability to choose between two or more – what are at least perceived to be – competing values. For if such were the case, value conflicts could in principle be resolved by either conceptual or cognitive means alone. The example of Aeschylus play, however, has already demonstrated that this cannot be true. Even if Agamemnon would have had some kind of divine ability to perceive and understand his situation in a way unavailable to him as a mortal, he may have decided the issue with (even) more certainty of being right than he actually did, but he could not have escaped his predicament without the horrible cost he caused and the subsequent blame attributed to him by the plays equivalent of a moral community. In fact, the whole concept of *cost* seems difficult to conceive if conceptualised *exclusively* in conceptual or epistemological terms. In sum: value conflict, operationalised as the inevitable incurrence of *real* cost, seems to be a necessary condition for dirty hands. But it must be clear that it is a necessary and not a sufficient condition. Dirty hands may be inconceivable without there somehow being cost involved, but cost is not yet equivalent with blame. The context of normative conflict just constitutes a background against which the problem of dirty hands can be characterized.

It has been argued that what may turn cost into blame and thereby normative conflict into dirty hands, is the distinction between moral and non-moral value.[242] In this view, only a loss in moral value will result in blame, which, as was argued earlier, *is* normatively significant, while a loss in non-moral value will just constitute cost which on its own *is not*. But taking the moral non-moral distinction as characteristic for dirty hands creates more problems than it solves, if it solves any at all. This is, first of all, because making such a bold distinction is questionable in the first place. For one thing, an attempt to separate the many different guises of normativity in terms of moral and non-moral value may be about as feasible as an attempt to classify everything one encounters on a market as either herring or fruit.[243] It is difficult to see

---

[241] See: Williams (1973, 1981); Berlin (1978, 1979); Nagel (1979); Raz (1986); Chang (1997)
[242] Williams (1973:166-186)
[243] Cf. Pitkin (1972:219)

how such an attempt could be successful without at some stage creating arbitrary results. But even if we could safely assume that a distinction between moral and non-moral value could be completely non-arbitrary, the question remains what it is precisely that this distinction would achieve. In trying to resolve the issue it would be very difficult to avoid begging the question: i.e. making dirty hands a moral problem because it involves moral value.[244] It seems better, therefore, to avoid the issue of morality altogether.

This is even more so since there is a way to turn cost into blame, and thereby value conflict into dirty hands, without the need for any prior distinction between moral and non-moral value. For this to become clear we must revert to the earlier conclusion that in order to have any normative significance blame must be socially expressed, albeit not necessarily and certainly not exclusively by the agent himself. If blame is to have normative significance it should rather be expressed by those surrounding the agent. Since we aim to transform cost into blame, the same should therefore hold true for costs. But why would cost be socially expressed in a way that blame generally is? And what would make cost *socially* significant to the extent that its social expression would become warranted? These seem difficult questions to answer *in abstracto*. Perhaps a concrete example will help to focus in on what is at stake here more closely. Imagine a person that chooses to finish an important academic paper he is writing rather than to attend his aunt's funeral overseas. This person would, in all likelihood, not bother us with the real costs he incurs as a result of his choice. More importantly, however, there would be little reason for us to be interested in the choice he makes in the first place. This would be different, however, if that person were the chair of an academic department deciding for the all faculty of that department that the completion of papers trumps honouring private commitments. In this case the person causing the cost will have something to explain to his fellow academics. The latter may even justifiably blame this person if they think he made the wrong decision.

What this examples suggests is that what can turn cost into blame and thereby normative conflict into dirty hands, is that deciding the issue or taking the action causing the cost happens on behalf of *others* than the agent executing it: when, to put the same in a more general wording: the action or decision taken is *representative* for the interested parties. There is no ground to attribute any blame to an agent if this agent acts purely on his own behalf, with no consequences for the interests of others whatsoever. The ground to attribute blame in such cases only comes into being with the involvement of others, whether that be intentionally or as a matter of fact. This *representative* feature is therefore necessary to turn cost into blame. But although necessary, the representative feature may not be sufficient. The example given above also illustrates that for cost to turn into blame it is additionally required that the action is *actually* disapproved of by the interested parties. The introduction of this *contingent* element, however, does not prevent the representative feature from being taken as a characteristic for dirty hands. For although the representative feature is not a sufficient condition for blame in a *causal* sense, it is sufficient as the *ground* for its

---

[244] See Nusbaum (1986) for comparable arguments.

attribution; the latter being relevant from a *normative* perspective. Moreover, the representative feature (R) is a characteristic for dirty hands (D) against the background of all value conflict (V) if the condition is fulfilled that any value conflict (V) that is not a case of dirty hands (D) does not involve representative action (R), and this may be the case regardless of whether the action concerned is *actually* disapproved of by those represented by that action (see figure 3.1).
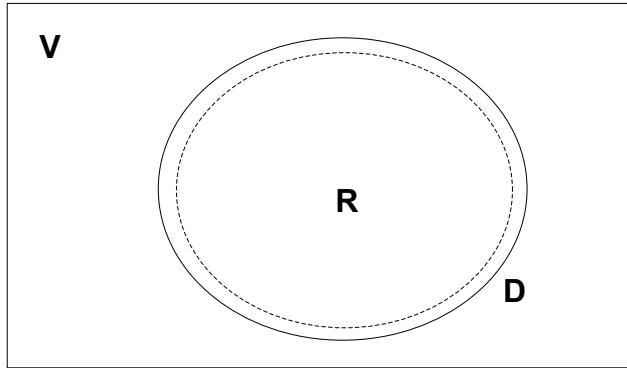

Figure 3.1 Dirty Hands: a characterization

I propose, therefore, that what characterizes the problem of dirty hands against the background of value conflict is the *representative* feature of the cost-causing-action. It must be noted, again, that the characterization of dirty hands in terms of value conflict and representative action is not referentially unambiguous. It was already explained that it may very well be exhaustive nor exclusive. What the characterization of dirty hands does accomplish, though, is to create a perspective from which dirty hands can be distinguished from other forms of value conflict, thereby providing a conceptual perspective that both ancient and contemporary literature are lacking.[245]

With the advantage of (conceptual) hindsight it is possible to explain why this perspective has been lacking for so long. It appears that while the conditions under which the problem was first developed have changed dramatically, the debate on dirty hands and related problems has never really outgrown the assumptions of its classic and ancient origins. The "face to face" features of the classic Greek polis in terms of

---

[245] It must be noted that although both Walzer (1973) and Thompson (1987) implicitly identify the representative feature as an important component of dirty hands, both fail to systematically incorporate this feature in a conceptualisation of dirty hands: Walzer because he relates dirty hands to representative action only as a matter of contingency; Thompson because he seems to confuse representation with democracy, thereby excluding all kinds of non-democratic representative action such as management decisions in corporations. Both contributions are, in my view, nevertheless the most systematic to date in the literature on dirty hands.

which the problem is commonly understood make the conceptual refinement proposed here more or less impossible, because it blends what I believe are different kinds of normative conflict into uniformity.[246] In Aeschylus' play, for example, the person to be sacrificed – not coincidentally – concerns Agamemnon's own daughter. This makes his predicament primarily a personal tragic conflict and maybe even a moral dilemma; for it is not at all clear how the highest of personal obligations should be weighed against the most important of public duties. What is blurred, though, by the rather dominant personal dimension of the normative conflict in which Agamemnon finds himself, is that the action causing the cost is also *representative* for interested parties other than Agamemnon, such as Iphigenia who eventually had to bare the costs, but also for the men under his command for they would have had to pay had Agamemnon decided not to sacrifice Iphigenia. In modern society – if only because of reasons of scale – those who are sacrificed for the benefit of some overriding social or political concern are very seldom the daughters of those deciding or doing what has to be done. They are more likely to be just anonymous victims, having no relationship whatsoever with those who cause their suffering. This fact of modern life, therefore, not only makes it possible for a conceptual distinction such as the one proposed here to be made more clearly, but it also makes it more necessary. For with the divergence of personal grief and the pain of others caused by these tragic actions, some sort of natural break on choosing lightly in these cases went lost. New ways have to be found to compensate for this loss. New conceptual perspectives may guide us to these ways. The characterization of dirty hands in terms of value conflict and representative action may very well be such a perspective.

An important advantage of this characterization over understandings of that do not distinguish dirty hands from other kinds of normative conflict, is that it explains at a conceptual level why the problem of dirty hands is so often associated with politics.[247] We cannot, as some authors seem to do, dismiss this obvious relationship as a mere empirical coincidence.[248] That would deny the fundamental and irreversible changes that have taken place in many different kinds of human associations; most noteworthy in society at large. These changes necessitate a conceptually more refined approach to the different conditions under which normative or value conflict applies; one that does justice to the increasingly *impersonal* nature of relations in human association. The conceptualisation of dirty hands proposed here does just that. It relates dirty hands to politics, not as some accidental empirical coincidence, but rather – as we shall see later on – in more systematic and perhaps even conceptual terms. For politics, almost by definition, is concerned both with normative conflict and

---

[246] See: Laslett (1963). Howard (1977) has pointed to a relationship between Laslett's characterization of Greek political thought and the problem of dirty hands. He does not, however, share the view I aim to develop in the text. His point is merely that the development from the Greek polis into modern society has asked much of our moral rules and intuitions "when they are extrapolated form individual to public cases".

[247] Cf. Walzer (1973), Thompson (1987), Williams (1981)

[248] Walzer (1973)

representative action. It is therefore not just that the problem may occur more often in politics because of the adversarial way politics tends to be organized and the kind of persons that are attracted to it.[249] It is rather because politics – involving the kind of representative action that it does – inevitably enters a context of normative or value conflict, coping with which seems to be a major function of politics in the first.[250] Before focussing on the more systematic relationship between the problem of dirty hands, on the one hand, and authority and intervention, on the other, I first want to answer the question precisely how problematic the problem of dirty hands is for our understanding of normativity.

The Depth of the Problem

An important presumption implicit in the foregoing characterization of the problem of dirty hands is that it is a *real* problem indeed. It is a real problem in a banal sense, first, in that it seems practically unavoidable. Its characteristics are such ordinary features of our day-to-day reality that they seem difficult if not impossible to avoid. The fact that we often make decisions on the basis of at least some crude form of cost-benefit analysis suggests that value conflict – operationalized as the inevitable incurrence of *real* cost – is a common assumption we make about the way we are committed to the world as it is. At least equally common is representative action, that is, action that is in some way undertaken on behalf of others.[251] This is a rather straightforward consequence of the fact that man associates in communities and modern man in many different ones at the same time. The problem of dirty hands is a real problem in a more philosophical sense secondly, because once encountered it is inescapable. As was argued earlier, it cannot be dismissed as an error of judgement or the result of some kind of epistemological or conceptual confusion. For even if we take away all confusion and error, we may still have to act bad only to do good, and be guilty of a wrong even when that wrong can be excused in terms of its consequences.

But it is not just that the problem of dirty hands is a real problem and just that. The problem of dirty hands involves something more than merely a nasty contingency for any particular agent confronted with it. It also challenges our most general and systematic views as to what norms are and how and why they can claim obedience from those addressed by them in the first place. To say the same thing more concisely: next to a real problem, the problem of dirty hands is also a *conceptual* problem for the theory of normativity.[252] It is a conceptual problem because up till now, no general and systematic position in that theory has proven able to either solve it or render it harmless. This is a far-reaching claim to make. Because it involves all

---

[249] Williams (1981)

[250] Cf. Easton (1965)

[251] Cf. Pitkin (1967, 1968)

[252] Which I take to be roughly equivalent with what is commonly referred to as meta-ethical theory.

viable theories of normativity that have been formulated up till now, it is not possible to argue its tenability *in extenso* here. Within the framework of this study any argument in favour of this claim is necessarily limited both in scope and in depth. At best it could touch only superficially on a limited sample of what are really caricatures of comprehensive and elaborate theoretical positions. The results of such an exercise may nevertheless be telling. Both the 'sampling' and the 'construction' of caricatures can be done with the purpose of the whole exercise in mind. By selecting only those positions in the theory of normativity that are *prima facie* conducive to the problem of dirty hands in the first place, and by subsequently enlarging only those features of these positions that bear on the subject at hand while at the same time ignoring features that do not, something of a *crucial test* for the claim made above can be designed. For if even these relatively conducive theories of normativity cannot deal with the problem, then it is not to be expected that any theory of normativity can. In constructing this crucial test, three area's in the theory of normativity appear particularly susceptible to the problem of dirty hands: (a) the area delimited by deontology on the one hand an consequentialism on the other, (b) the realm described by what will subsequently be called theories of normative reasoning, and c) the domain of value pluralism.

Between deontology and consequentialism
The problem of dirty hands is commonly presented and analysed in terms of a clash between deontological normative reasoning on the one hand, and consequentialist or utilitarian normative reasoning on the other. The wrong that is committed is generally justified by consequentialist arguments, while deontological arguments subsequently play a crucial role in determining that even while the wrong committed is on the whole the better thing to do, it is still a wrong. It is even claimed that it is only possible to condemn a practically necessary action as wrong on the basis of norms that claim obedience come what may.[253] But while combining consequentialist and deontological theories of normativity in an *ad hoc* manner may yield good examples of dirty hands cases, neither of these individual theories, nor the more systematic combinations of the two theories acknowledge it to be a problem at all. This is most obvious for a straightforward and undiluted deontological theory of normativity. According to such a theory, a norm depicts an action as either right or wrong and always so. As a consequence, a deontological theory of normativity does not allow for one and the same action to be both right and wrong at the same time. In deontological terms: there is nor can be a problem of dirty hands. There is just right and wrong, regardless of what the world may have in store for us.

It was already said that deontology is not alone in dismissing the problem of dirty hands as a form of conceptual confusion at best. The same is true for its

---

[253] Nagel (1972:126), for example, considers a deontological position on the rules of war "often the only barrier before the abyss of utilitarian apologetics for large-scale murder." In his view, deontology "operates as a limitation on utilitarian reasoning, not as a substitute for it". Ibid. p.128

consequentialist counterpart at the other extreme. Roughly spoken, a consequentialist theory of normativity judges the desirability of an action by its consequences. If, from the perspective of the community as a whole, these are valued positively overall, the action in question is considered right. If they are valued negatively it is judged wrong. As a result no action can be both right and wrong at the same time. Consequentialism simply explains away the clash between two competing norms by assuming one of the two norms invalid. It has few problems with this because consequentialists consider normativity as little more than a kind of shorthand for more complex consequentialist analyses. Thus, according to a consequentialist theory of normativity, torturing a terrorist in order to save the lives of innocent civilians does not constitute dirty hands. If the negative consequences of not doing so outweigh the negative consequences of torturing this terrorist, it is simply the right thing to do. If not, it is plain wrong.

It was already hinted at that it are not just the straightforward and undiluted guises of consequentialism and deontology in which there is no place for the problem of dirty hands. The same holds true for more moderate combinations of these two theories such as: (a) rule utilitarianism and (b) weak consequentialism.[254] Rule utilitarianism distinguishes itself form plain utilitarianism in that the object of utilitarian analysis are norms and not actions. Once it has been established that a norm is justified in terms of utilitarian arguments it is considered valid and binding whatever happens, thereby making irrelevant considerations of expediency.[255] In this respect there is no difference between rule utilitarianism and deontology. Once it has been established that something is wrong in general it cannot be right at the same time anymore.

Weak consequentialism is different from both plain consequentialism and rule utilitarianism in that it does not aim to replace deontological normative reasoning with consequentialist considerations. It rather sees the latter as a limitation on the former.[256] In general, weak consequentialists can afford to be rather indifferent to the sources of obligations, as long as it is acknowledged that circumstances matter with respect to the validity of a norm's claim to obedience.[257] If these are such that complying to a norm results in consequences far worse than those that are the result of breaking the norm, the norm loses its bindingness and thereby its validity under those circumstances. In short, there exists no problem of dirty hands as far as weak consequentialists are concerned. Given certain circumstances there are just norms that make a legitimate claim to obedience and norms that do not. In this view, acknowledging the problem of dirty hands to be a problem at all would just add to the confusion already present in any situation that warrants reconsidering a generally valid norm.

---

[254] Nielsen (1996)
[255] Brandt (1972)
[256] Not that this is the mirror image of Nagel's (1972:128) position.
[257] Nielsen (1996)

Theories of normative reasoning
It is not only along the lines of the opposition deontology – consequentialism that theories of normativity have been diluted, so to speak, to accommodate for the circumstances under which norms are to be applied. There exists a line of normative thinking that is relatively indifferent to this opposition: one that is equally applicable to both consequentialist and deontological theories of normativity as well as to combinations of the two. According this view, it is never quite clear at face value what it is precisely that a norm claims from the subjects addressed by it. The reason for this is that this is dependent on the circumstances under which it is to be applied. In this line of normative thinking, norms depict only *prima facie* obligations.[258] What it is precisely that a norm claims from subjects only becomes clear in its confrontation with practice. In this view, establishing what should be done in a particular situation is the result of a process of *normative reasoning* for which both norms and the circumstances under which they are to be applied are important inputs.

Theories of normative reasoning come in different variants: ranging from complex to relatively simple and straightforward. According to one of the simpler views, all general norms need to be adjusted to practice before they can make a particular claim to obedience. The reason for this is that although norms that have a high degree of abstraction and generality may be well suited for purposes of communication and education, they are not very useful in determining what is right and what is wrong in a particular situation.[259] For that purpose, norms must still be adjusted to the variety of different circumstances under which they are applicable. One way doing this is to try to incorporate in the norm all possible exceptions to it. This generally results in either very complex norms or in whole (sub)systems of norms. The core of such systems, then, consists of the most general formulation of the original norm – i.e. the form in which it is suited best for purposes of communication and education – while all kinds of application-norms constitute the fringe.

A more complex theory of normative reasoning makes use of the concept of *reflective equilibrium*. Originally developed by Rawls,[260] the reflective equilibrium approach to normative reasoning is the keystone of a strong current in the theory of normativity: one that has a strong presence in the contemporary literature on normativity in general. In very bold terms, the concept *reflective equilibrium* refers to the outcome of a process of normative reasoning in which general normative principles are systematically and reflectively confronted with facts and intuitions with respect to the case at hand. This process goes on until the general principles and the basic intuitions are mutually adjusted such that there exists a coherent set of beliefs on which judgement can be based. The reflective equilibrium approach is attractive for both its broad applicability and its familiarity. It is able to avoid some contestable assumptions while at the same time it resembles the most widely accepted conception of the scientific method – often referred to as "the wheel of science" or "the received

---

[258] The terminology derives from Ross (1930)
[259] See: Hare (1972)
[260] Rawls (1971)

view"[261] – in that both the reflective equilibrium approach and this conception of the scientific method combine deductive and inductive reasoning in a way that gives priority to neither.

It is difficult to see, though, how theories of normative reasoning can contribute to either solving or coping effectively with the problem of dirty hands or with any other kind of serious normative conflict for that matter. With respect to the more simple variants such as the one discussed above it must be noted that these generally rest heavily on the assumption that exceptions can be adequately incorporated into a norm or a whole system of norms. It was argued in the previous chapter, however, that this assumption is untenable. It erroneously dismisses rule-following problems as technical, while it was already established that referential ambiguity is indigenous to normativity and constitutes a fundamental rather than a technical problem. Although the more complex theory of normative reasoning that was also discussed here proceeds form from less controversial premises, it is equally problematic for one pressing reason. It is difficult to imagine how a reflective equilibrium can be achieved between the general normative principle that torture is wrong and the basic intuition that in particular circumstances it may nevertheless be the expedient thing to do. Reconciling such opposite claims would clearly be to much to ask from the reflective equilibrium approach to normative reasoning. Besides that, it would probably also over-stretch any minimally consistent normative system.[262] What is a problem for theories of normative reasoning regardless of their complexity is that they approach normative conflict as an epistemological or conceptual problem rather than a real problem, that is, as a problem that can be solved once the correct analytical tools and concepts are available and all relevant information is provided for. That this assumption is clearly mistaken needs no further argument at this stage. In summary, then, we may conclude that theories of normative reasoning have little to offer in coping with the problem of dirty hands. Although they do not, as other theories of normativity do, simply deny its existence, they fail to appreciate the full depth of the problem.

Value pluralism

A wholly different position in the theory of normativity proceeds from the assumption that – at any given point in space and time – there exists a whole variety of different ultimate values that cannot be reduced to a single unifying principle. This assumption often involves more than the mere observation that the world as it is shows there to be many different values that may ultimately not be compatible with each other. It typically involves the more far-reaching thesis that a notion according to which all that is good is ultimately one and the same thing is incoherent with some of our most common and deeply held views and intuitions: notably that the concepts of real *cost* and *loss* are not a priori meaningless.[263] This position is commonly known

---

[261] Cf. Wallace (1969); Lipsey (1963)
[262] The issue of minimal consistency is addressed later in this paragraph.
[263] See: Williams (1979); Berlin (1990)

as *value pluralism*. As is the case with most general theories, value pluralism stands for a family of related theories rather than for a single unified body of theory. For present purposes, however, it is sufficient to roughly distinguish only two kinds of value pluralism.[264]

A first kind of value pluralism applies to the level of individual norms or values. At this level, it acknowledges that individual norms or values – freedom and equality for example – can not be reduced to each other nor to any deeper comprehensive value. Instead, this kind of pluralism characterizes the normative universe by the plurality of guiding stars; each of which may independently contribute to a good human life, but none of which takes priority over all others by virtue of its integrating and overriding attributes. That norms and values are often made subservient to their contribution to some notion of 'the good life' in this approach illustrates that it is nothing new. The roots of this approach to pluralism can be traced back as far as the writings of Aristotle.[265] It is evident that value pluralism in general denies nor forbids there to be a problem of dirty hands. In fact, it was argued that value pluralism is a necessary assumption for normative conflict as such; at least for normative conflict as it is understood here.[266] But this kind pluralism in particular does more that just acknowledge – and at least in part explain – the possibility of normative conflict. A relatively recent contribution to the development of pluralist theory also attempts to adapt the theory of normativity to the possibility of normative conflict at a conceptual level. It does so by questioning two fundamental and commonly accepted presumptions in that theory: (a) that *ought* implies *can*, and (b) that norms must always be *action-guiding*.

It is something of a dogma in the theory of normativity that one is not obliged to do things that one cannot do. This seems obvious and is hardly disputed when *can* is understood in a causal sense. Even when one does not believe modal concepts like possibility (i.e. can) and necessity (i.e. cannot not) to be very meaningful, [267] one can always fall back on the commonly accepted position that any action or non-action that is caused or causally prevented from taking place does not really constitute an action and therefore *should not* be normatively evaluated. According to Stocker, however, this dogma does not apply for *practical* necessity, that is, to the impossibility to comply with two or more mutually exclusive normative claims at the same time.[268] Here it is not the action itself that is *determined* but it is the choice between different alternatives of action that is *constrained*. In such cases, Stocker argues, it is perfectly legitimate to say that an action is both right and wrong at the same time. It is right because from an *overall* perspective it is the proper thing to do. It is wrong because at the same time the *partial* claim that it is wrong does not dissolve with the overall judgement. In Stocker's view, therefore, the theory of normativity needs to be

---

[264] The view developed here is roughly the same as the one advanced by Susan Wolf (1992)
[265] Cf. Nussbaum (1986); Stocker (1990)
[266] See: Stocker (1990:241-277)
[267] See Quine (1963)
[268] Stocker (1990:11)

adjusted to the extent that it can accept what he calls "double-counted impossible oughts" as a normal feature of normativity.[269] To say the same in different words: normative theory must (a) learn to distinguish causal from practical necessity and (b) allow for double counting certain (partial) normative evaluations; allowing them to be counted once for the purpose of forming an overall judgement – in which it can be overridden – and once again in its own right – in which it cannot.[270]

This rather far-reaching adaptation of the theory of normativity invites an even more radical adaptation of that theory. Double-counted impossible oughts can only be incorporated into the theory of normativity if at the same time the requirement that normative evaluations be action-guiding is dropped. For it remains that normative evaluations can be action-guiding only once, even if one permits their double counting. If, for example, the moral community judges that my torturing a terrorist is practically necessary but nevertheless wrong, it is either the feature of necessity that is guiding me, or it being wrong despite it being necessary, but never both because then I would not be guided at all. According to Stocker, therefore, *non-action-guiding* normative evaluations are an inevitable and integral part of a pluralist theory of normativity. This is not only true for non-action-guiding *agent* evaluations, as these are already a relatively common feature in some virtue ethical approaches and are not considered conceptually very problematic.[271] It is also true for non-action-guiding *act* evaluations: i.e. normative evaluations of actions instead of agents. In the position that Stocker advances, therefore, norms *need not* be action guiding. Normative conflict in general and dirty hands in particular refer to circumstances under which at least *some* norms are not. Because the requirement that norms be action-guiding under all conditions is let go of, value conflict and dirty hands cease to be conceptual problems in the theory of normativity. What remains in Stocker's view, are just problematic contingencies that any self-respecting theory of normativity should be able to cope with.[272]

Although the effort to conceptually incorporate dirty hands in the theory of normativity is valuable for it's illumination of some of the most commonly held and widely accepted assumptions regarding normativity, it is defective nevertheless. It is defective, first, because it fails to acknowledge the inherently *social* nature of normativity. Although it is certainly necessary to analyse the condition of value conflict from a *single* agent perspective to determine its full depth – thereby making visible that normative conflict involves more than mere conflicting positions of different agents[273] – it is hardly *sufficient* to do so. For what a single-agent analysis necessarily leaves out is that norms – by their very nature[274] – address a *multiplicity* of *different* agents at the same time. They not only address the agent choosing what action

---

[269] Stocker (1990)

[270] Ibidem

[271] See: Nussbaum (1986) and Solomon (1992)

[272] Stocker (1990)

[273] The necessity of the single-agent approach to value conflict is argued by Williams (1981)

[274] Cf. Glastra van Loon (1956, 1958, 1980)

to take but also all agents that are somehow in a position to sanction that behaviour, albeit in a different way. That norms address a multiplicity of agents at the same time is true for all norms but it is most obvious for legal norms.[275] A criminal law, for example, not only addresses those who are in a position to commit a certain crime, but also the criminal justice system to prosecute that crime and society at large to support and empower the criminal justice system to do so. It remains therefore, that norms are *always* action-guiding, but not to all agents in the same way all of the time. Dispensing with the requirement that norms be action-guiding would be to throw away the child with the bathwater. Only if norms are action-guiding can they be effective, and only if norms are effective can they be binding. To take away the requirement that norms be action-guiding would therefore come down to dispensing with normativity altogether.

Stocker's effort to conceptually incorporate normative conflict in the theory of normativity is defective, secondly, because it does not account for the holistic feature of normativity and the demands that this makes on whole normative systems. It was argued earlier that norms can only have meaning as part of a whole normative system and never in isolation. This holistic feature of normativity, however, relates problematically to a pluralist approach that proceeds from the plurality of *individual* norms or values, for if these individual norms or values are not in some way systematically related to each other they will lack sufficient meaning. It would not be a solution to attempt to relate the plurality of norms to each other without destroying their plurality, by giving tolerance a central position in a normative system for example.[276] The reason why such an attempt would fail is that a normative system requires a certain minimal amount of internal *consistency* to be effective, that is, it must be able to represent, create or uphold mutual normative expectations in social interaction. It has already been explained that if a normative system is not effective it can be binding nor action-guiding. Because of this it is difficult to see how normative conflict in general – and dirty hands in particular – can be incorporated into a pluralist normative system without creating havoc to the internal consistency of that system. As was already touched upon in the discussion of the reflective equilibrium approach above, it seems extremely difficult to harmonize the norm that torture is wrong with the position that it may nevertheless be necessary, *within* the confines of a minimally consistent normative system. Pluralist normative systems, in short, must always meet the requirement of *minimal consistency*. If they do not, they will not be effective and therefore can not make a legitimate claim to obedience.

A second kind of pluralism is able to avoid problems relating to the requirement of minimal consistency. In this approach to pluralism it are not individual norms or values that cannot be harmonized by reducing them to a single principle, but rather whole normative systems for which this is true. Important groundwork for this approach to pluralism has been done by Isiah Berlin.[277] Tracing it back to the writings

---

[275] Cf. Kelsen (1945)

[276] Tolerance could also be realized at the level of institutions, however. See: Walzer (1997)

[277] Berlin (1953/1997, 1958, 1978, 1979, 1981, 1990, 1997)

of Machiavelli, Berlin has shown this second kind of pluralism to be an important contribution to modern Western thought in general and to liberalism in particular. What this approach to pluralism essentially comes down to is that it denies the possibility of combining all kinds of different norms or values within one viable normative system at will. By proceeding from the premise that the presence of some norms in a normative systems sets limits on the acceptance of other norms in that same system, it acknowledges both the holistic feature of normativity and the demand for minimal consistency that goes with it. Thus, to borrow an example from Machiavelli, the Roman pagan value system he so much prefers over that of Christianity cannot be fruitfully combined with the latter within one political entity. For Christian subjects will not live comfortable under pagan rulers who wholeheartedly accept that sometimes they must be bad in order to do good. In Machiavelli's view, the subjects must be pagans too, or else "they will accept to uncomplainingly the rule of mere bullies and scoundrels".[278]

The problem, however, with this second kind of value pluralism is that it takes the social and holistic features of normativity so seriously that it leaves little room for normative conflict in general and dirty hands in particular. The demand for consistency *within* the normative systems in such a pluralistic framework is so strong that serious conflicts between norms or values in that system are all too easily resolved in favour of one side of the conflict. Thus, again in the eyes of Machiavelli, it is not so much that the norms and values that guide the rulers may clash with those guiding the ruled. It is rather that the former are primary with respect to the latter, for without the former the latter could not exist at all.[279] Although this second form of pluralism does acknowledge the real possibility of value conflict *between* normative systems as wholes, it can only do so in relativistic terms.[280] The question whether this kind of holistic value conflict can be accepted at a theoretical level is thereby reduced to the question whether relativism is an acceptable position in the theory of normativity. I do not want to answer this question here, however. For now it seems justified to draw the preliminary conclusion that like the other positions in the theory of normativity discussed here, pluralism also faces serious problems in coping with the problem of dirty hands. Although it acknowledges the problem to be *real* in both the sense that it (a) does not dismiss it as a conceptual confusion and (b) can not be conceived of as a purely epistemological problem, it seriously underestimates the *practical* problems norms or systems of norms face in coping with the problem of dirty hands. As these practical problems relate to the effectiveness of norms, they are not a peripheral issue in the theory of normativity. It was argued earlier that the effectiveness of norms touches the very heart of normativity. The primary conclusion,

---

[278] Berlin (1953/1997:300)

[279] Freedom, for example, is generally taken by Machiavelli as being free from foreign rule; i.e. as freedom of the city state rather than individual freedom. On many occasions, Machiavelli clearly claims that the latter derives from the former. See in particular his *Discourses* (1516/1970) on this issue.

[280] Wolf (1992:785-798)

then, is that the problem of dirty hands is more than just a nasty contingency for any particular agent confronted with it. The evidence presented has shown that the problem dirty hands is also a serious conceptual anomaly as theories of normativity tend to cope with it either by (a) simply denying its existence, (b) failing to grasp the full depth of the problem, or (c) seriously underestimating the practical problems that dirty hands poses for norms or whole systems of norms. In general, therefore, the tension between normativity and dirty hands remains as dirty hands cannot be resolved or rendered harmless at a theoretical level. But perhaps we should not even try to. Perhaps we should approach the problem of dirty hands from a more pragmatic perspective: a perspective enabling us to learn to live with the problem of dirty hands instead of trying to solve it once and for all. Before adopting a more pragmatic approach in the next chapter, however, we first need to address the question how the problem of dirty hands relates to authority and intervention.

Intervention, Authority and Dirty Hands

In the previous chapter I have explained that norms must be effective – i.e. generally applied and on the whole complied with – in order to be action-guiding and binding. I have claimed that since not all norms we reflectively endorse at any point in time are spontaneously effective, intervention may be necessary to make these norms effective. I have also explained that the intervention necessary may vary from the exercise of the benign collective ability that men have when they act together, on the one hand, to the application of the coercive and violent capacities that some men have to make others do something against their will, on the other. The question now at stake is whether such intervention is normatively constrained. My answer, to be sure, is that ultimately it is *not*. My argument is that intervention may very well involve making dirty hands. Dirty hands – as conceived of in this chapter – thereby constitutes the *extreme case* of intervention, in that it is what intervention may *ultimately* come down to.

I have explained that the extent and invasiveness of the intervention necessary to realize a norm – i.e. to make that norm effective – depends on the structure of positive mutual expectations underlying that norm. The stronger these mutual expectations are infected, so to speak, with conflicts of interests (or values), the more intensive and invasive the intervention must be to secure that norm's effectiveness. The question is whether such intervention may go beyond or even against the norms applying at any point in time. It seems already quite clear that intervention is constrained at least by the norm for the realization of which it is necessary. But this need not be so. First, the norm to be realized may be, and typically is not effective yet, and hence does not constitute a *normative* constraint to intervention. Since the norm itself does not determine what would be necessary to realize it, these constraints are empirical rather than normative anyway. Second, it cannot be excluded that one must act against a norm in order to realize that very same norm. That this may indeed

be so is reflected by the saying that it takes a thief to catch a thief. The reality of contemporary drug policy demonstrates that this saying has a firm empirical rooting.

I have explained that the prohibition of drugs is based on a structure of mutual positive expectations that resembles a vicious variety of the prisoner's dilemma. The reason for this was that the criminalization of the use, possession and trade of narcotics have put a premium price on the drugs prohibited, which in turn motivates people to produce and trade drugs precisely because it is prohibited. I have emphasized that because of these counterproductive stimuli, upholding drug laws requires enduring intervention. There is also another problem, however, that makes upholding drug laws even something of a Sisyphus effort. For drug offences are victimless crimes in the sense that they typically produce no victim that is prepared to bring charges against the offenders. In fact, both the dealer and the drug user have a perfect alignment of interests where it concerns concealing their criminal interaction. In conjunction, these two features of drug laws make intervention on behalf of their realization both very intensive and highly invasive. It is a well-known fact that drug offences can only be prosecuted by giving law-enforcement officers a far-reaching mandate to discover evidence for prosecution. Not only must the privacy of those suspected of these crimes typically be violated by phone-taps and intensive observation, but often more invasive measures are necessary to secure conviction. These measures include infiltration by undercover law-enforcement officers, officers acting as pseudo-buyers and seducing criminals to testify against their criminal associates in return for reduced sentences or even immunity from prosecution. It was no surprise, therefore, that in the Netherlands the 'war against drugs' created by far the most serious crisis in law enforcement in modern history, with overzealous law-enforcement officers conducting illegal searches, running criminal infiltrants, intentionally letting tons of illegal drugs go through to the market, and systematically concealing these actions from public prosecutors, judges and the public at large.[281] If there is one thing clear at this stage, then, that there is a price to pay for upholding drug laws. An important part of this price is made up by the violation of other norms and values we are committed to.

This chapter made clear that such costs are everything but exceptional in modern life. The facts of the matter are that in modern society we are committed to a plurality of norms and values not all of which can be realized simultaneously. I have explained that when the latter conditions obtain we speak of normative or value conflict, and that this involves something deeper than a mere conceptual confusion or an epistemological problem that can – in principle – be solved by having more information or by refining our conceptual apparatus. While it is certainly true that normativity is *ontologically* constituted by *our* commitment to the world, this does not mean that any inconsistency in that commitment – i.e. value or normative conflict – can be solved with conceptual or epistemological means alone. I have claimed that there are real and existential problems involved here and that our common or garden understanding of the notion of *real* cost demonstrates that we perceive of these

---

[281] See on this: Fijnaut (1994); Van Traa (1996); Blom (1998)

problems in ontological rather than epistemological terms.[282] The example of drug laws demonstrates that intervention may very well tread into a context of normative conflict. Another example was given in the previous chapter. This example involved a norm making evolutionary theory a compulsory subject in high school biology lessons. The gist of this example was that the structure of mutual positive expectations underlying this norm can be characterized as a zero-sum game, that is, a game in which one man's gain is another man's loss. The reason for this is that having and enforcing this norm will not be possible without violating religious values according to which evolutionary theory constitutes outright blasphemy. These and many other examples demonstrate that intervention may ultimately be unconstrained by the norms there are at any point in time, and that it may therefore very well result in the intervenor making dirty hands. This is not only because upholding one norm may only be possible by breaking another, but it may even be necessary to (occasionally) break a norm in order to realize that very same norm. It is highly significant in this regard that a norm against violence is ultimately upheld by at least the threat of the same kind of violence the norm speaks against. Dirty hands, in sum, constitute the extreme case of intervention. The question is whether this extreme case tells us something *conceptually* significant about intervention and authority. There are, in my view, three lines of argument supporting the conclusion that it does.

First, there may be a conceptually significant relation between intervention and the problem of dirty hands for the general analytical reason that the extreme case of some phenomenon may very well highlight features *characteristic* of that phenomenon.[283] In physics, for example, we often make use of extreme or pure cases in theory development even if it is perfectly obvious that such cases will never actually obtain. We do so because we want to exclude presumably irrelevant factors from disturbing our general understanding of the phenomenon studied. The same could hold true for intervention and dirty hands. It can be argued that the distinctive features of intervention become visible only when intervention transcends the confines set by the norms that apply at any point in time. In essence this is Schmitt's argument for taking the exception rather than the rule as the foundation of normative legal order.[284] Although I have earlier dismissed Schmitt's decisionism as simply too radical, there seems to be more to his views than appears at first sight. It is clear, however, that this methodological argument cannot convince on its own account. It is compelling only in conjunction with arguments pertaining to the content of the relationship between the problem of dirty hands and its defining features, on the one hand, and intervention and authority, on the other.

A second line of argument focuses on the representative characteristic of the problem of dirty hands and its relation with intervention and authority. I have said that representation has become increasingly important in modern society even if only

---

[282] I elaborate on the distinction between ontological and epistemological subjectivity in chapter V in my discussion of institutions. See on this: Searle (1995)

[283] Cf. Davie (1980)

[284] Schmitt (1922/1985)

because of problems of scale.[285] Since representation *necessarily* involves authorization – i.e. endowing some agent with the right to act on one's behalf – there is already a clear conceptual link between the characteristic feature of dirty hands and authority. It is easy to see how this link may subsequently apply in the appropriate context – i.e. one of normative conflict – and thus constitute dirty hands problems in everyday representative relationships. The management of a firm, for example, may very well find the problem of dirty hands materializing while coping with the normative conflict of efficiency versus fair remuneration dividing the two central corporate constituencies, that is, shareholders and employees.[286] Likewise, a municipal government acting on behalf of all its citizens must inevitably make dirty hands in deciding whether or not to build a road through a forest in order to boost the ailing local economy. While all may benefit from the economic spin-off of this road, many may value this benefit less than the value they attribute to having an untouched forest. Representative relationships, in sum, may enter a condition of normative conflict when different constituencies endorse different norms that cannot be realized simultaneously. It is quite clear that such problems belong to the core of what representative relationships are about in modern society.[287]

A third line of argument focuses on how the condition of normative conflict relates to intervention and authority. The previous argument held that representative relationships may very well enter into a context of normative conflict and that such conflicts are common to what representative relationships are about in modern society. This line of argument, in contrast, turns the relationship between normative conflict and representation the other way around. It holds that representation and authority are a common and effective 'instruments' for coping with problems of normative conflict, and that authorities may very well be instituted precisely because of this. This is most evident, first, for adjudicative authority. It must be clear from my discussion in the previous chapter that even in purposefully designed legal normative systems there is room for applicational ambiguity and normative conflict. In order to cope with these problems adjudicative authority is indispensable in any sustainable normative legal order. It is also true, second, for legislative authority. It has been argued, for example, that it is a central purpose of the modern state to authoritatively contain religious or even moral disputes to the private sphere in order to safeguard the freedom of all.[288] In a similar vane it has been claimed that both the defining characteristic and the raison d'être of the modern state lie in its authoritative allocation of those values that cannot be either efficiently, effectively or fairly allocated by forms of spontaneous social order.[289] More or less the same logic can be seen to apply to the theory of the firm.[290] It can be argued that a central function of

---

[285] Dahl & Tufte (1973)
[286] Cf. Wempe (1998); Wempe & Kaptein (forthcoming)
[287] Cf. Pitkin (1967, 1968)
[288] Cf. Nagel (1987/1990)
[289] Easton (1965)
[290] A theory of the firm explains why there are firms within markets in the first place.

corporate authority is to impose unity upon the inevitably diverging normative orientations of the firm's constituencies, which prevent them from effectively coordinating and cooperating in a market context. Merely being "a nexus for contracting relations" may already require corporate authority.[291] It is no coincidence, finally, that in modern society the right to intervene with coercion and violence has been monopolized by the state. For the delegation of the exclusive right to apply violence in cases where it is necessary to secure the effectiveness of normative order to some central authority, *detaches* that violence to a certain extent from the sphere to which the norm against violence applies, such that the interventionist action itself does not undermine the effectiveness of the norm. The general conclusion, in sum, is that coping with problems of differential ultimate valuation is a central function of practical authority next to solving problems of coordination and combating problems of motivation.

Conclusions

The characterization of the problem of dirty hands in terms of normative conflict and representative action has made us see why intervention and authority are *ultimately* unconstrained by the norms that apply at any point in time and space. This is because the condition of normative conflict makes it possible that we can realize – i.e. make effective – one norm only at the cost of breaking another. The reality of contemporary drug policy is a case that proves the point. The same example also demonstrates that intervention on behalf of the effectiveness of a norm may sometimes even require breaking that very same norm. Intervention may therefore ultimately result in the intervenor making dirty hands. I have argued that the problem of dirty hands should consequently be conceived of as an *extreme* case of intervention, in that making dirty hands is what intervention may ultimately come down to.

I have subsequently claimed that this extreme case is more than some accidental and exceptional contingency. The reason for this is that it highlights some highly significant conceptual features of both intervention and authority. It shows, first, that representative relationships in general, and authority in particular, may very well enter a condition of normative conflict when different constituencies differentially endorse two or more mutually exclusive norms. I have argued that such conflicts are common to what everyday representative relationships are about in modern society. Second, it reveals that next to coping with coordination problems and combating problems of motivation, dealing with problems of differential *ultimate valuation* it is a central function of authority in modern society. Thus forms of *adjudicative* authority, first, are inevitable in any normative order because even the most carefully designed normative systems leave room for applicational ambiguity and normative conflict. Forms of *legislative* authority are central, second, because in any human association where there

---

[291] This is the central thesis of a paper still 'under construction'. The point of this paper is that the capacity to bind stakeholders indirectly to contractual obligations is already what authority consists of. The clearest on this is Green (1988), but see also MacMahon (1994, 1995)

are potentially many mutually exclusive guiding stars, some stars will have to be authoritatively endorsed and others rejected in order to escape from chaos and paralysing normative conflict. Forms of *executive* authority, finally, are highly functional in modern society because they enable us to detach making dirty hands in the realization of normative order to a certain degree from the domain to which the norms of that order apply, such that the interventionist action itself does not undermine the effectiveness of that order. I have claimed that the monopolization of violence by the modern state is no coincidence in this respect.

Because intervention and authority are ultimately unconstrained by the norms that apply at any point in time, there is no fixed demarcation line between power, on the one hand, and legitimate power on the other. Hence conceiving of authority in the classic Weberian sense as legitimate power is highly problematic. It is an interesting question how this insight reflects on modern ideals such as the Rechtsstaat, constitutional government, bureaucracy and organization. I will not address these issues in this study, however, as what is at stake here is our understanding of authority itself rather than any particular applications of it. With respect to the latter the overall conclusion must be that we are left empty-handed in our efforts to give a general, conceptual answer to the question of legitimacy that is indissolubly tied up with our understanding of authority. Not only did we have to conclude that there is no fixed conceptual boundary between our normative understanding of authority, on the one hand, and legitimate power on the other, but also the final frontier between legitimate power and power as such cannot be defended in conceptual terms. It follows that we should abandon our attempt to conceptually come to grips with the question of legitimacy and explore the possibility of more pragmatic solutions for what is now the *quest for* rather than the *question of* legitimacy. I leave such exploratory efforts to the following chapter.

IV

Responsibility for Authority I: Requirements and Complications

Introduction

The great American statesman and former president Woodrow Wilson once argued that "there is no danger in power, if only it be not irresponsible."[292] In the context of what this study is about this quote almost speaks for itself. It suggests that perhaps we should not even look for a *conceptual* solution to the quest for legitimacy that authoritarian governance confronts us with. The advice seems to be that instead of finding out if and how authority can be legitimate in any definite theoretical sense, we should rather learn to cope with authority by *making* it responsible. The promise implicit in this suggestion is that perhaps responsibility can relate to authority within hierarchy as demand relates to supply within a market, that is, that the practical problems pertaining to the legitimacy of authoritarian governance will be cleared once authority and responsibility are in some sort of equilibrium.[293] This suggestion is worthwhile exploring. Clearly, the notion of responsibility is one of mankind's most marvellous inventions. It is perhaps the central axis around which our social lives evolve in modern society. Modern moral community, legal systems, markets, organizations, but also friendships, marriages and many of our most common and taken for granted institutions would simply be inconceivable without some robust notion of responsibility being operative in them. But responsibility is arguably more than just a central and necessary part of our modern social machinery. It is also in an important sense constitutive of our modern self-image as reasonable and autonomous beings that perceive of their lives as ends in themselves rather than as mere means to some superhuman cause. Perhaps, then, if we could somehow make this highly significant notion do only half the work for authoritarian governance that it has already done for the emancipation of man, we can afford ourselves the luxury of becoming more relaxed about the remaining, mainly theoretical anxieties associated with authoritarian governance.

---

[292] Cited in Bovens (1998)
[293] Cf. Arrow (1974)

Embracing a more pragmatic approach in our quest for legitimacy does not mean that we can dispense with theory altogether. It just means that theory now focuses on something else; something that stands in a more pragmatic relation to what it was formerly concerned with. That something is the idea of responsibility for authority. It is clear that such an idea could be of central importance in coming to understand political, administrative, and executive responsibility. It could function as a kind of a meta-theoretical mould from which these more concrete forms may subsequently be conceptualised in greater detail. The problem, however, is that such an idea is not already available. Although much has been written on human responsibility in disciplines as diverse as philosophy, law and psychology, the nature of responsibility for authority is very much an undiscovered terrain. A major reason for this may be derived from the previous chapters. The nature of authority itself is subject to conceptual problems and ambiguities. If it were not for these problems there would be no reason to resort to an understanding of responsibility for authority in the first place. But although we are pretty much in the dark about the nature of responsibility for authority as such, many potentially constitutive elements of such an understanding are already available. The practice of responsibility, first, is ubiquitous in modern society and typically unproblematic. This practice is not just bound to the personal sphere, as we often speak of the responsibility of managers, politicians and even of whole administrations and apparently do so meaningfully. It is common sense to commence our effort to come to a conceptual understanding of responsibility for authority from such phenomena even if they are conceptually not understood very well. A second relatively obvious starting point for our analysis involves the abundant literature on the nature of human responsibility. Although this literature is generally exclusively about human responsibility and is often of a deeply metaphysical nature because of the close relationship between the concept of responsibility, on the one hand, and concepts such as freedom and determinism, on the other, some recent developments in this literature enable us to circumvent much of what may not be very relevant to understanding responsibility for authority. That this literature should contain something of value to us is promised by the assertion that the etymology of the notion of responsibility derives from the legal and political sphere.[294] Perhaps the deep metaphysical theorizing on the nature of human responsibility is even parasitic upon some prior unproblematic social understanding. Be this as it may, the journey ahead of us is clearly not an easy one. For not only do we have to show that there can be something like a conceptually consistent notion of responsibility for authority, that is, a notion consistent with and complementary to a conceptual understanding of individual responsibility, but we also have to make plausible that such a notion is capable of performing the pragmatic task we have in mind for it. That is, we will somehow have to demonstrate that responsibility for authority is indeed a potent and viable medicine that can help our modern hopes of reasonableness; autonomy and responsibility survive practical necessity of authoritarian governance in modern

---

[294] Pennock (1960:5)

society. Because this task is very complex indeed, let me sketchily point out how I will go about in completing it.

The first step in our journey will involve taking stock of the meanings of the concept of responsibility that are implied in some of the most salient practices and institutions of modern society. As these meanings diverge significantly, I will impose on them a conceptual dimension that will enable us to reduce them to three basic understandings of responsibility. In conjunction, these three basic understandings constitute a conceptual understanding of responsibility that can be arrived at through both instensional and extensional strategies of conceptualisation. The next step consists of confronting this conceptualisation with the additional demands that a notion of responsibility for authority requires. At face value, these demands are at odds with our conceptualisation of responsibility. Yet upon closer inspection they are no less problematic for understanding individual responsibility. The chapter therefore ends with an air of paradox. Although it is unlikely that the necessary requirements pertaining to a conceptual understanding of responsibility will actually be met, the practice of responsibility is ubiquitous in modern society nevertheless.

The Notion of Responsibility

Although the notion of responsibility is central to many of our modern social practices and institutions its precise meaning seems far from being clear. As we shall see below, it is not very difficult to come up with a number of quite different meanings that are clearly related but not in any straightforward way. One reason for the conceptual ambiguity surrounding the notion of responsibility is that it is central to a number of actually quite different practices and institutions that may or may not share many other features besides the central role of this concept itself. If we want to make sense of the notion of responsibility, then, we cannot get around exploring the meanings implied in these practices and institutions. This by no means implies that what responsibility is – or is not – can be determined by surveying the linguistic surface of these practices and institutions alone.[295] The reason is that responsibility is in no unimportant measure also what we what we *want* it to be regardless of how things are as a matter of fact. Because of the conjunction of this *normative* involvement inherent to the notion of responsibility, on the one hand, with the endemic disagreements we humans have on what this commitment pertains to, on the other, the notion of responsibility seems to have all the symptoms of what I have earlier referred to as an "essentially contested concept";[296] a concept the meaning of which is at stake in an ongoing debate that will arguably never come to an end because of the different perspectives and interests that are inevitably involved in this debate. That it is not to be expected that the precise meaning of responsibility will ever be agreed upon once and for all does not imply that anything goes where talk of responsibility is

---

[295] As theorists like Winch (1959) and Pitkin (1972) have claimed. See on this also: Green (1988)

[296] Gallie (1958), Connolly (1974)

concerned. Clearly, what can meaningfully be understood as responsibility is constrained both by existing practice and by arguments. Let me begin with surveying the most salient meanings implicit in these practices and allow me to turn to the arguments later on.

According to Bovens,[297] at least five different meanings of responsibility can be discerned at the linguistic surface of our most common and important practices and institutions in modern society. Although in conjunction these meanings constitute quite a large domain to which the notion of responsibility meaningfully applies, the following description of this domain does not claim to be exhaustive. There may very well be common and generally accepted alternative meanings of responsibility that I am unaware of and that fall outside the sphere of application located here. Nor are the different meanings of responsibility listed here mutually exclusive. Some of them overlap significantly while others clearly entail one another. In describing this domain, then, I am well aware of the possibility and even the likelihood of this description being incomplete and perhaps even to a certain extent incorrect. Yet given the explorative nature of our journey even a fallible description will be quite valuable. Let me begin, then, with pointing out and describing the five landmarks within this domain that were discovered by Bovens and to a significant extent already by others before him.[298]

1. A first familiar landmark within this domain involves the use of the notion of responsibility as a rough equivalent for the notion of a *cause*.[299] Such is the case for example when we say that fatigue was responsible for the poor performance of our national football team or when it is said that the global increase in the use of fossil fuels is responsible for the changing climate in North-Western Europe.

2. The notion of responsibility is often used, secondly, to denote or define a *task*, *function*, or *role* actors can or should fulfil; often but not necessarily in some sort of organizational context. The notion of responsibility is used in this way, for example, when the statutes of a management school say: "The dean is responsible for the general state of affairs in the school" or when a police captain tells one of his officers: "From now on you are responsible for making sure that no one crosses this line!"

3. A third meaning for the notion of responsibility distinguished by Bovens involves using it as a synonym for *accountability* or *liability*. To say that someone is responsible in this way means that we (will) hold that person accountable or liable for a certain state of affairs that we think is in some way related to his actions or lack of them. It quite normal, for example, to dismiss the CEO of a poorly performing company because we think he is accountable for what we think is an undesirable state of affairs.

---

[297] Bovens (1998:22 ff.)

[298] Bovens himself primarily turns to Hart as a source of inspiration. See: Hart (1968: 211-230)

[299] Bovens (1998:24-25); Hart & Honoré (1959/1987: 62-83), Hart (1968:214-215), Wolf (1985: 275 ff.; 1990: 40 ff.), Feinberg (1970b), Thompson (1980: 909-912)

4. A fourth meaning of the notion of responsibility refers to certain necessary qualities or attributes of actors that we commonly believe to precede the possibility and desirability of holding some one accountable and liable for what he has done or did not do. The way in which criminal law states the conditions under which someone is eligible for criminal punishment is paradigmatic for using the notion of responsibility in this way. We generally do not think of minors or mentally incapacitated persons as being fully responsible for their actions and commonly withhold punishment and retribution precisely because of this.

5. A fifth and final meaning of responsibility listed here is at stake when we say: "X is a very responsible person." What we mean to say here is that X is in some deep and special way *praiseworthy*; not only because of what he did or did not do – i.e. for behaving consistent with the normative expectations we have towards him – but also in a more enduring sense that tells us something about X as a person rather than merely about his actions. Using the notion of responsibility in this evaluative way makes responsibility a *virtue*; an attribute actors should desire for its own sake. Bovens refers to this particular meaning of responsibility as *active* responsibility because it has to be achieved or strived for by an actor him self, while he reserves the term *passive* responsibility for the kind of responsibility that is attributed to an actor by others; often, but not necessarily, only *after* the act or facts to be evaluated.[300]

Although these five meanings of responsibility differ quite substantially and thus constitute five distinctive landmarks that enable us to roughly locate the domain to which the notion of responsibility meaningfully applies, I think we must and can be more precise in describing this domain. Ideally, we would create some kind of semantic map for the notion of responsibility; a map that not only locates its domain of application but that would also help us to navigate within that domain. Such a map, however, would need at least two dimensions: two perspectives from which the landmarks described above can be systematically related to each other by providing each of them with a set of coordinates based on these two dimensions. The problem, however, is that no such two dimensions are readily available. At best there is only one dimension that seems to make sense in this respect. This dimension was already touched upon above. It concerns the level of *normative involvement* that underlies the different understandings of responsibility, that is, the extent to which some form of normative commitment is more or less significant in determining what precisely is at stake with each of the five 'species' of responsibility listed above. Although this dimension does not constitute the kind of map we are in need of here it is valuable nevertheless. For what it can do is to help us interpret these different species as differing in degree rather than kind. This would provide us with at least some sense of direction within the domain located. But next to giving us a sense of direction, this dimension of increasing normative involvement also enables us to reduce the five species listed above to three basic understandings of responsibility underlying the

---

[300] Bovens (1998:26 ff.)

different species.[301] Let me describe these three basic understandings and elaborate in particular on the degree in which they can be seen to differ. What I hope will result is an initial course for us to navigate on in our quest for a theoretically consistent and pragmatically viable understanding of responsibility for authority.

Responsibility: Three Basic Understandings

At the least demanding level of normative commitment the notion of responsibility is used for ascribing certain facts to a cause. Two examples of using responsibility in this way were already given above. Both examples depicted a causal relationship between two phenomena of which one is the antecedent and the other its consequent. When the notion of responsibility is used merely to denote a causal relationship without even an implicit reference to practical considerations – i.e. considerations pertaining to (human) action and agency – I hold that we cannot really speak of responsibility. In my view, referring to causal relationships in this way adds nothing that is not already contained in the vocabulary of causation while it does involve a risk of creating confusion as to what precisely is being said. In such cases it is arguably more appropriate to stick to the vocabulary of causality; the vocabulary as it is central to the discourse of natural science for example. But that a causal relationship between two or more phenomena does not constitute a sufficient reason to talk of responsibility does not imply that causality is not in some important way relevant to what responsibility is about. As we shall see later on, causal attribution constitutes a necessary condition for a *conceptual* understanding of responsibility.

If a causal relationship referred to is in some way related to either the possibility or desirability of *intentional* action, however, talk of responsibility becomes meaningful in its own right. It becomes meaningful because of the normative involvement implied in what seems to be a mere statement of fact. This involvement functions as a searchlight, so to speak, highlighting certain conditions from others as being relevant to our cognitive and evaluative practices in some special way. As is demonstrated in the two examples given above, this involvement often remains implicit and can pertain to both the possibility and the desirability of (human) action. The most basic form of involvement is already contained in ascribing certain facts to a specific kind of cause, that is, to an *intentional* actor.[302] For in causally ascribing certain facts to an

---

[301] The three basic meanings I distinguish in the following text are quite similar to the distinction made by Susan Wolf (1985: 275-286) between *superficial*, *practical* and *deep* responsibility. As we shall see later on, however, the three basic meanings of responsibility that I distinguish here involve more than what Wolf had in mind.

[302] The question what exactly constitutes intentionality is at stake in the philosophy of mind and action. Although I consider myself unqualified to say anything intelligent about this debate, my use of the notion of intention here roughly corresponds with what is known as the perspective of folk-psychology; i.e. the theory underlying the everyday ability of people to account for other people's behaviour by attributing content-full thoughts (or intentional states) to them. See: Dennet (1981); Searle(1983)

intentional actor we open up an inside or so called 'first-person's-perspective' on this relationship that makes it the object of *understanding* over and above merely being the object of *explanation*.[303] What understanding adds to explaining is a *projection* on this relationship of intentional considerations. More in particular, this projection involves the kind of intentional considerations we ourselves – as interpreters of the facts – are prone to. But such a projection does more than just constitute a first-person's-perspective on a causal relationship. As it draws this relationship into the space of practical reasons guiding (human) action it also makes this relationship susceptible to our evaluative practices; for to *understand* practical reasons presupposes some ability to evaluate them in a normative sense.

An example of a stronger form of normative involvement implicit in using the notion of responsibility to denote a causal relationship can be found in the testimonies of witnesses as these are submitted as evidence in a court of law. Although such testimonies generally have the ambition – and often succeed in – denoting causal relationships in a 'purely' descriptive manner, the pragmatics clearly demonstrate that a normative involvement is implicitly present in such statements of fact nevertheless. Not only are attributions of intentional states common and often quite essential to testimonies, but an even stronger normative involvement can be derived from the context in which such testimonies are made. For how could a witness testify of some facts and not of others if there were not some normative searchlight active from outside these facts making them relevant in some special way? And more generally: why would someone testify in court in the first place if this testimony was not in some normative sense relevant to human action and human association? These examples demonstrate that what is characteristic for the use of the notion of responsibility at this least demanding level of normative involvement is that this involvement is present, yet only implicitly. As a rule, it needs to be inferred from the *pragmatic* context in which the notion of responsibility is used. As I have argued above, however, this normative commitment is essential for using the notion of responsibility meaningfully nevertheless. Without normative involvement talk of responsibility is but a potentially confusing substitute for the language of causality.

Talk of responsibility becomes even more meaningful at a second, here intermediary level of normative involvement once the relationship between an actor, on the one hand, and his actions and its perceived consequences, on the other, are placed within an encompassing – sometimes purposefully created – social system of which the notion of responsibility is a central functioning part or *mechanism*.[304] There exist many social mechanisms in modern society that are relevant to the understanding of responsibility that is at stake here. A few were already touched upon

---

[303] For a very accessible and rather straightforward account of this distinction, see: Hollis (1994)

[304] I use the notion of a mechanism in a very general sense here while perhaps I should rather be more specific about what I mean with the term 'mechanism'. Space forbidding an elaboration on this issue, however, I refer to Jon Elster's work on this topic (1989,1999). In particular the latter work gives some very good examples of what I have in mind here.

above (sub 2) when discussing understanding responsibility as a *task*, *function*, or *role*. To understand the notion of responsibility in this way characteristically involves *explicit, action-guiding, normative* expectations towards an actor. These normative expectations denote or define what this actors' responsibilities are. Thus we define the responsibilities of a judge within a legal system, for example, by making explicit our normative expectation that judges are to make their judgements according to law and on the basis of a fair and impartial hearing of all relevant parties.

Merely defining responsibilities by making explicit the normative expectations we hold towards an actor, however, is not yet sufficient to constitute responsibility as a *functioning* social mechanism. What is additionally required is that we actually hold the actor to these expectations. In other words: we must make and actually hold an actor *accountable* for what he is expected to do or refrain from doing if the mechanism of responsibility is going to do any work at all. To illustrate this point, take the example of designing what we intend to become an efficient and effective law-firm. In designing this firm we cannot rest assured after having meticulously positioned each partner, all associates, the apprentices, and so on, in the complete web of explicit normative expectations of which this design consists. No organizational design, however brilliant, has ever been a sufficient condition for organizational success. What must additionally be done is to complement these normative expectations with some sort of sanctioning practice, that is, with *actually* holding the persons performing specific tasks or occupying particular functions within this design accountable or liable for what they are expected to do. Within the context of this study this is already a quite familiar kind of conclusion. It was argued quite extensively that only when intentions are connected to consequences that are, on the whole, consistent with them, can normative expectations be expected to function at all.[305] The *mechanism* of responsibility, then, comes out of conjoining responsibility as a task, function, or role (sub 2) with that of responsibility as accountability or liability (sub 3).

It needs to be noted that although normative expectations at this intermediate level of normative commitment are generally explicit and action-guiding, they may not and often do not constitute an *ultimate* kind of normative involvement. Even if we do not know what exactly would constitute an ultimate kind of normative involvement it is quite clear that the normative expectations that are characteristic for understanding responsibility as a social mechanism need not be just that. In fact, these expectations are commonly instrumental rather than ultimate because they are characteristically thought of as committing or binding for an actor only if they are socially efficacious; often towards some overarching goal or purpose. Let me illustrate this point with the aid of an example.

The example involves the notions of *strict* or *vicarious* liability in civil law.[306] In many legal systems it is quite common that actors are held liable for things they cannot reasonably have done or prevented from happening. Anyone who owns a dog knows that he cannot completely control this animal's behaviour. Yet at the same

---

[305] See chapter II of this study.
[306] See: Hart & Honoré (1959/1987: 62-68); Hart (1968: 215 ff.); Feinberg (1970b)

time it is sensible legal practice to make the dog-owner liable for damage caused by this dog. The reason behind this is that it is arguably the owner of the dog that is most efficacious in preventing harm from happening and that making him liable thereby assures the most extensive control over the dog. Roughly the same logic holds true for the Dutch practice of making the drivers of cars liable for at least fifty percent of the damage resulting from an accident involving a car and a cyclist, regardless who was actually at fault.[307] Again, it is arguably the driver of the car who is most efficacious in preventing harm caused by accidents because a car can cause much more damage to a cyclist than vice versa. Would we eventually have to conclude from empirical evidence that our assumption were false – because this measure stimulated a certain degree of recklessness on behalf of cyclists for example[308] – we would not hesitate to change this specific form of liability in a way that we see fit.

A third broad understanding of responsibility that I wish to distinguish here involves the most extensive form of normative involvement. In contrast with understanding responsibility as a social mechanism in which the normative involvement is explicit yet often only instrumental, this third understanding of responsibility does involve a normative involvement that is both explicit and ultimate in the sense that the former was not. It involves an ultimate normative involvement because it pertains to both the question whether and how our practices of holding actors responsible can be ultimately justified, and to the inherent desirability of being a responsible actor. As such, it captures both the necessary qualities and attributes actors must possess in order to justifiably be held responsible (sub 4) and responsibility as a virtue (sub 5).

It was already hinted at that we generally do not consider minors and mentally incapacitated persons to really be responsible for their acts or omissions. The reason for this is that we believe that both categories of actors lack certain qualities that we think are necessary conditions for being (held) responsible in the first place. These qualities involve both a kind of basic freedom and a minimal capacity to reason. If either one of these conditions is lacking, we run into serious trouble with our attributions of responsibility. For can we *really* say that someone is responsible for committing a crime, for example, if we know the offender's physical and psychological make-up are predisposed and conditioned in such a way that he could not have done anything other than what he actually did? And can we justify attribution of responsibility and administering punishment to someone who does not understand what the law requires of him, not just in any particular case but in a general sense? It is clear enough that both cases raise serious doubts whether attribution of responsibility and administering punishment can be justified even if it is

---

[307] A law to this effect was actually proposed in 1997 by the Dutch minister of Justice, and again, albeit modified, by her successor in 1999. It never formally became law, however. In the absence of a formal law to that effect, roughly the corresponding policy has evolved in jurisprudence. See: NRC Handelsblad, 26 april 2001, p. 33

[308] Which was one of the objections against the proposal mentioned earlier

completely clear that: (a) the actor did break the law and it is known further that (b) punishment does contribute to the prevention of precisely the kind of crime the actor has committed. What is at stake here, in other words, is not what Susan Wolf calls "superficial responsibility",[309] for this can be established or imposed regardless of the necessary conditions specified above being satisfied.[310] Rather, what is at stake here is what Wolf calls 'deep responsibility'; the kind of responsibility that involves the question whether or not the events caused by an actor *are really up to him*.[311]

The question what exactly constitutes deep responsibility is a matter of great debate, primarily in philosophy but evidently also in other disciplines such as legal theory and psychology. Although the jury is still out on this debate, it seems clear at least that this question is ultimately a normative one.[312] Without an actor being *able* to understand what it is that may be expected of him and without this actor being *able* to act accordingly, it is difficult, if not impossible, to *justify* holding this actor responsible for what he does or does not do nevertheless. That the question what constitutes deep responsibility is ultimately a normative question does not imply that deep responsibility is inherently a *moral* notion. Again, what is at stake with deep responsibility is whether or not the facts we ascribe to an actor are really up to him; that the deeply evaluative attitude we hold and perhaps express towards this actor does not merely address what the actor did or did not do, but also what we believe this actor could have done given the kind of actor that he is and having the kind of abilities that he has.[313] To be sure, such an attribution of deep responsibility can apply both to moral and non-moral evaluations. The latter is the case, for example, when we evaluate the performance of a football player in relation what to he could have done given the abilities he is known to have. Only if we know this player could have done better or worse than he actually did, does he deserve *deep* praise or blame respectively.

---

[309] Wolf (1990:40 ff.)

[310] It is questionable, of course, whether holding someone responsible that does not posses the necessary attributes for responsibility will keep this same person from offending again. So there is what looks like an efficacy-argument of in favour of these necessary conditions for responsibility here too. Cf. Hart (1968: 229). It is clear, however, that this efficacy-argument cannot be extended to the aggregate level. At this level punishment also has a symbolic value and is therefore also directed at others than the offender himself. In that case holding this offender responsible despite his lack of relevant capabilities does not necessitate these capabilities in order to be efficacious.

[311] Wolf (1990: 40 ff.)

[312] Cf. Strawson (1963); Wolf (1990); Wallace (1994). An important advantage of this point of view is that it allows one to disregard much of the metaphysical theorizing on the nature of human freedom and responsibility that has dominated this debate for so long. Not only do I believe, with Strawson, that much of this theorizing hard to reconcile with our everyday understanding of human freedom and responsibility, but I also think it to be largely irrelevant to an understanding of responsibility for authority.

[313] Wolf (1990: 41)

But responsibility in this deep sense does not just constitute an ultimately normative understanding of responsibility because of the requirement that attributions of it be justifiable in an ultimate sense. It also does because it denotes what we believe to be a *central human virtue*: something we should all want to be or become. To talk of responsibility as a human virtue arguably entails more than just referring to the necessary conditions for deep responsibility. For isn't it true that we would not consider a man responsible if he behaved badly, not just occasionally, but as a rule? And don't we actually mean that someone is in a deep sense praiseworthy for what he *does* and who he *is* when we say that he is a responsible person? These examples suggest that deep responsibility is a normative concept in a much more involving manner. More in particular they suggest that our attributions of deep responsibility are intimately connected to *actual* blameless or even praiseworthy behaviour and that it is problematic to connect deep responsibility with blameworthy actions. Susan Wolf comes to this conclusion when she holds that we really cannot be certain that a blameworthy action is just that or rather a symptom of a much more serious condition: that of not having what it takes to be responsible.[314] If we want to connect deep responsibility with blameworthy action, then, we must somehow *presume* an actor to be responsible absent evidence to the contrary, and we must do so for the ultimately normative reason that our commitment to the human virtue of responsibility is so strong that to presume otherwise is in itself normatively undesirable.[315] But although our commitment to this presumption is normative it is not invulnerable to the facts. We can easily imagine it being undermined by empirical conditions such as recurrent deviant behaviour for example. Arguably, then, the connection between the necessary conditions for deep responsibility and the virtue of responsibility is an intimate one, even if we do not precisely know how this is so. I will come back to this issue later. For now it is safe to conclude that at the far end of our underlying dimension of normative involvement lies an ultimately normative understanding of responsibility; an understanding that commits us not only to ultimately justify our applications of the mechanism of responsibility, but also to responsibility as a virtue.

Let me pause and look back on the path that has taken us here. I have started out exploring various meanings of responsibility implied in some of our most salient practices and institutions in modern society. This exploration resulted in five different species of responsibility that constituted five distinctive landmarks for us to navigate on in our quest for a notion of responsibility for authority. But since these landmarks were inconclusive in deciding which direction further to take, I have imposed on them a dimension according to which these species can be interpreted to differ in the extent to which they entail a certain normative involvement. Taking this perspective enabled us to sketch three basic understandings of responsibility that each involve a

---

[314] See Wolf's asymmetrical view of freedom and responsibility (1980, 1990:79-81)

[315] Cf. Strawson (1963) but also the more specific claim made by Sie (1999) that acknowledging the possibility of an act being the expression of a normative conflict (normatively) requires the presumption of the actor being responsible.

different degree of normative involvement. This involved, first, what from now on I will call a *cognitive* understanding of responsibility.[316] This understanding assumes an implicit normative commitment that enables us to understand certain causal relations as in a special way significant to our normative evaluative practices. It was argued that this minimally involved understanding these causal relations as the (possible) result of *intentional* action. A second basic understanding concerned what I see as a *regulative* understanding of responsibility. This involved responsibility as a social *mechanism*. This understanding involved a conjunction of there being (a) explicit action-guiding normative expectations and (b) some sort of sanctioning practice to uphold these expectations. It was argued that the normative expectations characteristic for this regulative understanding of responsibility do not necessarily involve an ultimate kind of normative involvement. Instead, this involvement is often only instrumental because the normative expectations involved are subject to their being socially efficacious. At the end of our dimension of normative commitment, finally, lay what I consider an inherently *normative* understanding of responsibility. This understanding involved both an ultimate justification of our practices of holding actors responsible and the virtue of responsibility. It was shown that the former and the latter are intimately related to each other.

It is central to note that if one wants to apply the notion of responsibility throughout the domain demarcated by the *de facto* use of this notion – and it must be said that this (i.e. extensional adequacy) constitutes something of a standard requirement for *concepts* in general – then it must satisfy the basic requirements posed by *each* of the three understandings of responsibility discussed above. To put it in other words: the *concept* of responsibility can at least be partially defined in terms of these basic requirements, which are then necessary – but not sufficient – conditions for this concept. Thus our cognitive understanding of responsibility, to start with, requires that one must be able understand certain facts as the (possible) result of the acts or omissions of an intentional actor, thereby making such *ascription* a necessary condition for correct use of the concept of responsibility. And our regulative understanding of responsibility, secondly, necessitates there being both an unambiguous normative background consisting of explicit action guiding normative expectations and an accountability practice that can uphold these expectations. Our normative understanding of responsibility, finally, demands that the actor to which responsibility is attributed satisfies certain necessary conditions in order to justifiably be held responsible.

It is not just for reasons involving *extensional* correctness of application, however, that these necessary conditions apply. They can also be arrived at following an *intensional* strategy. Such a strategy would pertain to asking the meta-ethical question what conditions must be fulfilled in order to *reasonably* hold responsible an actor (A) for certain facts (F), given the way we intensionally understand the concept of responsibility. It is quite clear that an adequate answer to this question would

---

[316] The particular names I use to denote these three basic understandings of responsibility anticipate the theoretical perspective on responsibility that I will propose later.

resemble if not be equivalent to the answer we have arrived at following an extensional strategy. It would not reflect our understanding of the meaning of responsibility to hold A responsible for F if (a) F could not be ascribed to A as an intentional actor, (b) if there were no unambiguous normative background on the basis of which F can normatively be evaluated, and (c) if A does not have what it takes to be a responsible actor nor values to be so. Moreover, it is quite clear that A would not be *actually* held responsible for F if there were no one to *hold* him responsible. Given that these requirements for the *concept* of responsibility are robust over both extensional and intensional strategies of conceptualisation, then, the next question is if and how a notion of responsibility for authority can satisfy these necessary conditions, that is, whether there can be a *concept* of responsibility for authority and, if so, what such a concept would look like.

The *Concept* of Responsibility: Requirements and Complications

Although we now have a rough idea of the domain to which the notion of responsibility meaningfully applies and have gained some sense of direction within this domain, the question remains where to go exactly in our quest for an understanding of responsibility for authority. What troubles us here is that the provisional 'semantic map' of the notion of responsibility we have created thus far is but a one-dimensional one. It is therefore incapable of locating anything but points on this dimension itself. Ideally, we would have an additional dimension on this map; one constituted by an unambiguous understanding of authority. Understanding responsibility for authority would then be a matter of studying the intersection of these two dimensions. The troubling fact of the matter, however, is that understanding authority is precisely what is at stake here and that what we have discovered thus far is clearly insufficient to conceive of anything close to what we are in need of here. That we cannot follow this ideal strategy, however, does not threaten our expedition as a whole. For what we can still do even with a problematic and ambiguous understanding of authority is to look whether what we do know about authority and intervention is compatible with *any* understanding of responsibility. Arguably, this is what was required anyway. It was argued that a *conceptual* understanding of responsibility for authority is possible only if it can fulfil the basic requirements that hold for *each* of the three understandings of responsibility discussed in the previous paragraph.

The problem of ascription

It was explained that a cognitive understanding of responsibility minimally requires the ascription of facts to an intentional actor, that is, one must be capable of *understanding* these facts as the (possible) result of the acts or omissions of an intentional actor. The problem, however, is that such ascription is highly problematic in case the actor is an authority. The most important reason for this is that authority coincides with power. We have already seen that for authority to make sense at all it

must at least be able, that is, have the power to do what it is meant and perhaps even justified to do.[317] But, as explained, power is a very complex notion and that to operationalize, identify, let alone measure it in some way relevant to the requirement of ascription is everything but simple and straightforward.[318] Things become even more complicated once we need to connect power to intentions in the context of ascribing a certain state of affairs to an authority.[319] These issues constitute what I will call *the problem of ascription*.

An example of what is at stake here involves a common feature of everyday political reality. Assuming there are good reasons to devise a constitution and confer to an administration the authority and power to decide, act and take care of certain things on behalf of us, the question subsequently becomes relevant whether it does so to our satisfaction and whether the administration is responsible (in a cognitive sense) for what happens to be the case. Although it is arguably a poor example given the lack of governmental authority over economic affairs in modern liberal democracy, the responsibility of an administration for the state of the economy is probably the most intensely debated issue modern democracy that shows precisely how problematic *understanding* the role of government is in such contexts. If we have learned anything about social reality – and in particular about the *aggregate* level of social reality at which authoritative agents often operate in modern society – it must be that the phenomena in need of understanding generally have a very large number of different causes and an even greater amount of consequences. To identify the net contribution of an administration in such a context is typically highly problematic. One must keep in mind that we do not have a *natural* birds-eye point of view from which we can untangle the complex causal relations typical for an aggregate social level. Generally, such a point of view needs to be purposefully constructed. This commonly involves a collective and highly organized effort.[320]

Things get even more complicated once we realize that what authorities do and are meant to do hardly ever involves the work of a single man. Instead, authoritative agency typically involves some (a) collective and (b) organized agency thereby making it even more difficult to ascribe a given state of affairs to an authority. For both the collectivisation and organization that are – at least in an empirical sense – constitutive of authoritative agency, obfuscate, confuse, and to a non-trivial extent disassociate power and intentionality in authoritative agency.[321] Can we really say, for example, that the abuse of power over civilians by an official is attributable in any causally relevant sense to the organization that employs him or to those in charge of that organization? And how do this man's intentions relate to those of the

organization or its management? Can we really understand them as being systematically related? And, perhaps more importantly, how do organizational intentions relate to organizational power and capabilities in general? Are they as closely adapted as intentions and capabilities typically are in (normal) human actors? These questions demonstrate that it is often highly problematic to *understand* whatever causal relationships we can unravel as the result of *intentional* action. For although we have a relatively unproblematic everyday understanding of what it is to be an intentional *human* actor, this is typically not the case with collective and organizational actors. Now I do not mean to say that collective and organizational agents cannot have intentions. On the contrary, I think there are very good reasons to hold that they can.[322] But I do say that understanding this kind of intentional agency is often problematic, particularly in the context of ascription. Even where intentions have been made explicit by authoritative agents do we need to be very cautious because they are often coloured by the competitive context in which authorities are typically appointed in modern society. All too often the results are unrealistic expectations that can hardly be considered intentions at all.[323]

Although these questions already represent serious problems, the problem of ascription becomes even more acute once we consider non-action; i.e. what authorities did not actually do but allegedly could and arguably should have done. For this necessitates counterfactual speculation on both the power and intentions of authoritative agents. With respect to the former it was already said that power is a notoriously difficult concept to apply even in the actual world and we can only imagine the kind of problems we will encounter once counterfactual speculation on power as a (mere) capability is required. At least equally problematic are intentional considerations in such cases. As explained, it is already difficult to connect complex causal relations with intentional considerations but to connect the latter with non-events is clearly more difficult than that. Let me summarize what I think the problem of ascription essentially comes down to. The conditions that are at least in an empirical sense constitutive of authoritative agency seriously frustrate understanding a given state of affairs as the (possible) outcome of intentional authoritative agency. In practice, the requirements for such ascription – i.e. for applying our cognitive understanding of responsibility – are such that they can be met only in exceptional situations. It is clear that this hardly constitutes a sound basis for a *conceptual* understanding of responsibility for authority.

It is important to note that the problem of ascription is not just a problem for authoritative agency. It also applies to individual human actors. The condition that threatens the application of our cognitive understanding of responsibility to human actors, however, is metaphysical rather than practical. It involves the thesis of determinism. In laymen's terms this thesis holds that any given state of affairs is the necessary result of given initial conditions and the laws of nature that apply to these

---

[322] Cf. Gilbert (1989), Tuomela (1995), Pettit (1993/1996, 2001), Bratman (1993, 1999)
[323] Knight (1960:187)

conditions.[324] This thesis is closely connected with the perspective of the natural sciences. The problem, however, is that it is not just about the things we commonly associate with natural science. It is also about human action and everything that it involves. If the thesis is true, even human thought and action are the result of a given state of initial conditions and the natural laws that apply to them. It is intuitively not difficult to see how the truth of determinism threatens even our normatively least involving understanding of responsibility.[325] For if everything an actor intends and does is fully determined, why should we ascribe any facts to *that* actor in particular? What reasons can we have for stopping short of the causal forces determining this actor and his actions in our explanatory efforts? Why would intentionality be a relevant condition for our practices of ascription in the first place? Wouldn't determinism make human thought and intentions mere epiphenomena? And wouldn't all understanding ultimately reduce to explanation under such conditions? What could be the reason for applying our cognitive understanding of responsibility nevertheless? Even if the thesis of determinism turns out to be false this would not solve our problem, for if our thoughts, intentions and acts are not determined by anything at all, why should they be any actor's thoughts, intentions or acts in particular? Doesn't the arbitrariness that the falsity of determinism implies frustrate ascription as well as the whole point of the exercise for roughly similar reasons that determinism does?[326] These are all difficult issues and I do not wish to explore them any further here. My point is merely to demonstrate that the application of our cognitive understanding of responsibility is problematic for both authoritative *and* human agency; that is unless some other basic understanding of responsibility is involved, but that is a point I want to reserve for later. Let me first answer the question whether the requirements for our *regulative* understanding of responsibility can be met by authoritative agency.

The problem of practical necessity

It was argued that this second basic understanding of responsibility involves responsibility as a social mechanism. This mechanism comes into existence by the conjunction of there being: (a) explicit action guiding normative expectations and (b) an accountability practice that involves some sort of sanctioning supportive of these expectations. Let me focus on the former first and say a few words on the latter later. Earlier in this study it was argued that one important reason for authoritarian governance to persist despite the process of modernization involves the condition of value conflict and the normative conflicts associated with this condition.[327] It was argued that the requirement that norms be effective in order to be binding necessitates the deadlock in certain normative conflicts to be broken in order to save the effectiveness (and ergo bindingness) of at least one of the normative claims in

---

[324] For an authoritative source on the thesis of determinism see: Honderich (1988).

[325] The position that freedom and responsibility are incompatible with determinism is appropriately called 'incompatibilism'. Cf. Van Inwagen (1983)

[326] Cf. Wolf (1990); Young (1991)

[327] See chapter III

conflict. It was also argued that authority – i.e. the constitution of a representative *agent* that decides for and on behalf of us – is an important way of coping with this kind of normative conflict. It was explained that authoritative representation was historically additionally stimulated by the development of the face-to-face society of the Greek polis into the large anonymous communities typical for modern society. These developments made representation necessary if only for reasons of scale.[328] The problem now is whether this central function of authority in modern society is *practically* compatible with the requirement of there being action guiding normative expectations constituting a functioning mechanism of responsibility. I believe this to be highly problematic.

I have already argued that authority is in a systematic way – short of necessity – related to what has become known as the problem of dirty hands; i.e. the problem that under conditions of unsustainable normative conflict authorities must sometimes act bad in order to do good.[329] In such conditions authorities must act against at least one of the action-guiding normative expectations in conflict in order to save the effectiveness and therefore bindingness of another. But if we want authorities to do that – if we even believe that it is part of authority's raison d'être to do so when the circumstances require it – then how can the mechanism of responsibility be fruitfully applied here given that this mechanism relies on there being explicit action guiding normative expectations? For isn't the problem precisely that whatever is normatively expected is not action guiding for the authority?[330] And if this is true, against what normative background should we hold authorities accountable then? A straightforward mandate or delegate notion of representation will hold little ground here,[331] given that often the whole point of instituting an authority is to cope with discontinuities between what individual actors *value* as their interests and what can be *realized* at the aggregate level. Moreover, the mere expectation that authorities break the deadlock under circumstances of unsustainable normative conflict is simply too thin to normatively constrain authorities in any significant way, while the possibility of creating some kind of *second order* normative expectations at the level of the authority is severely constrained by the requirement that these may not undermine the (effectiveness) of what first-order normative expectations there are. It is good to keep in mind here what Berlin thought to be Machiavelli's most original advise: that to mix two normative systems – one for the rulers and one for the ruled – inevitably results in the worst.[332]

Again, this problem – which from now on I will refer to as *the problem of practical necessity*[333] – is not unique for authoritative agency. We have already seen while

---

[328] Cf. Laslet (1963); Dahl & Tufte (1973)

[329] See chapter III

[330] See my discussion and criticism of applying the notion of 'non-action guiding act evaluations' in chapter III. Cf. Stocker (1990)

[331] Cf. Pitkin (1967), Pennock (1968)

[332] See Berlin (1953/1997: 300 ff.) but also chapter III

[333] Cf. Stocker (1990)

characterizing the problem of dirty hands *vis á vis* notions such as 'tragic choice' and 'moral dilemma' that normative inconsistency and 'moral luck' do not just apply to authoritative agency but are rather pervasive facts of the human condition as such.[334] Not only does this problem therefore also apply to individual human actors, but it has become both more pervasive and more acute in modern society. One reason for this is that modern man typically fulfils many different roles at the same time and therefore has equally many different 'responsibilities' that do not always nicely cohere or coincide. Next to the problem of 'dirty hands' and that of "many hands",[335] this problem has been named the problem of "e*ntangled* hands",[336] because the entanglement of the normative expectations that go with the different roles modern man typically 'plays' simultaneously can seriously frustrate the functioning of the mechanisms of responsibility associated with each of those roles. Now admittedly, this problem may be less acute for individual human actors because "individual responsibility attaches to persons" and not to roles and can therefore be followed "wherever he or she goes".[337] But that individual responsibility attaches to persons doesn't make this problem go away. At best, it mitigates it. At the end of the day even natural persons can be torn apart by conflicting normative expectations as classic Greek literature and philosophy have already shown so well.[338] The problem of practical necessity thus applies both to individual human actors and authoritative agency, albeit in different degrees and in different ways.

The problem of constituency
With respect to the accountability and sanctioning that constitute the second conjunct of the mechanism of responsibility, it must be noted that this assumes there to be some party to which one is accountable and that is capable of the sanctioning required to uphold the action guiding normative expectations that constitute the first conjunct of this mechanism. This is assumption is already implicit in the etymology of the notion of responsibility which comes from the Latin verb *respondere* and which means 'to answer'.[339] For to answer in any sense of the word assumes there to be someone to answer to.[340] The problem, now, is who precisely this someone is. The relevant difference here is not between individual responsibility and responsibility for authority but between moral responsibility and the responsibility for authority, as each seems to assume a different party to which it is liable to answer. This is what

---

[334] See chapter III but also Berlin (1953/1997, 1958, 1978, 1979, 1990, 1997); Nagel (1979); Williams (1973, 1978, 1981); Stocker (1990) and Nussbaum (1986)
[335] Thompson (1980)
[336] Kaptein (1998), Kaptein & Wempe (forthcoming).
[337] Thompson (1980:914)
[338] Nusbaum (1986); Stocker (1990)
[339] Bovens (1998:23)
[340] Cf Anscombe (1958), Roland Pennock (1960:5 ff.), Van Gunsteren (1974), French (1979/1991:140).

Gutman and Thompson have called "the problem of constituency".[341] The constituency of moral responsibility is formed by the moral community from which no one may reasonably be excluded, while for responsibility for authority there always seems to be some kind of privileged constituency consisting of those on behalf of which the authority is instituted: the so called *principals*.

Although the problem of constituency does indeed involve some practical discontinuities between moral responsibility and responsibility for authority it equally does so for different forms of individual responsibility. Some moral norms are considered of such importance that their sanctioning cannot be left to the spontaneous reactive attitudes and actions that are thought of by some as constitutive of morality.[342] These norms therefore tend to become codified in criminal law and their sanctioning is left to some organized agency that acts on behalf of all of us. All modern legal systems know the function of a public prosecutor. But although a public prosecutor is an authoritative agent that acts on behalf of all of us, it is also a *principal* with respect to those he holds criminally responsible.[343] A public prosecutor is therefore a *constituency* for legal subjects. Moreover, this particular constituency has an *exclusionary* nature. It does not allow other constituencies to also hold criminal offenders accountable. Now of course this does not mean that we are not allowed to moralize on those cases public prosecutors have an exclusive right to prosecute. It does mean that other constituencies are excluded from the means that can make our sanctioning of criminal behaviour effective in the way that the proper functioning of the mechanism of responsibility requires. We typically do not see this as a problem in modern society, though. In fact we commonly praise the exclusionary nature of criminal justice as one of the most valuable achievements of modern civilization. Thus rather than creating some kind of conceptual cleavage between responsibility for authority and individual responsibility, then, the problem of constituency merely seems to reflect a practical trade-off between the scope of accountability and its effectiveness in practice: a trade-off that is equally relevant to both individual responsibility and responsibility for authority.

The problem of responsiveness

The requirements that follow from a normative understanding of responsibility, finally, are arguably the most difficult to satisfy for authoritative agency. They involve both the necessary conditions for justified attribution of responsibility and responsibility as a virtue. Since we have already seen that the two are intimately related we can focus here on the former thereby automatically drawing in aspects of the latter. Earlier it was explained that in order for an actor to justifiably be held responsible this actor must at least have a capacity to reason and some basic freedom, that is, he must have the capacity to know, understand and reflect on what is expected

---

[341] Gutman & Thompson (1996:128), but see also French (1979/1991:140)

[342] Notably by so-called dispositionalists, see Van Willigenburg (2000)

[343] Cf. Knight (1960:172)

of him and to subsequently act accordingly.[344] It is already quite clear that these requirements are threatened by some of the problems discussed in this paragraph; in particular by problem of practical necessity. Since I have already discussed the latter issue at some length I will focus here on two other, though clearly related issues. The first of these issues involves what I will call the problem of responsiveness. This issue pertains to the question whether an authoritative agent has what it takes to be appropriately responsive to the interests of its principals. The second issue pertains to the capability required to subsequently act accordingly. This problem involves attributing responsibility to the structural and organizational features that are at least in an empirical sense constitutive of authoritative agency.

It was already explained that the problem of practical necessity frustrates the functioning of the mechanism of responsibility because what is often lacking with authoritarian governance is an unambiguous normative background against which authorities can be held responsible. But even if it is not clear by what normative light an authority can be held responsible it remains that it must minimally be *responsive* to the interests of its principals. That is, it must see and understand what these interests are and it must be capable of forming its intentions in the light of these interests. Although this requirement seems rather straightforward, there is no simple way to ascertain whether it is adequately satisfied or not. The main reason for this is that the context of practical necessity typical for authoritative agency complicates the relation between the interests of the principals and what the authority does or does not do.[345] Thus it may very well be the case, for example, that an authority is responsive to its principals while at the same time it acts against their self-proclaimed interests nevertheless. An example of this is when – contrary to the wishes of the shareholders – the management of a company decides not to pay out dividend in a given year because of a big investment it wants to make and which in their opinion is central to the survival of the company. That management does not do what its principals desire in this case says nothing about its capability to understand and attend to their interests. For all we know it did consider these interests carefully and appropriately while at the same time deciding against them on *their* balance of reasons.[346] It is of course also possible that management somehow lost track of the interests of the shareholder – because of a preoccupation with growth for example – and that it has become unresponsive to the interests of its principals. The point is that it is not easy

---

[344] Cf. Wolf (1990:92): "the ability to know the know the True and the Good, that is, to form ones values and plans in the light of them, and the ability to convert ones values and plans into action. If one lacked the first component, the choices one made would be blind. If one lacked the second, one's choices would be ineffective".

[345] Which, it must be noted, is true for any representative relationship, but that is besides the point here. See Pitkin (1967, 1968)

[346] That what counts as reasons and what constitutes the balance of reasons at the collective level may differ from what are considered reasons and what constitutes the balance of reasons at the individual level – even if the collective level rests exclusively and exhaustively (i.e. supervenes) on the individual level – is argued and explained later in this paragraph. See: Pettit (1993; 2001)

to see which one of the two scenario's obtains but that to accept authoritative agency requires the *presumption* of responsiveness nevertheless. Without this presumption there would be no reason ever to accept any kind of authoritarian governance. For why would one do so if there is no reason to think that it even *can* appropriately attend to ones interests? The question, then, is not whether the presumption of responsiveness must be made, but rather what would make this presumption either plausible or justified.

There are no easy answers to this question. Unlike our inherently normative commitment to the presumption of individual freedom and responsibility there is no such commitment to the presumption of authoritarian responsiveness. We simply do not value authoritarian governance for its own sake and are therefore not a priori committed to what it entails. Yet like in the case of human freedom and responsibility whatever presumption we do make is vulnerable to the facts as they are. No presumption, however inherently desirable, can prevail in the light of overwhelming evidence to the contrary. What seems to make the relevant difference between these two cases, then, is the *extent* to which we are prepared to revise our presumptions in the light of the facts. Our humanist predisposition towards other human beings, on the one hand, keeps us from judging them irresponsible and treating them as objects of instrumental manipulation all too easily, while we certainly believe there to be a limit to the kind of behaviour that is compatible with being responsible in a deep normative sense.[347] Our commonly instrumental attitude towards authoritative agency, on the other hand, gives us little reason to be tolerant of the facts. We commonly have few moral qualms about an objective and instrumental attitude towards authoritative agency as we often eager to redesign and reorganize if we are unsatisfied with what it 'produces'. But despite this differential propensity to revise our presumptions in light of the facts, both kinds of presumption are similar in that they hinge on the same social mechanism: that of *trust*.

Trust, in short, is a way to cope with the behavioural uncertainty we have towards others.[348] In essence it involves the expectation of favourable (future) actions or conditions even when there are no clear rational grounds to do so.[349] In this particular case trust involves a way to cope with the uncertainty we inevitably face with respect to the responsibility of individual human actors and the responsiveness of authoritative agency. In line with our differential normative commitment towards human responsibility and authoritative responsiveness respectively, trust involves both something we are committed to in a primary sense and something we consciously aim to create in order to live together in a peaceful and cooperative way. To put it in different terms: "trust is both a human sentiment and a modality of human action".[350] What characterizes the analogy with the mechanism of trust here in particular is, first, that both our presumptions and trust are vulnerable to the facts. In

---

[347] Cf. Strawson (1963); Wolf (1990)
[348] Luhmann (1979, 1988)
[349] Cf. Gambetta (1988: 213 ff.)
[350] Dunn (1988:71)

the same way that our presumptions may very well be proven wrong trust can be betrayed.[351] That this comes with a price does not need any further argument. But what is characteristic, secondly, is that there is also a price for not making these presumptions; that is, a price for distrust. This price is to sell short the *potential* for human flourishing in community and cooperation. Without trust not only would we loose out on the fruits of human community and cooperation, but the *inherent* value of what it is to be human would suffer dramatically. It is only through trust then that we are able to overcome a predominantly instrumental and objective attitude towards other human beings and can have the kind of complex human associations we actually have.

But although trust is both empirically ubiquitous and pragmatically viable, it does not constitute a final solution for the problem of attributing a deep, normative understanding of responsibility both to human actors and authoritative agents. For 'pure' trust is sensitive and vulnerable to many influences that either may or may not have anything to do with the question of deep responsibility. The mechanism of trust will therefore not function resiliently, given that it is prone to indiscriminate external disturbances. It could function resiliently enough, though, if there existed some sort of 'flanking social mechanism' that was geared particularly to producing the kind of trust that our deep, normative understanding of responsibility is in need of.[352] Arguably, such a mechanism is even available. It consists of our *regulative* understanding of responsibility. Now clearly the mechanism of responsibility that is central to this understanding can support trust by holding to account those whom we trust to be responsible or responsive. It thereby significantly helps us to overcome the so called 'type-I errors', that is, the error of not presuming responsibility or responsiveness where the opposite is actually called for. But one must note that it can do so *only* by making us *more* vulnerable to 'type-II-errors', that is, the error of presuming responsibility or responsiveness when such a presumption is not warranted by the facts. Now this may not be very problematic in case we are presuming an authoritative agent to be responsive because we are not committed in any inherently normative way to such a presumption. But it would be problematic

---

[351] Gambetta (1988: 218-219)

[352] The notion of a 'flanking mechanism' is relatively straightforward. Think, for example, of the mechanism by which we stop an engine of a car by interrupting the supply of electricity to it. Such a mechanism can, and actually is quite useful for preventing theft and joyriding. But this mechanism could not prevent theft and joy-riding on its own. If there were just a button on the car's dashboard that switches of and on the electricity supply to the engine both thieves and joy-riders could have their way with the car. What is additionally needed then, is another mechanism, that of an ordinary lock for example, by which we could prevent the switch from being operated by anyone other than the owner of the car. The lock would then be a 'flanking mechanism' relative to that of manipulating the electricity supply to the engine. Although this example demonstrates a *conjunctive* flanking mechanism there also exist *disjunctive* flanking mechanisms. Think for example of the many reserve circuits that are purposefully built into airplanes to increase their safety. On the general idea that the presence of multiple social mechanisms contributes to the resilience of phenomena see: Brennan & Hamlin (2001)

were we to be forced to presume individual actors responsible when these actors simply do not have what it takes to be responsible in any deep, normative understanding of responsibility. Although it cannot be denied that some of the irresponsible – notably children – can be *made* responsible by subjecting them to the mechanism of responsibility, this is certainly not true for all those who presently lack what it takes for being responsible. The mentally incapacitated, for example, will arguably not respond well to such treatment, nor would it be justified to hold them responsible nevertheless. Thus we must draw the ironic conclusion that individual responsibility is more problematic than responsibility for authority because what may secure our trust in human responsibility may also be a source of grave injustice. Fortunately, no such irony is warranted in securing trust by holding accountable authoritative agents.

The problem of the organizational constitution of authority

But even if we can trust an authoritative agent to be responsive to the interests of his principals, that is, if we can somehow be adequately assured that he is capable of seeing and understanding what these interests are and of forming its intentions in the light of these interests, he still needs the capacity to actually do what it has decided to do. We have already seen in this regard that authority necessitates power, and that power in this context involves both collective and organizational constitutive features.[353] The question arises, then, whether we can attribute responsibility to these features, that is, whether we can attribute responsibility both (a) fully *and* (b) justifiably to authoritative agents at the same time. We already know that to rely exclusively on individual responsibility in this context may fall short of what is desirable and may even leave us empty handed. This is first because of the 'many hands' that are commonly involved in authoritative agency and which may dilute individual responsibility roughly in proportion to the number of hands involved, and secondly, because of the structural or organizational features that an effective and efficient collective effort necessitates and which may create a context of practical necessity in which individual responsibility may very well dissolve completely.[354] Since a lot is already known on the problem of collective responsibility,[355] let me focus here on the structural or organizational features of authoritative agency and on how they impact on our attribution of deep responsibility.

That organizational features of authoritative agency can make individual responsibility more or less dissolve and make some sort of attribution of responsibility to organizations desirable can be explained best with the aid of an example. The example involves a supreme court that must decide whether or not to grant a stay of execution to a convicted murderer and rapist.[356] Let us assume that this

---

[353] See chapters I and II of this study.

[354] French (1979), Thompson (1980, 1987); Wolf (1985)

[355] A relatively recent overview of this debate is given by May & Hofman (1991)

[356] Although I have borrowed this example from Pettit (2001, unpublished manuscript), I have modified it somewhat to fit the particular argument I want to make here. Essentially the

supreme court consists of three judges: A, B and C and that legal doctrine provides three grounds for granting a stay, either one of which is a sufficient ground. These grounds involve: (i) the availability of new evidence that creates reasonable doubt on the convict's guilt, (ii) some reasonable suspicion that the conviction was secured by a violation of the convict's civil rights, and (iii) considerations of a humanitarian nature. Table 4.1 summarizes what each judge thinks about each separate ground.

	Judge A	Judge B	Judge C	Collective judgement:
New evidence?	No	No	Yes	No
Violation of civil rights?	No	Yes	No	No
Humanitarian pardon?	Yes	No	No	No
Individual judgement:	Yes	Yes	Yes	Yes/No

Table 4.1 The doctrinal or discursive dilemma.

What is clear from this table is that the outcome of the decision will differ with the decision procedure followed. If the judges vote on the overall decision they will decide in favour of a stay. If the decision procedure is organized in such a way that each ground is deliberated on separately and collectively and subsequently decided on by ordinary majority voting the court will decide against it. Let us assume that the latter scenario obtains and that a stay of execution is denied. The convict is executed at the scheduled execution date. Let us assume, finally, that five years after the execution a new kind DNA-test is performed on the evidence presented at the murder trail. It is concluded from this test that the DNA found on the victim's body is not that of the man convicted but is that of another man. This man is subsequently identified and arrested and confesses to the murder. Who or what is to blame for this failure of justice?

What is clear from this example is that no individual judge can be blamed, for each judge would have voted in favour of a stay had they voted on the issue as such. But since they voted on each of the individual legal grounds instead it is the *organization* of their voting rather than their individual reasons and decisions that is ultimately responsible for the result. Now of course one could object that the way in which the vote was organized was up to the judges and that they are therefore – albeit

---

example is an expression of what has become known as the Condorcet-paradox. That my example involves judicial decision-making is no coincidence. Pettit also gets his ammunition from the context of collegial judicial decision making: Cf. Kornhauser & Sager (1986); Kornhauser (1992); Kornhauser & Sager (1993); Chapman (1998a, 1998b); Brennan (2000)

indirectly – ultimately individually responsible for what they decided.[357] But to that the judges could reply that although the organization of the vote was indeed *formally* up to them, it was not up to them individually because they were *materially* required to vote on each individual ground for a stay of execution given the legal obligation to consider and base their decision upon the law and prior judicial decisions. They could argue that this requires that they act as *one* body because both the law and the previous decisions of the court are binding on them *as a body* and not on each judge individually. After all, the composition of the Supreme Court may change over the years with some judges leaving and new ones being appointed. They could argue, furthermore, that this requires the *collectivisation* of their deliberation on the cases brought before them and that this in turn necessitated the design and implementation of an appropriate decision-making procedure.[358] They could even hold that, because of the obligation to base their decisions upon the law and on prior judicial decisions and their structuring their decision making process to optimally comply with this obligation, they form intentions *as* a supreme court and *not as* individual judges, and they would be right at this, given that the reasons the court has with respect to certain decisions need not be reducible to the reasons the individual judges may have with respect to the same decisions.

But if an organization can form intentions, that is, if it can reason and decide on issues *as* an organization, then surely it can be responsible as an organization in a deep normative sense too.[359] It could be responsible because of (a) its internal decision structure that aggregates the reasons of its individual constitutive parts into a whole in such a way that it need not be reducible to its parts again and (b) that it does so in a manner that is rational at the organizational level.[360] Moreover, that an organization can be deeply responsible in this way would nicely serve our needs, given that individual responsibility commonly falls short in adequately accounting for organizational actions and their consequences and sometimes even leaves us empty handed. We should not allow ourselves to get carried away by wishful thinking here, though. There are also some serious complications that may prevent us from attributing deep responsibility to an organization even if we grant that organizations can be intentional actors and that they can even constitute metaphysical[361] or institutional[362] persons. Let me point out the two salient objections in particular.

The most important objection against attributing a deep, normative understanding of responsibility to organizational agents comes from comparing organizations with

---

[357] Assuming it is possible and appropriate to derive individual responsibility from collective inaction in this context. For arguments supportive of such an assumption in *general* see: Held (1970/1991); May (1990).
[358] Pettit (2001), Pettit & List (2001)
[359] Cf. French (1979, 1984)
[360] French (1979:144); Pettit (2001)
[361] French (1979)
[362] Pettit (2001)

sociopaths.[363] Although the latter are clearly capable of rational deliberation and rational planning they are typically thought of as incapable of *feeling* and *understanding* what is right about right and what is wrong about wrong. Thus they may *know* that they should not steel, rape and murder because they have been told that it is wrong to steel, rape or murder, but they do not *feel* that these things are wrong or *understand* why it is that these things are wrong. What sociopaths seem to be missing, then, is some inner faculty that enables them to track what is right and what is wrong by *their own* lights. What they know about right and wrong they know from others and they know it in a way that does not yet constitute an *understanding* of what they know. For a sociopath typically cannot take the perspective of a *sensitive* "I" that it would need to reflectively project upon himself the kind of considerations that are necessary to *understand* what is wrong about what it cognitively knows to be wrong. Sociopaths, in short, do not have consciences.[364] They therefore cannot be responsible in a deep, normative understanding of responsibility.

The point is, now, that organizations are typically like sociopaths in this regard. Like a sociopath an organization can be both plan-full and rational in the sense that there may be a rational relationship between what it does at this time and what it does later, and between what it does in this context and what it does in another. But also like sociopaths organizations lack the sensitivity to really *understand* what it is that is right about right and wrong about wrong. Thus like sociopaths, organizations do not have consciences. Whatever sensitivity there may exist within organizations exists with its human members. But unlike intentions that *can* exist at an organizational level by *supervening* upon the acts and intentions of the organizational members,[365] there is no such thing as organizational sensitivity and conscience supervening upon the sensitivity and consciences of the organizational members. Sensitivity, particularly regarding the moral sentiments, remains a distinctively human trait; a trait we simply do not see at the organizational level. Organizations will therefore always depend upon humans to really understand what is right and what is wrong. Organizations in this view therefore cannot be deeply responsible agents.

But, and this is our second objection, perhaps neither do we really want organizations to be deeply responsible agents. We have already seen that our deep normative understanding of responsibility is embedded in, and partially constituted by other normative considerations. These include the virtue of responsibility and the inherently normative presumption that humans posses that virtue absent evidence to the contrary. It was explained that this presumption was warranted because to presume otherwise would be undesirable in and of itself. The problem, now, is the extent to which we commit ourselves to such related normative commitments by saying that organizations can be responsible agents in a deep, normative sense of the word. Thus we must seriously ask ourselves, for example, whether we really are prepared to extend to organizations the kind of legal rights we thus far have

---

[363] Wolf (1985:278 ff.)
[364] Wolf (1985:278)
[365] Pettit (1993, 2001)

exclusively reserved for human actors, merely because we want to hold organizations *criminally* liable for their acts and omissions.[366] The point is that it may very well be both wise and appropriate to treat organizations with precisely the kind of general instrumental and objective attitude we find so repulsive and dehumanising when it is directed towards human beings. We should not forget that organizations are human artefacts and that they therefore do not *automatically* deserve the same kind of respect that human agents do.

Conclusions

Let me review the major line of argument of this chapter and formulate some preliminary conclusions. What was at stake in this chapter was the question whether there could be such a thing as a *conceptual* understanding of responsibility for authority. It was argued that whatever understanding of responsibility for authority we do have can be conceptual *if and only if* it can fulfil *each* of the requirements underlying the three basic understandings of responsibility discussed in this chapter. What has become clear in this paragraph, however, is that it is unlikely that this is the case. We have run into several formidable problems that all stand in the way of turning our understanding of responsibility for authority into a *conceptual* whole. We have successively discussed: the problem of ascription, the problem of practical necessity, the problem of constituency, the problem of responsiveness, and the problem of attributing deep responsibility to organizational features of authoritative agency. I think it is justified to conclude from this discussion that the joint impact of these problems make the conditions in which a conceptual understanding of responsibility for authority obtains *exceptional* at best. But this chapter has also resulted in another remarkable finding. Not only are these basic requirements difficult to fulfil simultaneously for a notion of responsibility for authority, but they are also highly problematic for an understanding of responsibility as such. We have seen that – with exception of the problem of the organizational constitution of authority – each of the problems discussed also impacts negatively on our understanding of individual responsibility. The finding that a conceptual understanding of responsibility obtains only in exceptional circumstances is highly remarkable, however. It is clearly at odds with the facts as we know them to be. For in everyday life the practice of responsibility is ubiquitous in modern society and typically quite unproblematic. Perhaps, then, it is not our understanding of responsibility that is at odds with the facts as they are, but rather our insistence that that understanding be *conceptual*, that is, the demand that *all* requirements following from the three basic understandings of responsibility must be met simultaneously. But if the notion of responsibility is *conceptually* problematic as such, the question remains how the practice of responsibility should be conceived of then. This question is at stake in the next chapter.

---

[366] Thompson (1985:213-214)

V

Responsibility for Authority II: A General Institutional Perspective

Introduction

The previous chapter ended with a paradoxical conclusion. Although a *conceptual* understanding of responsibility was too demanding in the sense it is not very likely that the necessary conditions that obtain for such an understanding will *actually* be met simultaneously, everyday observation shows that the practice of responsibility is ubiquitous in modern society nevertheless. It is clear that this paradox needs resolving. It is also clear where a resolution must come from. Given that we do not want to give up on the presumption that our conceptual understanding be *about* something, the facts must eventually trump our conceptual commitments. Sidetracking our ambition to come to a conceptual understanding of responsibility makes sense only, however, if there is an alternative available; one that is capable of overcoming the problems that undermined taking a conceptual perspective. Arguably, there is such an alternative. Recent developments in a number of different disciplines have led to a revival of the theory of *institutions*. The claim I want to defend in this chapter is that an institutional understanding of responsibility is theoretically more powerful than a conceptual one, while at the same time being practically more viable. Before I can elaborate on this claim, I must first explain what I take institutions to be. Since there are quite a few different approaches in the theory of institutions, I will focus on three features that these approaches have in common first, and subsequently address the diversity of approaches by briefly discussing three distinctive branches of institutional theory. In my view these three branches represent three basic *mechanisms* by which institutions work. Not coincidentally, these three mechanisms cohere nicely with the three basic understandings of responsibility I have discussed earlier. Sketching the outline of a general institutional theory of responsibility will therefore proceed on familiar grounds. After having sketched this outline in some detail, I elaborate on my claim that an institutional understanding of responsibility is superior over a conceptual one. I conclude this chapter with exploring in broad strokes what a general institutional theory of responsibility for authority may look like by relating it

to the currently dominant agency theoretical perspective on responsibility for authority.

The Concept of Institution

The notion of institution - like that of responsibility - has acquired quite a number of different meanings in academic debate. Even a quick glance of the prevailing literature on institutions shows that these meanings differ significantly along disciplinary lines. Thus the question what precisely is an institution gets a different answer inside the disciplinary boundaries of economics[367], for example, than it would get within the disciplinary confines of sociology,[368] political science,[369] law,[370] or organization theory.[371] Rather than following these disciplinary lines I aim to give a theoretically more systematic account here of what I believe institutions consist of. This is not just because disciplinary boundaries are often drawn arbitrarily, but also because contemporary research and theory development in social science and the humanities is increasingly boundary crossing. This is true in particular for theory of institutions. Not only has institutional theory developed interactively over different disciplinary boundaries, but arguably the most interesting applications of institutional theory have taken place in the interdisciplinary study of organizations.[372] I will focus on a number of general features on which the different branches of institutional theory more or less converge first and come back later on some systematic differences between these branches.

Although the notion of institution has been around for quite a while in the study of man and society, the nature of institutions can be understood best, I think, against the background of relatively recent developments. Somewhere towards the end of the first half the twentieth century the presumption became dominant that, like the natural sciences, the sciences of man and society should be *positive* sciences; i.e. sciences based on facts. In putting this presumption to work, however, two additional assumptions were made somewhat overzealously. This was, first, that the study of man and society can and should be conducted in the same way that natural phenomena are studied – i.e. by taking an external so called third-person's-perspective

---

[367] Accessible and relatively recent reviews of institutional theory in economics are given by: North (1991); Hodgson (1998); Aoki (2000) and Williamson (2000).
[368] Institutional theory in sociology more or less coincides with that in organization theory (but see note 6). Recent overviews are given by: Powel & DiMaggio (1991); Scott (1995), and Brinton & Nee (1998).
[369] A comprehensive and recent overview is given by Peters (1999).
[370] Recent institutionalists in Law would include: MacCormick & Weinberger (1986) and Morton (1998)
[371] Recent overviews are given by: Powel & DiMaggio (1991) and Scott (1995).
[372] Within the theory of organization roughly two institutional paradigms have emerged. There is a more economic institutional paradigm dominated by the work of Oliver Wiliamson (1975, 1985, 1991, 2000) and there is a broader social science approach, which is dominated mainly by sociologists. Cf. Powel & DiMaggio (1991); Scott (1995); Tolbert & Zucker (1996)

on social reality – and second, that the natural unit of analysis in such studies was the individual human being. These assumptions have become known as 'behaviourism' and 'methodological individualism' respectively.[373] Although these assumptions have greatly facilitated the development of the social sciences by introducing a certain degree of academic rigor and unity to the relevant fields through spurring empirical research and rigorous analysis of theories through formalization, it gradually became clear that the assumptions themselves did not square with the facts as they are. Thus rather than being able to reduce social order and the central social phenomena encountered in day to day reality to the actions and dispositions of individual actors, it was found that social reality contains 'chunks' of irreducible social 'matter' and that this 'matter' significantly shapes the behaviour of individual actors. To put it in other words: it was gradually discovered and understood that individual action is *socially constrained* and that the social constraints to individual action warrant research and theory development on their own account..

Not surprisingly, the most central concept in this re-emerging field of social theory and research has become the notion of institution. Institutions can be understood – with an air of circularity – as those conditions by which man is *socially constrained* in his actions. Although I will address the question how institutions constrain individual action later on, let me give a few paradigmatic examples here of what I have in mind. A first example is both basic and straightforward. It involves buying a ticket for the bus. Although economics has stereotypically presented consumer choice as *autonomously* given, it could certainly not make sense of something as common as buying a bus ticket on that basis. For in performing this apparently simple and everyday action quite a number of social constraints are already operative that make this action possible in the first place. It is assumed, first, that exchanging money gives one certain rights and that buying a bus ticket gives one the right to ride the bus in particular. Moreover, what counts as money and what constitute busses is apparently also given in this example. It must be clear that what enables us to make and subsequently act on such assumptions is not up to us individually. For these assumptions to work they are *necessarily* contextual in the sense that they transcend the scope of any action in particular.[374] This point is nicely demonstrated in a second example involving economic institutions. Although (neo-) classical economics, again, has typically assumed the transaction as given and unproblematic, the practice of

---

[373] That institutional theory contrasts most sharply against the background of these two assumptions is held either explicitly or implicitly by many theorists of institutions. Explicit attention to this thesis is paid by Peters (1999:11 ff.). That this thesis is implicitly adhered to more widely is demonstrated by the fact that the behavioural revolution in the social sciences is often considered as a demarcation point separating 'old' from 'new' institutional theory. See for example: Powell & DiMaggio (1991:2 ff.); Scott (1995:1ff.); Goodin (1996:2ff.). I will not use the adjectives old and new here because my aim is not to position institutional theory as a perspective in social science, to which this distinction in my view belongs.

[374] Jepperson (1991) conceptualises institutions by opposing them to *actions* more or less along these lines.

'transacting' is neither unproblematic nor natural. Transactions aren't natural in general for exactly the same reasons that buying a bus ticket isn't natural in particular. More interesting, though, is that neither are transactions socially unproblematic. It is not just a theoretical possibility, after all, that payment does not result in the delivery of the good or service purchased. In fact, transactions where "clear agreement" is more or less automatically coupled to "clear performance" are rather rare.[375] If they do occur, they typically do so in a highly institutionalised context – such as a commodity or stock exchange for example – which nicely demonstrates the point I want to make here. For transactions characteristically involve a non-trivial risk of them being frustrated by the motivational shortcomings of the transactors. To cope with such behavioural risk under conditions of cognitive uncertainty – lack of foresight for example – a variety of economic institutions have evolved. These include amongst others: contract law, merchant banks, the firm, the limited liability company, and various hybrid (non-market) governance structures such alliances and franchising arrangements.[376] Without these institutions and the social constraints they impose on economic action economic systems could be nowhere near as complicated and efficient as they typically are in modern societies. But this example also demonstrates something else. This is that institutions do not merely *constrain* individual action. They also *empower* it precisely by imposing constraints. It is clear that without social constraints there could be no such things as bus tickets or smoothly functioning product markets. But once there are such things, they empower individual actors by increasing their options for action.

A second general feature of institutions is closely related to the first. Although institutions characteristically impose constraints upon individual action this does not involve just any kind of constraint. The constraint exercised by "brute facts" such as purely physical conditions are clearly not institutional.[377] Thus while the human ability to jump, for example, is constrained by both the force of gravity and the physical constitution of human beings, these constraints are social nor institutional. There is therefore something *ontologically* peculiar to the kind of constraint exercised by institutions. This is that they exist by virtue of *collective* human *intention*. Thus a bus ticket, again, does not constrain and empower us to ride busses because of it physically being a piece of paper. It does so only because of a collective intention imposing on a piece of paper the so called "*status function*" of being a bus ticket.[378] In a banal sense, then, it is true that a bus ticket *is* a bus ticket because we believe that it is a bus ticket. There is nothing mysterious or radical about this apparently circular claim. It isn't mysterious because it is a common feature of everyday reality that we impose status functions on physical objects and events over and above their intrinsic properties. Thus we call certain shorelines 'beaches' and consider certain syndromes of sounds to be music. The claim isn't radical because it does not imply that some

---

[375] McNeill (1974:738)

[376] Cf. North & Thomas (1973); North (1990, 1991); Williamson (1975, 1985, 1991, 2000)

[377] Anscombe (1958)

[378] Searle (1995:13ff.); emphasis mine.

radical subjectivism obtains with respect to social reality. While it is true that shorelines can be beaches only relative to human use and valuation, it does not make much sense to say that a given piece of shoreline *actually* used as a beach is a beach or not depending on my own subjective account of the facts as they are. A distinction must be made between the *ontological* and *epistemological* subjectivity of what John Searle has named "institutional facts".[379] While the latter are necessarily *ontologically* subjective in that they could not exist independently of human use and valuation, they are *epistemologically* objective in exactly the same way that stones are once they have been established by collective human intention.

To identify a third general feature we must elaborate a bit more on Searle's micro-ontology of institutions. As explained, institutional facts differ from brute facts in that a status function (Y) is (intentionally) imposed on an object or symbol (X) in context (C) such that the basic formula for an institutional fact is: X counts as Y in C.[380] Thus we take a piece of paper (X1) to be a vote (Y1) given the context (C1) under which *we accept* it as being a vote. In principle this formula can be iterated indefinitely whereby ever more complex institutional facts are created. Thus the Y1 term of the initial formula can become the X2 term at a higher level of complexion where we collectively *accept* that under conditions C2 (getting the majority of votes) Y2 (winning an election) obtains. It is important to note that I have printed the verb *accepting* in italics here. The reason is to emphasize, again, that institutional facts can exist only by virtue of us accepting them as such, that is, by the grace of our collective intention imposing status-functions on events, objects or symbols over and above this natural properties. The point I want to make now, is that this acceptance of status-function can take place in *various* ways and – in the view I will expound later – via various mechanisms. The *equifinality* of the various ways in which status-function-imposition can take place, then, is a third general feature of institutions. As we shall see, this feature helps to explain both the resilience of institutions and the existence of different branches of institutional theory. With respect to the latter it can be said that different branches of institutional theory simply emphasize different ways in which status functions are imposed on either brute or other institutional facts. Let me briefly describe three basic branches of institutional theory before returning to the question how institutional theory can nevertheless be seen as whole.[381]

Three Branches of Institutional Theory

Elements of a first branch of institutional theory were already touched upon above in the example of how economic institutions socially constrain potentially opportunistic

---

[379] Searle (1995:7-9)
[380] Searle (1995:43 ff.)
[381] As was the case with my discussion of the three basic understandings of responsibility earlier, I draw heavily on the distinctions within the field of institutional theory made by Scott (1995: 33 ff.)

individual economic action.[382] The branch of institutional theory for which this example was paradigmatic has been named the "regulative pillar" of institutional theory because it relies heavily on *intentional* social regulation of individual action by means of *sanctioning* and *coercion*.[383] Thus economic actors are believed to honour their commitments because the existence of economic institutions makes the expected cost of defection higher than its gains. Not surprisingly, this regulative approach to institutions has been developed and subscribed to mainly within the discipline of economics and rational choice theory more broadly. The central 'logic' by which this branch of institutional theory explains both the existence and the functioning of institutions is the so called "logic of consequentiality",[384] that is, the logic that works through the rational anticipation of the *consequences* of action. In this view, institutions are often purposefully designed "governance structures" who's primary function is provide countervailing stimuli in order to cope with the *motivational* shortcomings of the actors within their scope of operation.[385] As these governance structures commonly involve high operating costs – making explicit normative expectations, continuous monitoring and the incidental application of sanctions are both necessary and expensive[386] – they will only persist for as long these costs stay well below the benefits they create. Thus the logic by which institutions work in this view is the same one explaining their persistence.

A second branch of institutional theory can be described in contradistinction to the first. According to this branch individual action is not so much socially constrained by the anticipation of its social consequences but rather by what actors perceive to be *appropriate* given the circumstances. Appropriately, this branch of theory is thought to constitute the "normative pillar" of institutional theory.[387] The central logic by which institutions work and persist in this view has been named "the logic of appropriateness".[388] The difference between this logic and the former is that it does not work through rational anticipation of consequences but rather through *valuation* and *normative* expectations. At an everyday level of abstraction and

---

[382] As was briefly touched upon in the introduction of this study, the role of opportunism in explaining the institutions of capitalism is stressed by Williamson (1985). In Williamson's view, (a) opportunism and (b) bounded rationality are two *constants* in the province of economic action. The existence of different economic institutions or "discrete structural alternatives" (Williamson, 1991a) is then explained by identifying the variable causes of this institutional variance, which are (a) "asset specificity", (b) frequency and (c) uncertainty. (Williamson: 1975; 1985; 1991) The theoretical importance of opportunism, however, has also been criticized; amongst others by Noorderhaven (1995) and Goshal & Moran (1996)

[383] Scott (1995:35). The fact that I mention sanctioning and coercion separately here, does not imply that I believe there to be a sharp distinction between the two. As a matter of fact, I do not, but that is an issue that transcends the point I want to make in this paragraph.

[384] March (1981), March & Olson (1984, 1989, 1995)

[385] Cf. Williamson (1991a)

[386] Cf. Scott (1995:37); Pettit (1996a)

[387] Scott (1995:37 ff.)

[388] March & Olson (1984, 1989, 1995)

aggregation, the logic of appropriateness can be seen to operate in the existence and centrality of *roles* in social life.[389] Thus doctors, for example, may very well continue to treat patients even when it is completely clear to them that the treatment will have no effect, because treating patients is simply what is appropriate for doctors to do. But it is not just in this somewhat negative sense that social roles shape a significant part of our social lives. Society – even the face-to-face one of the ancient Greek polis – would simply not be possible if actors were forced to continuously deliberate on what is best for them to do. Roles, rules and norms, then, are central in reducing complexity of individual action while at the same time providing it with purpose and direction. But there is more to this normative branch of institutional theory than merely the functioning of roles, rules and norms in human association. For ongoing action in the context of concrete roles, norms and rules can create a normative reality in its own right once it becomes infused "with value beyond the technical requirements of the task at hand."[390] That valuation constitutes an important component of institutional reality is demonstrated nicely by the vigour with which democracy and the 'Rechtsstaat' are commonly defended against criticism in modern society. This vigour suggests that there is much more at stake here than merely the formal requirement that government action be constrained by legal norms and a system of governance in which governments are appointed into office through general and competitive elections. More precisely it suggests that the 'Rechtsstaat' and democracy have become very important political values on their own account.

A third branch of institutional theory has been labelled the "cognitive pillar" because of its emphasis on the social constraints contained within and working through processes of human cognition.[391] This branch of theory has evolved primarily inside the disciplinary intersection of the cognitive sciences,[392] sociology[393] and (cultural) anthropology.[394] Its central premise is that individual actors are necessarily socially constrained in the way they *understand* social reality and themselves as actors within that reality. Thus the focus here is on the way that *meaning* and *representations* of self and social reality are (continuously) shaped and (re-) enacted through ongoing social interaction. Like the normative pillar, the cognitive view of institutions asks attention for the centrality of rule following in the both the functioning and reproduction of institutions. But where the former emphasizes the valuational aspects of rule following, the latter focuses almost exclusively on the way in which rule following is central to cognition and the representation of (social) reality. This differential focus manifests itself in the kinds of rules considered relevant. While normative institutional theory primarily involves normative expectations and *practical* rules, the cognitive branch of institutional theory is concerned with the 'taken-for-

---

[389] Cf. Berger & Luckmann (1967)
[390] Selznick (1957:16-17)
[391] Scott (1995)
[392] Cf. Simon (1945); March & Simon (1958); Cyert & March (1963)
[393] Cf. Berger & Luckmann (1967), Gofman (1967)
[394] Cf. Geertz (1973); Douglas (1986)

grantedness' of everyday life and the (theoretical) rules constitutive of both social reality and our understanding of it.[395] If there is a dominant logic to be associated with this branch of institutional theory it is "one of orthodoxy"[396]; i.e. a logic according to which belief, meaning and understanding are fixated through ongoing social interaction.

A General Institutional Theory of Responsibility for Authority

I have already briefly touched upon how, in my view, these different branches of institutional theory tie together into a whole, that is, into a *general* theory of institutions. Although elements of such a comprehensive approach have already been formulated elsewhere,[397] the view that I propose here has the ambition to be more fine grained and precise, both in answering the question what institutions are and in explaining how they operate and are able to persist through time. I have already discussed what I believe to be the three basic characteristics of institutions. These are that: (a) institutions involve *social* and *contextual constraints* to individual action, that (b) are constituted by *intentional* imposition of *status-functions* on either brute or other institutional facts, and that (c) this imposition of status-function can come about in *various ways*. I have also discussed in broad strokes how different branches of institutional theory give different explanations of how status functions are imposed in everyday reality. I have pointed at the logic of consequentiality, the logic of appropriateness and the logic of orthodoxy as micro-explanations for the functioning and persistence of regulative, normative and cognitive institutions respectively. My proposal, now, is to understand these different explanations as separate yet complementary *mechanisms* of status-function-imposition that each – under differential conditions yet to be specified – can function on their own account, yet that more typically operate simultaneously. In this view, therefore, institutions are what D'Andrade has called "*overdetermined*…(..)...cultural meaning systems ..(.).. in the sense that social sanctions, plus pressure for conformity, plus intrinsic direct reward all act together to give a particular meaning system its directive force".[398] It is important to note that I have emphasized the verb 'overdetermined' here. Again, although each 'institutional logic' can independently be effective as a *mechanism* of status function imposition it is more typically found that two or more mechanisms work simultaneously. As we shall see, this *overdetermination* of action by multiple mechanisms is what makes institutions resilient and robust.[399] Before applying this general view to

---

[395] Cf. DiMaggio & Powell (1991) but also Searle (1969, 1995)

[396] Scott (1995:45)

[397] Most notably, I think, by D'Andrade (1984) but also by DiMaggio & Powell (1983) and Scott (1994, 1995)

[398] D'Andrade (1984:98); emphasis mine.

[399] For a general argument that the presence of multiple mechanisms contribute to the resilience of phenomena see: Brennan & Hamlin (2001)

the different basic understandings of responsibility, however, let me elaborate a bit on how these mechanisms work.

In general, institutional mechanisms work through either (a) *ex ante* screening (i.e. filtering) for options or (b) *ex post* sanctioning of behaviour.[400] Thus cognitive institutions and the logic of orthodoxy work by effecting the beliefs and options for action an individual actor has or perceives at any point in time. As explained, this mechanism screens or filters for those beliefs and options for action that are *socially* meaningful. Now it must be remembered that this mechanism does not merely constrain belief and action in the narrow sense of the word. It also empowers individual actors to have certain beliefs and options for action they could not have had without prior constraint. Thus in order to even conceive of a stock exchange, for example, there must first be a conception of capital, which in turn relies on an understanding of property. As a rule, constraining for beliefs and action at more natural and lower order levels empowers actors at higher order levels of abstraction and 'social construction'. Normative institutions and the logic of appropriateness, secondly, function by *both* screening for beliefs and options and the (non-intentional) sanctioning of actions taken or contemplated. Thus most citizens are law-abiding, for example, because they do not even conceive of breaking the law, while for others it is not so much this lack of conception nor the threat of external punishment but rather the (inherent) undesirability of being seen as a law-breaker that keeps them from illegal conduct.[401] Regulative institutions and the logic of consequentiality, finally, work mainly via intentional or non-intentional sanctioning. Thus the behaviour of firms in a market, for example, is socially constrained by the (aggregate) non-intentional sanctioning produced by consumers and other firms in that market whilst sanctioning is typically intentional within these firms, with organizational constituencies being rewarded or punished through the purposive interventions of management.

We can now see how the different basic understandings of responsibility I have discussed in the previous chapter constitute three alternative mechanisms by which the status-function 'responsible' can be socially imposed on actors and agents. Thus our cognitive understanding of responsibility, firstly, explains how social processes constitutive of human cognition label and fixate the status-function of being 'responsible for X' on what is (or can be) understood as the *intentional* cause of X. And our regulative understanding of responsibility, secondly, focuses on how considerations of efficacy in upholding certain normative expectations (N) in a given context (C) are decisive in imposing the status function 'responsible' (for the realization of N) on actors or authoritative agents interacting in C. Our normative understanding of responsibility, finally, points at the role that normative and valuational considerations play in imposing the status-function 'responsible' on actors

---

[400] I borrow the following basic account of institutions working through (a) screening and (b) filtering from Pettit (1996a). Elements of this approach are already foreshadowed in Elster (1989)

[401] The latter is referred to by Pettit (1993:226 ff.) as the "intangible hand" mechanism.

or agents. As explained, this involves both the question whether or not our attribution of responsibility can be justified in an ultimate sense and the inherent desirability of being considered a responsible actor or agent. We can also see more clearly now, in what sense an institutional understanding of responsibility differs from a conceptual one. Where the latter requires that the three basic mechanisms of status function imposition work in *conjunction*, the former allows these mechanisms to apply *disjunctively* dependent on the conditions at hand. The benefits of this are twofold. First, an institutional approach to responsibility is much more resilient and therefore empirically much more viable than any conceptual understanding of responsibility possibly could be. This coheres with our everyday observation that the practice of responsibility is ubiquitous in modern society. Second, whilst an institutional approach to responsibility can accommodate both responsibility for authority and individual responsibility within a single conceptual framework – i.e. that of a general institutional theory – it is doubtful whether there can be a conceptualisation of responsibility for which this is true. I will elaborate on each of these points.

There are roughly two reasons why an institutional approach to responsibility is more resilient and therefore empirically more viable than a conceptual one. A first reason is that an institutional approach enables us to adjust our attributions of responsibility to the particular circumstances at hand. This is possible because an institutional approach allows the different mechanisms of status-function-imposition to work disjunctively. Thus, in applying the notion of responsibility in conditions where the normative commitment at stake is relatively low, we can tolerate not all requirements pertaining to a normative understanding of responsibility being met simply because these requirements may not be very relevant to the case at hand. Such is the case, for example, when responsibilities are attributed within a group of friends on a weekend sailing trip. We can imagine our regulative understanding of responsibility being relevant here because sailing requires different tasks to be adequately fulfilled and coordinated therefore warranting an *effective* distribution of responsibilities over the crew members. And we can also see how our cognitive understanding of responsibility is highly relevant in this example because without being able to see the course and movements of the boat as the result of the *intentional* actions of the crew members there could be no such thing as sailing. Yet at the same time it is clear that our normative understanding of responsibility is less relevant here as there is commonly little need to attribute a deep normative understanding of responsibility to any of the crew members on a relaxing sailing trip with friends.[402]

A second reason why an institutional understanding is more resilient and therefore more viable than any conceptual understanding of responsibility possibly could be is slightly more complicated. In essence, the point is that problems or doubts with respect to the functioning of any of the three particular mechanisms of status-function-imposition can sometimes be resolved by the unproblematic functioning of another basic mechanism. Effectively this comes down to reducing one to the other.

---

[402] Exceptions noted of course. A day of casual observation in a Dutch water-lock will show that these exceptions are empirically everything but exceptional.

Because this point may be difficult to grasp *in abstracto* let me give a few examples involving some of the problems discussed in the previous paragraph. A first example involves our uncertainty over the presumption that a given authoritative agent has what it takes to be appropriately *responsive* towards the interests of its principals. It was argued that such uncertainty is coped with by trust as a social mechanism, and that this mechanism cannot be expected to function resiliently by its self, given that it is vulnerable to potential external disturbances. It was argued, furthermore, that a so called flanking social mechanism could be an effective remedy for this problem and that in this context in particular our *regulative* understanding of responsibility could perhaps function as such a flanking mechanism. Systematically subjecting authoritative agents to the mechanism of responsibility may eventually result in them being appropriately responsive to the interests of their principals even if this responsiveness itself comes about unintentionally.[403] Such a solution would not simultaneously undermine a normative understanding of responsibility in this context because contrary to what is the case with human actors, holding accountable authoritative agents regardless of what they are capable of does not bite with our deep normative commitments towards such agents. In a nutshell, then, the point is that in this particular case our doubts about the functioning of the logic of appropriateness can be coped with by relying on the adequate working of the logic of consequentiality.

A second example is closely related to the first as it involves the problem of ascription. Seeing certain facts as the consequences of acts or omissions of an intentional authoritative agent is often highly problematic for reasons explained earlier. Thus it is commonly difficult, for example, to ascribe the low stock value of a company to the management of that company even if only because of the many different determinants of stock prices in general. The problematic functioning of our cognitive understanding of responsibility in this context, however, can effectively be compensated for by the proper functioning of one or both of the other basic mechanisms of status-function-imposition. We could, for example, apply our regulative understanding of responsibility here in such a way that we knowingly hold authoritative agents accountable for things they could not possibly have influenced and we could do so for the reason that this assures the greatest degree of control over what we hold this agent responsible for.[404] In fact, it is quite common to dismiss the management of a poorly performing company even if we do not know to what extent performance is attributable to management. Although such a move appears to bite with our normative understanding of responsibility, the reverse could also be argued. It is often held that positions of authority come with additional responsibilities (*noblesse oblige*) and that resigning a position of authority when (normative)

---

[403] Appropriate responsiveness in this context may be an attribute that comes about in an *evolutionary* manner, that is, via unintentional variation and selection processes working through the mechanism of responsibility. A comparable account of the evolution of so called "background abilities" (i.e. non-intentional dispositions towards institutional facts) that are central to the functioning and persistence of institutions is given by Searle (1995:127 ff.)

[404] Van Gunsteren (1974)

expectations are not being met is primarily a question of honour and not of efficacy.[405] I will come back to this later on.

A third example demonstrates that it is not just true for responsibility for authority that the proper functioning of one mechanism of status-function-imposition can compensate for the problematic functioning of another. As explained in the previous paragraph, the thesis of determinism threatens even our normatively least involving understanding of responsibility. For if everything is determined by initial conditions and the natural laws applying to them, it would not make sense to say that an actor A is responsible for X, for A would be just as determined as everything else and there would therefore be no reason to stop short of the causes of A in attributing responsibility. More importantly, however, holding A accountable for X could ultimately not be justified because in no sensible way could X be understood as being really up to A. It has been argued, therefore, that our deep normative understanding of responsibility is incompatible with a deterministic worldview.[406] We need not despair, however, about our normative understanding of responsibility being undermined by a deterministic worldview. For other mechanisms than our normative one can arguably provide a solid enough foundation for our normative understanding of responsibility to rest on. It has been argued, first of all, that even in a deterministic world it could still make sense to hold A accountable for X because doing so may in some way be socially efficacious.[407] Thus even if all criminal behaviour were completely determined, for example, it could still be sensible to incarcerate criminals simply because incarcerated criminals will not be on the streets committing more crimes. It is clear, then, that our regulative understanding of responsibility could do some work here even if we can already see that it could not support the full weight of a truly normative understanding of responsibility. For the "objective attitude" that goes with seeing other human beings from the perspective of policy and manipulation is missing something that a normative understanding of responsibility necessitates.[408] This is that being responsible is a central (and constitutive) human virtue and that merely being human therefore warrants a non-objective attitude in its own right.[409]

Taking a *general* objective attitude to accommodate for a deterministic worldview also shows something else, however. This is that such a *general* attitude is incompatible with the facts of human interaction as we know them to be. Thus although, as Strawson argues,[410] it is quite common for us to *incidentally* suspend our reactive attitudes of blame and resentment in cases were harm or injury have been caused by

---

[405] In politics, this has become known as the 'Carrington Doctrine', after the British secretary of defence Lord Carrington, who resigned voluntarily after not having foreseen the Argentinean invasion in the Falklands in 1982 .

[406] Cf. Van Inwagen (1983)

[407] Cf. Nowell-Smith (1948); Hart (1968:229)

[408] Strawson (1963)

[409] This is not what Strawson (1963) argues, even though one could certainly read an inherently normative argument in his influential paper.

[410] Strawson (1963)

persons who do not have what it takes to be responsible – such as children or the mentally incapacitated for example – we simply cannot do so to all persons all of the time as would be required if we took the thesis of determinism seriously. It would be highly questionable whether we could even properly conceive of a world in which such an attitude generally obtained given how far this would remove us from the world we know to exist and in which human reactive sentiments play a central and constitutive role. Thus it is ultimately our phenomenological understanding of what it is to be human in a world like ours that prevents us from taking the thesis of determinism too seriously. At the very least it keeps us from too rigorously questioning our practices of holding actors responsible because these practices are in a cognitive sense constitutive of what it is to be human in a world like ours. This comes down to saying that our *cognitive* understanding of responsibility is ultimately a solid enough foundation for our *normative* practice of holding actors deeply responsible for what they do or omit from doing. From this it is but a small step to say that in attributing responsibility to humans *in general,* our normative mechanism of status-function-imposition is *ultimately* upheld by the proper functioning our cognitive mechanism.

A second comparative advantage of an institutional over a conceptual understanding of responsibility is that the former can accommodate both responsibility for authority and individual responsibility within the confines of a single conceptual framework – that of institutional theory – whilst it is doubtful whether there can be a concept of responsibility that applies to both. The reason for this is easy to grasp now. If the *concept* of responsibility applies at all, the conditions under which it does will be so demanding that it will not apply to very much. An institutional understanding, on the other hand, is more adaptive to the conditions to which it applies, because in contradistinction to a conceptual approach, it allows the different mechanisms of status-function-imposition to work disjunctively. It is therefore also more adaptive to the particular conditions pertaining to the functioning of authorities. Thus while our efforts to arrive at a conceptual understanding of responsibility for authority have led to very little, a lot has been gained by an *indirect* approach; an approach in which responsibility is understood *via* and *in terms of* a general theory of institutions. This institutional approach is theoretically as powerful as a conceptual one because it goes to the same depth in revealing the micro-mechanisms constitutive of responsibility. Yet at the same time it is more viable in practice because of its more relaxed requirements with respect to how precisely these micro-mechanism relate to each other. Although we can now understand why an institutional approach to responsibility (in general) is superior over a conceptual one, and can also see why it is more conducive to the conditions applying to authoritarian governance in particular, we are still in the dark about the question what an institutional understanding of responsibility for authority would look like. I now aim to shed some light on this matter.

Beyond Agency Theory

Now that we know where to look for a theoretically consistent and a pragmatically viable understanding of responsibility for authority, we can proceed and map the terrain located. The problem, however, is that there will not be much to see there once we arrive. A *general* institutional perspective on responsibility, and on responsibility for authority more in particular, has not been developed or otherwise subscribed to in the literature on either authority or responsibility, nor in the rich variety of literatures that crowd the province of institutional theory broadly conceived. That there may not be much to map and describe should not surprise us though. The whole point of the effort undertaken in this chapter and the previous one was precisely to locate and map a terrain of which little is presently known. Thus apart from the broad empirical variety of *extant* institutional arrangements by which authoritative agents are constrained and empowered by their principals in everyday reality, and that can, of course, be empirically surveyed and explored in all the phenomenological detail we can imagine relevant, there are really not many theoretical landmarks that can help us to navigate through the terrain located. There is one notable exception, however. Interesting theorizing has been done in a field that goes by the telling name of *agency theory*. This theory, however, is highly controversial. It has been criticized as fatally one-sided in assumptions, explanatory focus and explanations.[411] A general institutional understanding of responsibility for authority, in contrast, has the ambition of offering a very broad explanatory focus while at the same time being permissive in its assumptions. Perhaps, then, we can see the outlines of a general institutional theory of responsibility for authority *through* the perceived shortcomings of agency theory or so is at least the working assumption of this paragraph.

Although agency theory is sometimes presented as a young theoretical perspective,[412] which was developed only after the separation of ownership and control in the modern firm was put on the academic agenda,[413] its historical roots go back quite a bit in time. Already the seventeenth century political theorist John Locke identified the topic as highly significant by conceiving of the relationship between a sovereign and its subjects in terms of an agency (social) contract.[414] Although more contemporary agency theory stretches from the very formal and rigorous theorizing of economic principal-agent theory on cooperative relations in general, on the one hand,[415] to a broad variety of more behavioural research on agency relations within organizations on the other,[416] there seems to have persisted something of a common

---

[411] Cf, Perrow (1986: 224 ff.)

[412] Cf. Jensen (1983); Moe (1984)

[413] Berle & Means (1932)

[414] Locke (1690/1965) For the interpretation that Locke's social contract theory is an agency contract, see: Hampton (1986)

[415] For a review see: Holmstrom & Tirole (1989)

[416] For a review see: Eisenhart (1989)

focus of attention ever since Locke's early identification of this field of interest.[417] This focus is on the relationship between an agent, on the one hand, and those *on behalf* of whom he acts – i.e. the principals – on the other.[418] There also seems to be something of a general problem or dependent variable for agency theory to address and explain. This is the general question how the principal-agent relationship can be shaped best.[419] This problem can be addressed from either the narrow perspective of maximizing the principal's interests, or from the broader perspective of optimising the welfare implications of the relationship as a whole.

The explanatory focus of agency theory also betrays a central assumption underlying the theory. Both principals and agents are assumed to be constrained by bounded rationality and to *generally* act in their self-interest. For the remainder assumptions and *explanandi* vary and are often interchangeable. Thus in many behaviourally inspired (i.e. positive) contemporary versions of the theory it is additionally assumed that the interests of agents and principals diverge and are even in conflict.[420] Even critics of the theory consider this assumption more or less inevitable because without diverging interests there would really be not much to explain or focus on in agency theory.[421] As a rule, these theories subsequently aim to explain the conditions in which the divergence of interests between agents and principals are most likely to result in damage to the principal, [422] and search for ways to reduce this damage.[423] Other agency theorists, in contrast, assume information asymmetry between principal and agent to be the central assumption underlying agency theorizing with the principal typically being unable to adequately assess the actions and performance of the agent.[424] In this view, interest divergence is but one explanatory variable with the explanatory focus not so much on maximizing the

---

[417] Jensen (1983); Eisenhart (1989:59 ff.)

[418] Eisenhart (1989:59), but see Jensen (1983:327) who explicitly points at the individual as the central unit of analysis. In my view, Jensen confuses the unit of analysis here with the assumption of methodological individualism underlying much economic theorizing. Cf. Elster (1983)

[419] This core problem echo's a combination of voluntaristic and deterministic assumptions underlying agency theory. The voluntarism is manifestly present in the view that agency relations are the result of (voluntary) principal-agent contracting. See: Alchian & Demsetz (1972) The determinism generally comes with evolutionary arguments. See: Alchian (1950); Jensen (1983). I leave the question whether a theory can rely on both assumptions simultaneously unanswered here.

[420] Eisenhart (1989) but see the dialogue on this assumption between Albanese et al. (1997) and Davis et al. (1997b) in reaction to Davis et al. (1997a)

[421] Cf. Davis et al. (1997b)

[422] Involving the so called "agency costs" comprising: " 1) the monitoring expenditures by the principal, the bonding expeditures by the agent, [and] 3) the residual loss" by the principal. Jensen & Meckling (1976/1996:107)

[423] Cf. Eisenhart (1989)

[424] Cf. Alchian & Demsetz (1972); Pratt & Zeckhauser (1985), Arrow (1985)

principal's interests but rather on the welfare implications of the principal-agent relationships as a whole.

Although agency theory thus embraces a variety of different theoretical perspectives and spans a number of different disciplines,[425] I will – for purposes of exposition – from hereon focus on positive agency theory and the issue of corporate governance – i.e. the agency relationship within firms between management and stockholders – more in particular. The reasons for this are threefold. First, the relationship between stockholders and management approaches something of an analytically pure authority relationship with stockholders almost completely surrendering their judgement over the way their property is used while at the same time delimitating the authority relationship by remaining residual claimants.[426] Second, the topic of corporate governance has had widespread theoretical and empirical attention ever since the separation of ownership and control within the modern firm was put on the academic agenda. Not only does there therefore exist a large body of relevant literature,[427] but the relationship between stockholders and management has become more or less paradigmatic for (positive) agency theory.[428] A third reason for this rather narrow focus involves the purpose of discussing agency theory in the first place. Since this is not to contribute to agency theorizing but rather to bring into perspective *through* agency theory the rough outlines of a general institutional theory of responsibility for authority, I can benefit from the sharper contrast that a somewhat narrow perspective provides without at the same time distorting the issue to the extent that fallacious arguments result, or so I presume.

In a nutshell, then, the agency theory perspective on corporate governance involves the question how the relationship between management and stockholders can be structured such that stockholder-value is maximized. Although this perspective therefore takes the narrow perspective referred to above, this seems to be the rule within the field of corporate governance.[429] Seen from this perspective there are roughly three factors that threaten the maximization of shareholder-value. A first factor involves interest divergence between managers and shareholders. Even if managers are not typically found with their hands in the company tilt they may and often do pursue projects that are beneficial to them but are harmful to shareholders. They may do so for reasons of prestige or because they hope to enhance their value on the labour market. A second factor involves information asymmetry between shareholders and managers. Managers are generally in a much better position than shareholders to gather and interpret crucial information on the firm's performance, simply because they are functionally on top of it. This problem would persist even if

---

[425] For political science see: Moe (1984); Ferejohn (1999), for sociology see: Eccles (1985); White (1985)

[426] Friedman (1973). For the methodological relevance of extreme cases see chapter III

[427] For a recent and rather substantial review see: Schleifer & Vishni (1997)

[428] Eisenhart (1989:59)

[429] Eisenhardt (1989:59); Schleifer & Vishny (1997) but see Jacobides & Croson (2001) for a counter argument

all information were freely available to shareholders. The reason for this is that information relevant to judging performance is typically difficult to attain because of difficulties inherent to task programming and performance measurement.[430] A third factor relates to the first and involves diverging attitudes to risk. While stockholders can purposefully manage risk by diversifying their portfolio, managers typically have only one job and are therefore more vulnerable to firm-level outcome variance. Assuming that managers can only partly determine firm-level outcomes they will generally be more risk-aversive than stockholders are. In very broad strokes, then, these three factors set the stage for agency problems to occur and for agency theorists to find for ways to cope with them.[431]

In reply to this challenge, agency theorists commonly propose two kinds of remedies.[432] The first involves a co-alignment of shareholder's and manager's interests to alleviate the potential divergence of interests inherent to the shareholder-manager relationship. Thus it is nowadays quite common, for example, for executive compensation to be at least partly based on firm-level outcomes by means of bonuses, stock ownership or stock-option-plans for managers. But although performance based executive compensation may certainly help to alleviate agency problems in corporate governance, they can at best be only part of the solution. The greater risk aversion of managers makes outcome-based compensation relatively unattractive for them (and hence expensive),[433] whilst even the most effective performance contracts cannot dissolve all divergence of interests between managers and stockholders. If my earlier conclusion that it is a central function of authoritarian governance in modern society to deal with problems of differential ultimate valuation, holds true for firms as well, we may expect there to be definite limits to the extent that the interests of agents and principals can be aligned in authority relationships precisely because one of the main reasons for authorities to be instituted in the first place is to cope with problems of differential (ultimate) valuation.[434] A second remedy involves the design and application of monitoring and information systems by which stockholders can track and sanction the behaviour and performance of management. An important monitoring instrument in the modern corporation is the board of directors.[435] While managers make and execute day-to-day decisions, the board reviews these decisions on behalf of shareholder interests and can commonly do so in

---

[430] Cf. Alchian & Demsetz (1972/1996); Eisenhart (1985, 1989), Anderson (1985)

[431] Although agency problems can take a variety of forms, the problems of 'moral hazard' and 'adverse selection' are perceived to constitute the core of agency theory. Cf. Barney & Hesterly (1996:125). The former involves "hidden actions" and is primarily associated with interest divergence while he latter pertains to "hidden information" and is mainly associated with information asymmetry. Cf. Arrow (1985)

[432] Eisenhart (1989)

[433] Eisenhart (1989:62)

[434] See chapter III

[435] Fama & Jensen (1983)

the light of independently verified information.[436] There are limits, however, to what monitoring can achieve in coping with agency problems. First, monitoring does not go without cost and its utility is thereby constrained by its cost-benefit ratio. I have already pointed out, secondly, that problems of performance measurement and control can seriously inhibit effective monitoring. Things get even worse when principals are really in no position to judge the nature of performance, as is often the case in agency relationships with professionals.[437] Third, monitoring can have perverse effects with managers harming shareholder interests by pursuing behaviour required by invalid performance indicators.[438] Excessive monitoring, finally, may even corrupt managerial behaviour as pervasive distrust and an obsession with control can produce precisely the kind of behaviour it aims to combat.[439] The conclusion, in short, is that *pragmatic* value of (positive) agency theory is fairly limited even if only because the remedies it proposes are severely constrained by the basic relations constitutive of the theory itself.

But it is not just the pragmatic value of agency theory that has been seriously questioned. The reason why I have discussed agency theory in the first place is that the theory as a whole has been severely criticized on several fronts. According to one particularly fierce critic agency theory "does not have a clear problem to which it offers a solution" whilst at the same time it "is hardly subject to empirical test since it rarely tries to explain actual events or make predictions."[440] This critique is more or less in line with my earlier observation that assumptions and explanations vary across the field and are often interchangeable. Moreover, some serious inconsistencies in the theory's basic conceptualisation of the firm – i.e. as a "nexus for contracting relations"[441] – have led another critic to characterize agency as a mere "nexus of metaphors".[442] The most fundamental point in virtually all critique of agency theory involves the behavioural assumptions underlying the theory. While there can be little objection to agency theory's assumption that actors are constrained by bounded rationality, the conjoining assumption that actors are generally self-interest seeking or even opportunistic has met with considerable resistance.[443] Agency theory suffers the

---

[436] Independently gathered and verified information is supplied for by accountants. This is why Jensen (1983) sees accountancy as an inherent part of the structure of the modern firm.
[437] Cf. Sharma (1997). The relationship between professionals and clients is often sometimes seen as a pure agency relationship. See: Pratt & Zeckhauser (1985:2) Note that knowledge or information asymmetry *as such* involves a problem of *theoretical* rather than *practical* authority. See chapter I
[438] Jacobides & Croson (2001)
[439] Cf. Goshal & Moran (1996)
[440] Perrow (1986:224)
[441] Jensen & Meckling (1976/1996:109
[442] Campbell (1997:361)
[443] On the assumptions that agency theory shares with transaction cost economics see: Williamson (1988/1996) For critique on these assumptions Perrow (1986:231 ff.) and Goshal & Moran (1996) respectively.

same fate in this respect as economic (equilibrium) theorizing in general.[444] The gist of this critique encompasses two separate points. First, it is obvious that the assumption of *general* self-interest seeking – either with or without guile – is at odds with the facts of human behaviour as we know it to be. Even casual observation sufficiently demonstrates that humans display a wide variety of behaviours and are commonly prone to other-regarding motivations, even in a market context. To start off from such seriously skewed general behavioural assumptions is therefore bound to lead to invalid conclusions and inadequate explanations. But even if it is conceded that the general assumptions of self-interest seeking or opportunism are unrealistic, secondly, they may nevertheless become self-fulfilling and real if they are commonly perceived as real and consequently acted upon.[445] Given its assumptions, explanatory focus, and explanations, it is no exaggeration to say that underlying agency theory is a pre-occupation with rational control.[446] As was pointed out above, however, excessive control in authority relationships may very well lead to precisely the kind of behaviour calling for control in the first place. There exists alarming empirical evidence that the use of rational control is negatively related to the (intrinsic) motivation and commitment of agents while at the same time being undermining of trust and trustworthiness in agency relationships.[447] Even mild control measures may 'crowd out' processes of spontaneous coordination and compliance by signalling caution and distrust.[448] It has been concluded, therefore, that agency theory is a "dangerous theory".[449] For despite of its inveracity it may shape our world after its own image nevertheless. Agency theory, then, is at the very least a controversial theory.

Despite its controversial nature we should not put aside agency theory too easily, though. It is perhaps wiser to give agency theory some credit and to see it as a fundamentally *incomplete* rather than an invalid theory of agency relationships. An interesting attempt to complement agency theory in this respect involves what has become known as a stewardship theory.[450] Not surprisingly, stewardship theory takes issue first and foremost with the behavioural assumptions underlying agency theory. Instead of seeing human motivation as fixed and generally self-regarding, stewardship

---

[444] Cf. Pettit (1995a)

[445] Perrow (1986:235); Goshal & Moran (1996:21-27) A *general* institutional theory can make us understand why this is so. If actors act on the general *regulative* presumption of opportunism – which as Williamson (1985) argues is prudent because one never knows which actor is and which is not – then effectively the regulative mechanism of status-function-imposition *makes* the cognitive mechanism follow its trail. A *general* institutional theory gives us a fine-grained explanation of both self-fulfilling and self-denying prophecies in general, involving both regulative and normative mechanisms besides the cognitive mechanisms of status-function-imposition.

[446] White (1985)

[447] Enzle & Anderson (1993); Baker, Jensen & Murphy (1988); Peel (1998)

[448] Pettit (1996)

[449] Perrow (1986:235)

[450] Davis et al. (1997a)

theory assumes human motivation to be variable,[451] and both motivation and action susceptible to social and situational influences.[452] Consequentially, stewardship theory aims to offer more comprehensive and therefore more complete explanations of agency relationships encompassing both dispositional and situational explanatory variables. Thus managers, for example, who are intrinsically motivated by higher order needs and who work in an environment where they can maximally develop their responsibilities are considered more likely to behave as stewards in agency relationships than those who are primarily motivated by extrinsic lower needs and work in a context of monitoring and control.[453] Even if we grant that stewardship theory offers a more complete and veracious perspective on authority relationships than agency theory, it is important to note that it does so only at the cost of parsimony. Compared to agency theory, stewardship theory's assumptions, explanatory focus and explanations are less precise, less focussed, more complex and less determinate respectively. Although it is clear that veracity trumps all of these considerations in the end, it is also clear that the challenge for stewardship theory is to overcome its present status of merely being a broad critique of agency theory. It could do this by sharpening its theoretical explanatory focus and offering a more fine-grained understanding of agency relationships in general and authority relationships more in particular. A particularly interesting way by which this could be done is by taking the perspective that a general institutional theory of responsibility for authority offers.

Seen from this perspective the main problem with agency theory is its *exclusive* focus on a *regulative* understanding of responsibility for authority. Its ultimate aim to optimally design agency relationships through the means of monitoring and interest alignment can be interpreted as attempts to build the *mechanism* of responsibility into authority relationships.[454] Thus normative expectations and accountability structures are fitted together in such a way that the resulting mechanism of responsibility is optimally socially efficacious; in this case in optimising the principal's interests. Moreover, agency theory's explicit focus on information asymmetry between principals and agents in the advantage of the latter suggests the problematic functioning of our cognitive understanding of responsibility in this respect. This coheres with my earlier discussion of the problem of ascription according to which ascribing a certain state of affairs to the intentional acts or omissions of an authoritative agent is often highly problematic. One major reason brought forward was that we typically do not have a natural bird's eye perspective from which we can entangle complex causal relations at the often aggregate level at which authoritative agents operate. Although agency theory's explanations for the problem of ascription are clearly more fine-grained and concrete than my own observations in this regard,

---

[451] As was already suggested by Perrow (1986)

[452] Davis et al. (1997a:32 ff.)

[453] Davis et al. (1997a)

[454] Seen from the agency theory perspective again, an important way in which this mechanism subsequently functions is through the market for corporate control.

they are on the whole consistent with and supportive of them. Consequentially, agency theory's aim to match authority relationships with appropriate monitoring arrangements can be interpreted as the purposeful construction of artificial (bird's-eye) viewpoints from which the problem of ascription can be coped with.

Despite these rather obvious parallels between agency theory and a general institutional perspective on responsibility for authority there are also sharp contrasts between the two. These are, in my view, mainly attributable to one fundamental difference underlying the two theoretical perspectives. This difference pertains to agency theory's evident disregard for the dimension of *normative* involvement underlying the different mechanisms of status-function-imposition constitutive of *any* understanding of responsibility, and thus also of the regulative one that forms the heart of agency theory. Although it was argued in the previous paragraph that our cognitive, regulative and normative understandings of responsibility may each function as effective mechanisms of status-function-imposition on their own, it was also argued that each of these understandings is necessarily – yet differentially – embedded in an underlying dimension of human normative commitment. It was argued that even understanding a certain state of affairs to be the result of the *intentional* acts or omissions of an authoritative agent assumes a certain degree of normative human commitment, for to understand (human) intentions is already to be normatively involved to a certain extent. If this argument holds ground, then one of the founding fathers of modern (positive) agency theory has seriously put us on the wrong foot by insisting that a theory on agency relationships does not require us to make sense of human intentions.[455] On closer inspection, this claim even seems absurd. Without the assumption of differential intentions between principals and agents there would really not be much for agency theory to theorize about in the first place.

Even if we could somehow make sense of agency theory despite of its neglect of the dimension of normative involvement underlying the core mechanism of this theory – and which can be done by *unconsciously* subscribing to the behavioural assumption of general self-interestness [456] – the resulting perspective is indeed as myopic as agency theory is often accused of by its critics. For the general blindness to this underlying dimension makes it impossible to see both a cognitive and an inherently normative understanding of responsibility for authority as *extensions* – in opposite directions – of the regulative understanding forming the heart of the theory. The agency theoretical perspective therefore prevents us from seeing how these three basic understandings of responsibility can be *complementary* in the sense that the problematic functioning of one can sometimes be compensated for by the proper function of another. Given that much of contemporary agency theorizing is effectively about the problematic functioning of our cognitive understanding of

---

[455] Cf. Jensen (1983:322)

[456] I say unconsciously, because consciously subscribing to this assumption would necessarily involve referring to human intentions and would thereby involve an underlying dimension of human normative commitment.

responsibility in authority relationships this, is a potentially very costly omission. For in exclusively relying on our regulative understanding of responsibility in understanding and shaping authority relationships, agency theory makes these relationships less resilient and therefore less robust than they could have been had a normative understanding of responsibility been taken into account simultaneously. Our normative understanding of responsibility could complement our regulative one by letting the regulative mechanism of status-function-imposition of which this understanding consists be supported by our normative mechanism of status-function-imposition in cases were the former is known not to function properly; e.g. in contexts where rational control breads a greater need for control.

But it is not just in a pragmatic sense that a general institutional theory of responsibility for authority is superior over agency theory. The same holds true theoretically. A general institutional theory of responsibility for authority would be more realistic in its assumptions – allowing human agents to have different motivations and dispositions in different contexts[457] – while being both more focussed and comprehensive in its understanding of relevant phenomena at the same time. It would be more focussed because it could aim at understanding agency theory in terms of the institution of responsibility. It would be more comprehensive because it could account for all of the different micro-social mechanisms constitutive of this institution. A general institutional theory of responsibility for authority would pay attention in particular to how a normative understanding of responsibility for authority may complement the exclusively regulative understanding constitutive of agency theory. Before bringing this chapter to its final conclusion let me sketch in broad explorative strokes how this may be so. I make this sketch with the ambition of merely identifying a potentially fruitful avenue that may lead our thinking on responsibility for authority beyond the narrow regulative confines of agency theory. To actually discover where this route will lead us is beyond the scope of this study.

Trustworthiness and a Normative Understanding of Responsibility for Authority.

In order to demonstrate how a normative understanding of responsibility can complement the regulative understanding that forms the heart of agency theory in constraining authority relationships, I return to the topic of corporate governance that has served as a paradigm case throughout this chapter. The field of corporate governance is particularly interesting in this regard because agency theorists have consistently maintained that there is nothing inherently normative to understand about corporate governance.[458] This becomes most evident in agency theory's conceptualisation of the managerial agent as a self-interested utility maximiser who's potential opportunism must be combated with either monitoring or interest

---

[457] Conform the requirements of a stewardship theory of authority relationships for example.
[458] Cf. Jensen (1983); Demsetz (1995:28)

alignment.[459] Moreover, agency theorists have argued that there is no such thing as authority within the firm. In one of agency theory's founding and most widely cited texts it is explicitly held that "The firm…has no power of fiat, no authority, no disciplinary action any different….from ordinary market contracting between two people".[460] The firm, in this view, is but a "legal fiction"; a mere "nexus for contracting relations".[461] Thus despite the significant surrender of judgement that empirically characterizes the position of both shareholders and employees within the firm, the relationship between them and the firm's management is seen as an ordinary contractual one. For like is the case in a market transaction, both shareholders and employees can vote with their feet if they disagree with management's decisions. My hunch is that if a normative understanding of responsibility can be shown to be unavoidable even in this presumed market-like context, then surely it is inevitable in all (proper) authority relationships.[462]

With the advantage of theoretical hindsight, the core difficulty in agency theory's contractarian perspective is understanding why agency relationships exist at all and are pervasive in organizational rather than market governance.[463] If agents are fully self-regarding and agency relationships purely contractual there is really no reason to demarcate them from ordinary market transactions. The only reason that makes sense is the presumed information asymmetry between principals and agents, but if this prevents agency relations from being ordinary market transactions then surely it also prevents these relationships from being understood in contractual terms. The difference, agency theorists insist, is in the *ex-post* monitoring of the contractual relations. We have already seen, however, that monitoring and control can only take us so far and are particularly ineffective in contexts where information asymmetry is persistently high, which is typically the case in corporate governance. Again, it are the basic propositions of the theory itself that severely constrain finding solutions to the problems they identify. It does not seem to make much theoretical sense, therefore, to see manager-shareholder relations in contractarian terms because given their opportunistic disposition and a context of relatively ineffective monitoring managers

---

[459] Compare Fama's (1980:288) more symmetrical interpretation: "The firm is viewed as a set of contracts among *factors of production*, with each *factor* motivated by its self-interest."(Emphasis mine).

[460] Alchian & Demsetz (1972/1996:75)

[461] Jensen & Meckling (1976/1996:109)

[462] I do not take issue with agency theory's denial of authority here as I address that topic in a separate paper (Van Oosterhout, forthcoming). In this paper I argue that, ironically enough, authority within the firm already comes into existence by functioning as "a nexus for contracting relations". Campbell (1997: 360) anticipates some of the arguments I make in comparing agency theory with the *empirical* legal institution of agency in which executives enter into contracts with third parties on behalf of the company – thereby binding all organizational constituents.

[463] For a similar line of reasoning that ends up with agency theory's denial of the firm, see: Campbell (1997)

will rob blind their principals and get away with it too.[464] There is, after all, no *contractarian* reason for managers to keep to the agreement once it has been made.[465] Some notion of *fiduciary duty* constraining the agency relationship between shareholders and management therefore seems inevitable. Without such duties it seems unlikely that the separation of ownership and control within firms would have survived for so long.

That fiduciary duties are important in agency relationships is supported by the fact that they are actually quite common in modern corporate governance systems.[466] Most systems of corporate law in modern economies have provisions pertaining to disclosure rules, insider trading as well to more general open-ended obligations of loyalty towards the shareholder.[467] Although agency theorists cannot be accused of denying these facts, they claim instead that fiduciary duties can be understood in contractarianist terms.[468] They are commonly interpreted as either standard contracts that have been institutionalised and that subsequently have found their way into law, or as relatively cheap substitutes for agency contracts that would otherwise be too expensive to bargain and draw up because of the scope and scale of relevant but unforeseen contingencies.[469] From the agency theoretical perspective fiduciary duties "are the same sort of obligations, derived and enforced in the same way, as other contractual undertakings..".[470] We can already see, however, why invoking a contractarian logic in order to cope with contracting problems in agency relations is not very persuasive in a general sense, so I will save my words for the particulars of the case at hand. These, are that (a) it has never been shown that fiduciary duties have actually evolved in the way that agency theorists contend,[471] (b) despite of their *ex post* informational advantages agents are not very likely to voluntarily agree *ex ante* to the open-ended nature typical of many fiduciary duties,[472] and (c) such a view does not explain that while principal-agent contracts are often not empirical – as one would *eventually* expect – but only metaphorical,[473] fiduciary duties *are* commonly real life phenomena. The conclusion seems plausible, then, that both the actual existence and

---

[464] That there more truth in this proposition than we would like is feared by many. Schleiffer &Vishny (1997) estimate agency costs in publicly held firms to be substantial. An interesting proxy for these costs is suggested by Clark (1985:). Assuming that there is an effective market for corporate control these costs are assumed more or less equal to the costs of mergers or take-overs.

[465] Cf. Van Willigenburg (forthcoming, b)

[466] Schleifer & Vishny (1997) even identify legal protection of shareholder interests as one of the critical success factors in modern corporate governance.

[467] Clark (1985); Schleifer & Vishny (1997)

[468] Campbell (1997: 354)

[469] Clark (1985)

[470] Easterbrook & Fishel (1993:427)

[471] Clark (1985)

[472] That this may not be true with respect to duties of disclosure in democratic governance is argued by Ferejohn (1999).

[473] Campbell (1997)

the binding nature of fiduciary duties cannot be reduced to the contractarian logic underlying agency theorizing. Instead of deriving from actual or virtual contracting, fiduciary duties rather seem to *presumed* by contracting practices in principal-agent relationships. If anything, they are what make such practices efficacious and thereby socially prevalent. The way they do so is, by now, everything but mysterious. It is an insight widely subscribed to that contracting practices in general pre-suppose an embeddedness[474] in some social-normative – in modern society typically legal – framework without which there could be nothing like the practices we now refer to as contracting.[475] Contractarianism, in sum, is not a very fruitful conceptual path leading to an understanding the apparent normative aspects of principal-agent relationships. Arguably it is not even a particularly interesting path for understanding such relationships period.

A more promising route to understanding responsibility for authority, in my view, leads via the concept of trust.[476] I have already argued that trust and the social mechanisms that are related to it are crucial in understanding why people accept there to be authoritative agents in the first place, and I have pointed out in particular how our regulative understanding of responsibility can help make trusting authorities more resilient in practice. I have argued on a more general level that the 'pure' mechanism of trust is vulnerable to many external disturbances,[477] and that it is therefore necessary to look at what I have called 'flanking social mechanisms' that can help make the practice of trusting more resilient. I now want to ask attention to the converse relationship, that is, to the question how trust and its flanking social mechanisms may help to make our regulative understanding of responsibility for authority both practically more resilient and theoretically more complete. Although I realize that there is, by now, an ocean of literature on trust,[478] I will pay attention only to the *normative* mechanisms involved with trust. I do so because my aim here is to understand how a normative understanding of responsibility for authority may

---

[474] Granovetter (1985)

[475] Cf. McNeil (1974)

[476] This is also Dunn's (1988) reading of Locke. This in contradistinction to Hampton's (1986) interpretation referred to above which resonates better with a contractarian and agency perspective.

[477] Again, I conceptualize the *pure* mechanism of trust as a course of action in between the Scylla of distrust and the Charibdis of betrayal. I realize that my conceptualisation of trust may be unconventional, and that the mechanism I refer to is – from the perspective of convention at least – perhaps more properly addressed to as 'reliance'. I persist in my own view, however, in order to keep an open mind about *additional* mechanisms that may be helpful in understanding actual trust. Although, perhaps, unconventional I am not alone in this position. See: Blackburn (1998:37)

[478] An extensive and up to date review of this literature is given by Nooteboom (forthcoming). This paper is particularly interesting in this context because it aims to position trust in contradistinction to a contractarian perspective. A widely cited collection of essays is edited by Gambetta (1988). A current counterpart of this collection, with a particular focus on trust in governance, is edited by Braithwaite & Levi (1998)

complement the regulative one in understanding and shaping authority relationships. The issue, then, is whether there can be something as normative trust in authoritarian governance and how such trust relates to our normative understanding of responsibility.

In its most comprehensive understanding trust "is a four-place predicate"[479]: the trustor (X) trusts the trustee (Y) with respect to certain acts, omissions or the consequences thereof (A), under certain conditions (C). Different kinds of kinds of trust, that is, different combinations of trust and flanking social mechanisms can already be distinguished on the basis of this simple schema. A relatively common distinction is between personal and institutional trust.[480] Where the former is primarily based on personal mechanisms involving the attributes of X and Y as well as the relationship between X and Y, the latter is based on the conditions (C) in which X and Y have a trusting relationship and which make X presume that he can rely on Y. An example of the former is when I trust my brother to pay me back the money that he owns me because (a) I know he is not the kind of person that would default on his debts, and (b) our relationship is such that something like that simply doesn't happen. I trust the bank with my deposit, in contrast, because I know there exist a number of contextual mechanisms – both internal and external to it – that (sufficiently) assure that I get my money back even if some particular employee does not want to give it to me for personal reasons.

Despite this conceptual difference, both personal and institutional trust have inherently normative aspects, that is, both can function at least partly via normative mechanisms of status-function-imposition. That this is so becomes manifest when trust is betrayed and is relatively self-evident for personal trust. If my brother betrays my trust even an uninvolved third party will see sufficient reason for me to resent this betrayal. Though perhaps not so obvious this may also be true of institutional trust. Not only do we often speak of 'due process', the 'Rechtstaat' and 'democracy' in highly evaluative terms, but we are particularly prone to use such a normatively laden vocabulary in cases where trust in those institutions is betrayed.[481] We should therefore have no difficulties in seeing trust in authoritative agents from an inherently normative perspective. But to show that trust in authoritative agency may have inherently normative features is not enough. Some agency theorists will arguably even agree that trust in authoritative agency may have a normative dimension to it while at the same time insisting that self-regarding agents will abuse this trust regardless of how deeply committed we are. It is not enough, then, to identify normative features of trust from an external, *third*-person's-perspective. For unless we can show that trust has inherently normative features from the *first* person' perspective of the authoritative agent as well, we are bound to accept that such trust will lead to betrayal in the absence of contextual mechanisms that prevent it. To put it in other words: if

---

[479] Nooteboom (forthcoming)

[480] Zucker (1986); Brennan (1998); Pettit (1998)

[481] An excellent example of how this may be so is given by Martha Nussbaum (1985) in her chapter on the betrayal of convention.

we cannot show that trustworthiness is something that is inherently valuable for authorities, the conclusion seems more or less inevitable that there can be nothing but institutional trust in authoritative governance.[482] Such a conclusion would bring us a step closer to the narrow confines of agency theory again.

There are roughly three reasons why I think why such a step will not be necessary. These three reasons represent three basic mechanisms that make trustworthiness valuable for authoritative agents and that may together constitute resilient normative trust in authoritarian governance. Only in the first mechanism is trustworthiness inherently valuable. This is the mechanism of ultimate valuation. To be a head of state or the CEO of a publicly held company is for many people not just another job but a great honour and privilege that is accompanied by an equally great sense of duty and responsibility. Such people commonly take great pride in their position and are often intrinsically motivated to serve the interests of those on behalf they exercise their authority. That positions of authority may indeed have such intrinsic value and that there actually are people who are so committed is easily obscured by unconscious adherence to the behavioural assumption of *general* self-interestness. That trustworthiness may indeed be intrinsically valuable in positions of authority is demonstrated by the fact that authorities sometimes resign from office when they have failed to meet their principal's expectations even if their position has not been questioned by anyone. By doing so, they signal that trustworthiness is of central importance and that, in their view, it is a necessary condition for remaining in office. Contrary to popular myth it is not even unlikely that such intrinsically motivated people will end up in positions of authority in the first place. Because of their intrinsic valuation they are typically prepared to sacrifice a lot in order to get into office.

For those to whom reliance on this first mechanism may seem hopelessly naïve, there is a second mechanism that can make trustworthiness valuable even for authoritative agents who do not value it inherently. Given that their principals and the public at large *will* value the trustworthiness of their authorities, such agents will be susceptible to their attitude, that is, under conditions were positions of authority can be contested in some minimal way.[483] In this context the actual trust put in authoritative agents may give those authorities additional and sometimes even sufficient incentive to be trustworthy. This is particularly so if such trust is communicated to third parties. Pettit names this mechanism "trust-responsiveness" because trustworthiness may thus be elicited indirectly merely by the act of trusting it self.[484] The conditions under which this mechanism will work are not excessively demanding. Next to minimal translucency and some basic contestability of authority,

---

[482] Which is argued by Hardin (1998)

[483] Minimally this would require some freedom of speech in order to for reputation mechanisms to function properly.

[484] Pettit (1995b)

this mechanism is bound to work like all other reputation mechanisms, that is, without requiring intentional intervention or prohibitive operating costs.[485]

A third and here final mechanism involves a combination of personal and institutional trust.[486] Given a relatively strong form of contestability of positions of authority through elections or some other sort of institutional screening mechanism, it is not unlikely that trustworthiness will be a central variable that is screened for, and that trustworthy actors thereby have a relatively greater chance of getting in to authoritative office than actors that are perceived of as lacking trustworthiness. That trustworthiness may be an important selection variable is because of the large discretion that is typically afforded to authorities. It is often difficult to rely on the policy platforms in selecting candidates because these (a) may not discriminate *ex ante*, (b) may already be outdated when candidates assume office and (c) because authoritarian governance may very well be prevalent in general where a large degree of discretionary intervention may be socially efficacious and desirable (as is the case in context of persistent normative conflict). That trustworthy candidates have a higher chance of being selected is because principals will arguably be able to discern trustworthiness better than any random selection would,[487] and because trustworthy actors may, *precisely because* of their trustworthiness, be relatively more successful in aspects that are relevant for the principal's choice. Thus this mechanisms combines the dispositional variable of trustworthiness with institutional features that may enhance our trust in authorities by selecting for this dispositional variable.

The question remains, finally, how the normative trusts that results from the conjunction of these mechanisms relates to our normative understanding of responsibility and can thus become a constituent part of an inherently normative understanding of responsibility for authority. Without the ambition to be complete, let me point out two interesting analogues between the two. First, there is a striking resemblance between trustworthiness of authorities and the virtue of responsibility in the sense that both are inherently desirable. This becomes manifest when they are questioned. In the same way that there is no greater insult to an individual human being than openly questioning his status as a responsible human actor,[488] there seems to be nothing more offensive to an authoritative agent than to publicly question his trustworthiness. Trustworthiness may therefore be for authoritative agents what responsibility is for individuals: i.e. a central constitutive virtue. A second analogy is an extension of this point. In the same way that the presumption of responsibility is a necessary condition for individual human agents to take part in ordinary human

---

[485] This is mechanism is a paradigmatic example of what Pettit (1993) has earlier characterized as intangible hand. Contrary to "Iron Hand" of purposive intervention this mechanism relies on non-intentional sanctioning like Adam Smith's (1993) "invisible hand". But unlike the latter it does not function via the consequences of actions but rather via the attitudes of actors.

[486] Much of my discussion of this combined mechanism I derive from Brennan (1998) yet my interpretation and modifications would perhaps not be subscribed to by him.

[487] Brennan (1998:203 ff.)

[488] I thank Theo van Willigenburg for this example.

interaction, the presumption of trustworthiness is a necessary condition for the acceptance of authoritative agency in human society. Note that in both cases it suffices that this presumption is made regardless of whether the person in question actually has what it takes to be responsible or the authoritative agent at stake is actually trustworthy. As argued, this condition is more difficult to satisfy for authoritative agents than for individual human beings. Whereas the presumption of individual human responsibility is more or less the default presumption due to the deep normative commitment we have toward making such a presumption, there exists no such commitment to presuming authoritative agents to be trustworthy. It is sometimes argued that it should rather be the other way around. A default distrust or "eternal vigilance" towards authoritative agents is often held to be "the best way to keep government agents on track".[489] That the presumption of trustworthiness may be very difficult to earn for authoritative agents underscores its central importance, for authoritative agents commonly take great effort establishing at least a perception of trustworthiness. Having shown, then, how it is that a normative understanding of responsibility may complement the regulative one that is at present dominantly relied upon in understanding responsibility for authority, and having sketched the rough outlines of what a general institutional understanding of responsibility for authority may possible look like, I conclude with rehearsing the major line of argument of this chapter and the former.

Conclusions

At the beginning of the previous chapter, we started of with the conclusion that the ways to incorporate the quest for legitimacy within our understanding of authority were now definitely closed in one hand, and the mere promise that the idea of responsibility for authority could provide a pragmatic solution to our predicament in the other. The viability of this promise, however, was dependent on the theoretical question whether there could be something like responsibility for authority. To answer this question I have first surveyed notions of responsibility as these are implied in some of the most central practices and institutions in modern society and have subsequently reduced these notions to three basic understandings of responsibility, by identifying a dimension of normative involvement underlying these notions. These basic understandings involved a cognitive, a regulative and a normative understanding of responsibility respectively, with the first of these understandings assuming the highest normative involvement and the last one the least. The conjunction of basic requirements following from each of these basic understandings resulted in a conceptualisation of responsibility that was robust over both extensional and intensional strategies of conceptualisation. After subsequently having confronted this conceptualisation of responsibility with the additional demands that understanding responsibility for authority may require, the previous

---

[489] Pettit (1998:309 ff.) argues, I must say rather unconvincingly, that such default distrust of authorities is compatible with a reliance on the trust responsiveness of those in power.

chapter concluded that these requirements were not only too demanding for understanding responsibility for authority, but also for understanding responsibility *in general*. Several formidable problems were identified that make a conceptual understanding of responsibility apply in exceptional conditions at best. It is clear that this is at odds with the facts as we know them to be. Even casual observation will demonstrate that the practice of responsibility is ubiquitous in modern society.

In order to resolve this paradox the presumption that our understanding of responsibility must be of a conceptual nature was let go of in this chapter and an alternative was explored in some detail. This alternative consisted of a *general* institutional understanding of responsibility. In contradistinction to a conceptual understanding of responsibility according to which the requirements pertaining to the three basic understandings of responsibility must be fulfilled conjunctively in order for this concept to apply, an institutional understanding allows these requirements to be fulfilled disjunctively because it sees each of these understandings as *alternative* mechanisms of status-function-imposition. It was argued that an institutional understanding of responsibility was both theoretically and pragmatically superior over a conceptual one. It is theoretically superior because it can incorporate both responsibility for authority and individual human responsibility within the single conceptual framework of a general institutional theory, while it is not very likely that this is or can be made true of a conceptual understanding of responsibility. It is pragmatically superior because it explains how the institution of responsibility for authority can be made more resilient in practice by exploiting the complementary properties of different mechanisms of status-function imposition in different contexts.

In order to sketch the rough outlines of what a general institutional theory of responsibility for authority may look like, a more detailed analysis was made of how the regulative (mechanism) of responsibility that forms the heart of an agency theoretical perspective on authority relationships may be complemented by other institutional mechanisms. It was explained that agency relations in general, and authority relations in particular, are often troubled by the problematic functioning of our cognitive understanding of responsibility because there commonly is a large information asymmetry between agents and their principals. It is clear that this is commonly in the advantage of the former. In the context of an expected divergence of interest between principals and their agents – and it must be remembered that authoritative governance is often instituted precisely to deal with interest divergence and problems of differential ultimate valuation – this may not only lead to the interests of the principals being harmed, but also to significant deadweight (welfare) loss for all. It was argued that agency theoretical solutions to this 'agency problem' were either constrained by the basic relations of the theory itself, or were relying excessively on rational control. Since the latter may be particularly ineffective in a context of large and persistent information asymmetry – as is commonly the case in authority relationships – such reliance is often counterproductive. It was explained how a normative understanding of responsibility may help to overcome such

problems. The notion of trust was identified in particular as a potentially significant constituent part of an inherently normative understanding of responsibility for authority. It was explained how both responsibility and trustworthiness are central constitutive virtues of human and authoritative agency respectively. The underlying promise is that trust is or may become for authoritative governance what responsibility already is for the individual human subject.

VI

Coda

Introduction

I conclude this study by formulating the central theses that have emerged from the five chapters of which this study consists. These theses not only recapitulate the major findings of each chapter, but taken in conjunction they also constitute the skeleton of an argument against any unified and unambiguous theory of authority. More precisely, the major line of argument of this study rules out an understanding of the nature of authority that *theoretically* incorporates an answer to the question of legitimacy within that understanding. Each of the following thirteen theses of which this argument consists is briefly elaborated upon by recapitulating the arguments underlying them. The thirteenth thesis both amplifies the perspective on authority that has emerged from this study, and calls attention to the ultimately pragmatic way of coping with the quest for legitimacy that authoritarian governance confronts us with.

1
Although authoritarian governance is clearly in conflict with the fundamental modern demands of rationality, autonomy and individual responsibility, it is widespread in modern society and is even more ubiquitous today than it ever was in pre-modern society.

This thesis makes explicit the basic problem constituting the raison d'être of this study. In my view, it involves one of the most neglected and understudied paradoxes of modern society. In pre-modern society authoritarian governance was well embedded in a conjunction of historical conditions – both of an empirical and theoretical nature – that made it more or less the natural form of governance in human association. Transcendental metaphysics, first, made authority natural in a direct sense by placing it within the dominant holistic and hierarchical worldview, and indirectly by attributing to those *in* authority some kind of privileged access to the world behind appearances. Tradition, second, not only constituted a direct foundation for authority, but also played a more complementary role in connecting the abstract and intangible world of transcendental metaphysics with the day-to-day reality of the

masses. Moreover, tradition provided efficient and sustainable social coordination mechanisms in a society where knowledge and intellectual capacities were very unevenly distributed throughout the population. Non-trivial differences in power between those in command and those expected to obey, finally, gave authorities the capacity to sanction and coerce into obedience those who lacked the motivation to comply naturally. This more or less closed the circle as it rendered the institution of authority not only natural, but also resilient enough survive the inevitable occurrence of motivational divergence within human association.

The process of modernization, however, gradually dissolved these pre-modern foundations of authoritarian governance. First, a conjunction of a renewed interest for the concrete and particular and a methodical approach to the study of observable reality gradually undermined the hierarchical picture metaphysics had painted of the natural and social order. The new scientific worldview steadily displaced the metaphysical one because no society could afford to ignore the pragmatic successes that the advance of the scientific worldview had in store. The disenchantment of the world that characterized the process of modernization, second, also made the holistic and hierarchical *normative* universe of pre-modern society lose ground in favour of a more egalitarian normative space that was in line with man's newly acquired epistemological status as the 'seat of reason'. Within that space, reason, autonomy and individual responsibility became the three central coordinates. Authority is in deep conflict with each of these three modern values because it involves the surrender of (individual) judgement, while each of these three values requires that each man judges on his own account. Third, the economic, social and political changes triggered by these developments, and the de facto advance of the market in particular, demonstrated that there was a viable, more egalitarian alternative to the static hierarchical order of pre-modern society. Together these developments undermined the unquestioning obedience that is characteristic for authority, and led to the birth of the so-called question of legitimacy, that is, the presumption that authority is to be questioned before it is obeyed.

In spite of this presumption, however, authoritarian governance is widespread in modern society and is arguably more omnipresent today than it ever was in pre-modern society. The (nation-) state, first, is currently far from being dead, with both their number and the scope of the state in modern society growing significantly in the last century. Next to a more or less equivalent surge in local and functional public authorities, second, we have also witnessed the emergence of a variety of international authorities that are bound to become more important with the increase of mutual dependence in the modern economic world order. Even more significant than the persistence and development of public authorities, third, have been developments in the private sector. Not only is the institutional landscape of modern society inhabited by a broad variety of non-profit organizations that are central to the reproduction of modern society and whose political influence simply cannot be overestimated, but the ubiquity of authoritarian governance *within* markets has been such that the term organizational economy has become a more truthful predicate describing modern

economic life than the notion of market economy. Together with modernity's presumption against authority, these developments constitute an interesting paradox. For in spite of the modern doctrine that authority is to be questioned before it is obeyed, authoritarian governance is present in virtually every corner of modern society and is typically taken for granted. It is clear that this paradox is in need of explanation. It was the aim of this study to provide at least part of that explanation.

<div style="text-align:center">

2

***Explaining* authority falls short of appreciating what authority is about in modern society. *Understanding* the nature of authority requires addressing the *normative* question of legitimacy, that is, the question under which conditions an authoritative command is both action-guiding and binding for those addressed by it.**

</div>

I have offered three rough and ready explanations for the prevalence of authoritarian governance in modern society. Authority, first, is a familiar and effective means to cope with the high degree of specialization in modern society. It was explained that relying on the better judgement of specialists may be both highly rational and economically rewarding. The scope of this explanation proved very limited however. It explains only the prevalence of *theoretical* authority and therefore does not answer the question why extant forms of authoritarian governance typically go beyond what is required for theoretical authority to work. Because of the limited explanatory power of theoretical authority in general, I have focussed on *practical* authority – i.e. authority over questions of action rather than questions of belief – in the remainder of this study. A second explanation holds that authority is a comparatively efficient means to solve coordination problems in economic and social interaction. Such problems obtain when there are several possible courses of action that are roughly equally valued, but where the realization of that value is dependent on all actors acting similarly. Although this explanation proved clearly more powerful than the first one, it leaves out significant features that characterize authoritarian governance in modern society. In addition to solving coordination problems, therefore, a third explanation of authority focuses on authority as a means to cope with divergence of interests and problems of motivation in joint or collective action. Although this third explanation has proven quite powerful, it was argued that it relied too heavily on (a) unrealistic behavioural assumptions and (b) the rational control of behaviour. This excessive reliance undermines its ability to extend very far beyond the confines of the economic sphere in which at least the behavioural assumptions have some empirical rooting. It was concluded that what all these three explanations were lacking was a connection with (a) natural compliance and (b) the distinctively normative features that made authority such a powerful mode of governance in pre-modern society.

I have claimed that *understanding* authority therefore requires opening up the black box that authoritarian governance remains when it is looked at exclusively from an outside, so-called third-person's-perspective. This black box can be opened by

taking the insider's or first-person's perspective of those involved in the practice of either giving or receiving commands. What this perspective shows, first, is that authority must be made sense of in terms of reasons rather than causes. Without taking reasons into account, authority cannot be distinguished form coercion. Second, it reveals that commands must be conceived of as second-order *exclusionary* reasons for action that aim to exclude whatever relevant first-order reasons there are on the balance of reasons of those addressed by it. It was explained that someone who takes a command *on* the balance of reasons fails to understand its meaning. Making sense of authority in terms of this obligation to obedience highlights another central feature of authority relationships. This is that the actor giving the command apparently has a right to do so. This makes the question on what *normative* grounds an actor has a right to command and be obeyed central to our understanding of authority. The justification of authority is therefore part and parcel of what authority must be taken to mean. The stake of this study was to explore how the question of legitimacy impacts on our understanding of authority.

3
Although Raz's "service conception" of authority involves a valid general justification of authority, it is descriptively and normatively inadequate nevertheless.

I have discussed two *general* justifications of authority and have almost immediately rejected one of them. The position I have rejected without much further consideration involves the view that authority is justified by the consent of the governed. The reasons for my rejection were threefold. First, consent to authority is often not actually given. Virtually nobody has ever consented to the authority of the state, for example. But if consent may be implicit, second, then it may not be clear whether or not consent has actually been given. The point is that for an action to qualify as an act of consent there must be some background institutions *making* it an act of consent, while the presence of such institutions may not be taken for granted, particularly not in the context where consent to authority is at stake. Third, consenting to authority may be invalid. Consenting to authority can be invalid in quite a number of ways, but a powerful *general* argument against justification by consent holds that since consent is a form of content-independent commitment, it does not provide a content-full constraint on the scope of authority, while no unlimited authority can ever be valid. It was concluded that a valid justification of authority should therefore involve a content-full doctrine delimiting its scope in a content-full manner.

Joseph Raz' service conception of authority provides such a general justificatory doctrine. In his view, authority is justified *if, and only if*, it makes those subject to it better comply with reasons that apply to them already. Put in other words: a command is justified – and therefore authoritative – if it would be more effective or efficient for those subject to it to obey it, rather than to consider the reasons for

doing what the command requires directly. Although, insofar as I can see, the service conception is a valid justification of authority, it is a very restrictive doctrine nevertheless. It was claimed that neither the modern state, nor the firm – or any (formal) organization for that matter – are legitimate in this view, and hence do not really constitute authorities at all. It was concluded that the service conception of authority is therefore both descriptively and normatively inadequate. It is descriptively inadequate, first, because it cannot account for the majority of extant forms of authoritarian governance in modern society. It is normatively inadequate, second, because it cannot provide a solid normative foundation for many of the extant forms of authoritarian governance we typically take for granted in everyday life and value to have for utilitarian reasons.

In the light of these shortcomings this study set out to consider three distinctive issues that all bear on the justification of authority. The aim was to shed light on the claim to legitimacy that is tied up with our understanding of authority. The first issue involved the distinction between legitimate power and authority. Although the service conception defines authority in contradistinction to power, it was thought highly unlikely that such a *normative* understanding of authority can ever do without power. Power is indispensable both in solving coordination problems and in (occasionally) coercing those into obedience who do not comply naturally. The second issue involved our understanding of *legitimate* power. Legitimate power is power constrained by norms. The problem of dirty hands, however, questions our presumption that the kind of power that is associated with authority *can* be normatively constrained. The third issue involved our understanding of the notion of responsibility for authority. The idea of responsibility for authority was proposed as a pragmatic rather than a conceptual solution to the question of legitimacy.

4
Norms are like lines in the sand; they ultimately derive both their action-guidingness and their bindingness from the actions taken in respect to them.

The claim was that the central difference between authority and power involves the distinctively *normative* features of authority. The service conception is a normative understanding of authority because: (a) commands are *normally* understood in terms of their normative consequences, (b) it involves a doctrine of *legitimate* authority, and (c) authority is made sense of in terms of *reasons* rather than causes. I have claimed that the simplest way to conceive of the normative nature of authority was to see a command as an authoritative norm. It was argued, however, that that for a practical norm to be (a) action-guiding and (b) binding, it must be effective, that is, it must be generally applied and on the whole complied with. I have offered two arguments in support of the thesis that norms that are not effective are neither action-guiding nor binding.

The first argument was of a theoretical nature. It holds that knowing whether or not a norm applies involves rule following. The problem, however, is that to follow a

rule is not something that itself follows from the propositional content of that rule. This was, first, because in a applying a rule, we are in need of rules guiding its application, which in turn refer to other rules, which refer to still other rules, and so on. An endless regress of rules upon other rules seems inevitable here. But even if there were no threat of *endless* regress, second, the rule-following problem would remain intractable due to the necessary embeddedness of rules within other rules. The result of this embeddedness is that the rule in question can be made to accord with one set of conditions, at one time, and with contrary conditions, on another, solely by manipulating the internal relations between rules in the whole normative system. Due to this *holistic* feature of normativity, the application of a norm is *underdetermined* by its (propositional) content. This does not mean that the application of a norm is undetermined altogether, however. An alternative account of rule following was proposed in which following a rule is conceived of as a *practice*. In this view there is nothing prior to the actual practice of applying a norm that determines its proper application. Philip Pettit has given a plausible account of what such a pragmatic account of rule following must minimally involve. First, there must be certain habits of response that fixate, so to speak, the application of the rule in question. But since mere *habits* cannot be fallible, while normativity necessarily requires fallibility, second, there also must at least be the possibility of *ex post* correction of these habits of response. Such corrections are possible and relevant because there are good and bad conditions under which a norm can be applied. This requires that bad conditions are discounted for in applying a norm. Although *ex post* correction of habits of response can, in principle, take place inter-temporally within a single actor, it was argued that such correctional practices are typically of a social nature in the case of *practical* norms. It was also claimed that reflection and deliberation typically play a central role in these practices. It was concluded therefore, that without there being some practice in which a norm is actually applied and on the whole complied with, there can be no proper application of that norm, since it is this practice that determines its application in the first place. Hence no ineffective norm can be action-guiding.

The second argument builds on the first but was of a *practical* nature. It proceeded from the presumption that the major function of practical norms is to represent, create or uphold mutual normative expectations in social interaction. It was argued, however, that mutual *normative* expectations ultimately depend on there being a structure of mutual *positive* expectations that is, on the whole, consistent with what the normative expectations require. Without such a structure, or with a structure of mutual positive expectations that is inconsistent with what the relevant normative expectations require, norms lose their bindingness. They do so for two reasons. I have argued, first, that reactive sentiments or habits of response evolve into distinctively normative expectations only after they have been corrected for a (a) general and (b) resilient perspective that make them apply – in principle – to an infinite umber of other conditions. It is clear that such a general and resilient perspective is undermined if positive expectations do not accord with what the norm requires. Second, any serious inconsistency between normative expectations, on the

one hand, and whatever positive expectations there are, on the other, always constitute *general* invalidating conditions for the application of that norm, in the sense that anyone who is held to it may justifiably claim that it is unreasonable to do so. To say that there are *general* invalidating conditions, however, is to say that the norm lacks bindingness.

5
Might can make right.

The thesis that might can make right follows from the conjunction of the thesis that norms must be effective in order to be action-guiding and binding, on the one hand, and the thesis that norms may not be spontaneously effective – i.e. effective without some external *intervention* making them effective – on the other. I have explained that norms are ineffective when they do not rest on a structure of positive expectations that is, on the whole, consistent with what the norm requires. Although there are many ways in which this may be true, I have discussed five ideal typical situations deriving from basic game theory in which this may be so. These included: (a) coordination problems, (b) 'battle of the sexes', (c) chicken games, (d) the prisoners' dilemma and (e) zero-sum division games. The discussion of these ideal types showed that the stronger the conflicts of interests (or values) characterizing the structure of mutual positive expectations underlying a norm, the more intensive and invasive the intervention must be to secure its effectiveness. Since it may ultimately be intervention on which the effectiveness, and hence action-guidingness and bindingness of a norm depends, might can indeed *make* right. An example demonstrating this general thesis involved an impure coordination game like the battle of the sexes. In this example there are two mutually excluding conventional norms – say about the rules governing the issuance of academic degrees – that are valued differentially by two different constituencies – say that each country want its own rules to be adopted – while both constituencies prefer to have one set of rules over two or more. If the two constituencies are of roughly equal size and influence, coercive intervention may be necessary to select one of these sets of rules. Once one set of rules is selected, however, it does not pay anymore not to comply with these rules because sunk investments may have been written of and transition investments may have already been made. In such cases, therefore, it is ultimately might that makes right.

6
There is no conceptually solid distinction between authority conceived of by the service conception, on the one hand, and the classic Weberian conception of authority as legitimate power, on the other.

According to the service conception, authority is legitimate and hence normative *if, and only if,* obeying a command helps those subject to it better to comply with reasons that apply to them already, than if these reasons themselves are acted upon directly. The requirement that norms be effective in order to be action-guiding and binding, on the one hand, together with the fact that not all norms are effective without some intervention making them effective, on the other, already indicates that intervention may very well *create* reasons for action that actors do not have – typically because of lack of assurance – without such intervention taking place. It was explained that there is thus a real sense here in which might can make right and in which intervention can *create* authority as conceived of by the service conception. There is therefore no clear-cut conceptual boundary between authority and (legitimate) power. It was explained that this has two important consequences. That there is no such distinction, first, seriously undermines our efforts to conceptually come to grips with the nature of authority, for an important way in which the claim to legitimacy could be incorporated in our understanding of authority has now been cut off. It is clear that whatever claim to legitimacy can still be made, must extend to power or intervention as well. But if there is no conceptually airtight demarcation between authority and legitimate power, second, then neither can there be a conceptually sound distinction between "claim-rights" and "justification-rights" that is founded on it. Ultimately, there may be just power that is legitimate and power that is not.

7
In modern society the problem of Dirty Hands is best understood in terms of representative action (characteristic) in a context of normative conflict (background).

The distinction between power, on the one hand, and legitimate power – i.e. power constrained by norms – was at stake in chapter III of this study. In this chapter I set out to investigate the impact of an interesting anomaly to the claim that intervention is constrained by the norms that apply at any point in space and time. The anomaly involved the so-called problem of dirty hands, that is, the problem that political leaders or other persons in positions of authority must sometimes act bad in order to do good. It was explained that although this problem is often attributed to Nicolo Machiavelli, it is actually a philosophical insight that traces to the very origins of Western culture, as it was already explicit in some of the earliest texts available to us. Classic Greek philosophy and tragedy are a case in point. In spite of its ancient origins, however, the problem is presently still surrounded with conceptual vagueness and confusion. It has never become clear, first, which actual cases qualify as cases of dirty hands and which do not. Neither is it unambiguous, second, how the problem

of dirty hands relates to similar, yet arguably different problems, such as those denoted by the notions of tragic choice and moral dilemma. It was one of the aims of chapter III to bring some conceptual clarification to this matter.

The conceptual strategy adopted was that of characterization, that is, of defining the problem of dirty hands in terms of its necessary (background) and sufficient conditions (characteristics). It was explained that this strategy results in a *relative* perspective on the problem of dirty hands at best, because there is no fixed, pre-conceptual background against which it can be made sense of. It was claimed, first, that value or normative conflict is a necessary condition for dirty hands. This condition obtains when two or more action-guiding and binding norms claim two or more distinctive courses of action that are mutually exclusive. It was explained that such a condition involves something deeper than a mere conceptual confusion or some epistemological problem that can be solved once more information or conceptual refinement is available. I have claimed that because normativity is *ontologically* constituted by our commitment to the world, *epistemological* means alone may very well be insufficient for resolving inconsistency within that commitment. The requirement that action-guiding and binding norms be embedded in a practice in which that norm is generally applied and on the whole complied with makes clear that there is more to normative conflict than can be captured from a conceptual or epistemological perspective alone. For even if we would have all information and would be perfectly rational processors of that information, it may still be true that we have to fight wars and kill people to safeguard the values we stand for at any point in space and time.

I have proposed that what characterizes the problem of dirty hands against the background of normative conflict is the *representative* feature of the actions taking place in a context of normative conflict. I have explained that acting in a context of normative conflict necessarily involves incurring *real* costs. But costs do not yet constitute blame, which attaches to dirty hands in all understandings of the problem, even in the most inarticulate ones. I have explained that what turns cost into blame is when the cost-causing-action is taken by an actor who acts on behalf of those who have to incur the costs. For only then will the latter have sufficient ground for attributing blame to the cost-causing actor. In sum, then, the representative feature is a characteristic for dirty hands against the background of normative conflict if the condition is fulfilled that any normative conflict that is not a case of dirty hands does not involve representative action.

I have claimed three advantages of the proposed conceptual perspective on the problem of dirty hands. First, there is the conceptual advantage that extensional and intensional ambiguities surrounding the problem may be cleared to a certain extent. The former is done by stipulating two clear demarcation criteria that enable us to separate cases of dirty hands from cases that are not. The latter is done by making explicit the feature that distinguishes dirty hands from other paradigmatic forms of normative conflict – such as tragic choices and moral dilemma's – that are similar to it but are evidently sufficiently different at the same time. Thus neither tragic choices

nor moral dilemma's need involve representative action, for example, in more or less the same way dirty hands need not constitute much of a dilemma for those who have to make them. I have claimed, second, that making such a conceptual distinction has become necessary even if only because of the differences of scale that distinguish modern society from its ancient predecessors. It is clear that the concept of dirty hands has now outgrown the 'face to face' features of the ancient society in which the problem was first made explicit, and that the personal connotation of that environment not only blends it with similar yet arguably different problems, but also keeps us from seeing new ways to cope with old problems. I have explained that those making dirty hands today typically do not have to sacrifice their own daughters, and that with this divergence of personal grief from the pain caused to others by the actions necessary, some sort of natural break on choosing lightly went lost. New ways have to be found to compensate for this loss. I have claimed that the proposed conceptual perspective on the problem of dirty hands may guide us to these ways. A final advantage of the characterization of the problem of dirty hands proposed here is that it enables us to see the *conceptual* relationship between making dirty hands, on the one hand, and authority and intervention, on the other. I argue that this relationship keeps us from conceiving intervention and authority in the classic Weberian sense as *legitimate* power, that is, as power constrained by norms.

8
Authority and Intervention are *ultimately* unconstrained by the norms that apply at any point in space and time.

The characterization of the problem of dirty hands in terms of normative conflict and representative action helps us to see why intervention and authority are *ultimately* unconstrained by the norms that apply at any point in time and space. This is because the condition of normative conflict makes it possible that we can realize – i.e. make effective – one norm only at the cost of breaking another. The reality of contemporary drug policy is a case in point. The same example also demonstrates that intervening on behalf of the effectiveness of a norm may sometimes even require breaking that very same norm. Intervention may therefore ultimately result in the intervenor making dirty hands. I have argued that the problem of dirty hands should consequently be conceived of as an *extreme* case of intervention, in that making dirty hands is what intervention may ultimately come down to.

I have also claimed that this extreme case is more than some accidental and exceptional contingency. The reason for this is that it highlights some highly significant conceptual features of both intervention and authority. It shows, first, that representative relationships in general, and authority in particular, may very well enter a condition of normative conflict when different constituencies differentially endorse two or more mutually exclusive norms. I have argued that such conflicts are common to what everyday representative relationships are about in modern society. Second, it reveals that next to coping with coordination problems and combating problems of

motivation, dealing with problems of *differential ultimate valuation* is a central function of authority in modern society. Thus forms of *adjudicative* authority, first, are inevitable in any normative order because even the most carefully designed normative systems leave room for applicational ambiguity and normative conflict. Forms of *legislative* authority are central, second, because in any human association where there are potentially many mutually exclusive guiding stars, some stars will have to be authoritatively endorsed and others rejected in order to escape from chaos and paralysing normative conflict. Forms of *executive authority*, finally, are highly functional in modern society because they enable us to detach making dirty hands in the realization of normative order to a certain degree from the domain to which the norms of that order apply, such that the interventionist action itself does not undermine the effectiveness of that order. I have claimed that the monopolization of violence by the modern state is no coincidence in this respect.

Because intervention and authority are ultimately unconstrained by the norms that apply at any point in time, there is no fixed demarcation line between power, on the one hand, and legitimate power on the other. Hence conceiving authority in the classic Weberian sense as legitimate power has become highly problematic. It is an interesting question how this insight reflects on modern ideals such as the Rechtsstaat, constitutional government, bureaucracy and organization. This question was not addressed in this study, however. At this stage the overall conclusion was that we are left empty-handed in our efforts to give a general, theoretical answer to the question of legitimacy that is indissolubly tied up in our understanding of authority. Not only did we have to conclude that there is no fixed conceptual boundary between our normative understanding of authority and legitimate power, but the final frontier between legitimate power and power as such could not be defended in theoretical terms. It was suggested, therefore, that perhaps we should abandon our attempt to theoretically come to grips with the question of legitimacy and explore the possibility of more pragmatic solutions for what is now the *quest for* rather than the *question of* legitimacy.

<div align="center">

9

While it is unlikely that the basic requirements for the *concept* of responsibility to apply will actually be sufficiently fulfilled, the practice of responsibility is ubiquitous in modern society and is typically taken for granted.

</div>

In chapter IV the attempt to give a theoretical answer to the question of legitimacy that is indissolubly tied up with our understanding of authority was abandoned in favour of uncovering a more pragmatic solution to what turned out to be an ongoing quest for legitimacy. In this chapter I set out to investigate the possibility of an understanding of responsibility for authority on the promise that responsibility can perhaps relate to authority within hierarchy as demand relates to supply within a market, that is, on the prospect that the practical problems pertaining to the

legitimacy of authoritarian governance will be cleared once authority and responsibility are in some sort of equilibrium. Following a survey of the different meanings of responsibility that are implied in Western society's most important practices and institutions, and having positioned these meanings on an underlying dimension of differential normative involvement, three basic understandings of responsibility were identified. This involved, first, a *cognitive* understanding of responsibility. This understanding assumes an implicit normative commitment that enables us to understand certain causal relations as in a special way significant to our evaluative practices. It was argued that this minimally involved understanding these causal relations as the (possible) result of *intentional* action. A second basic understanding concerned a *regulative* understanding of responsibility. This involved responsibility as a social *mechanism*. It involved a conjunction of there being (a) explicit action-guiding normative expectations and (b) some sort of sanctioning practice to uphold these expectations. It was argued that the normative expectations characteristic for this regulative understanding of responsibility, do not necessarily involve an ultimate kind of normative involvement. It was claimed instead, that this involvement is often only instrumental because the normative expectations involved are typically subject to their being socially efficacious. At the end of our dimension of normative commitment, finally, lay an inherently *normative* understanding of responsibility. This understanding involved both an ultimate justification of our practices of holding actors responsible, and the virtue of responsibility. It was explained that the former and the latter are intimately related.

From these three basic understandings I subsequently derived four fundamental requirements or necessary conditions which must be fulfilled in order for the concept of responsibility to apply. Thus our cognitive understanding of responsibility, first, required that one must be able understand certain facts as the (possible) result of the acts or omissions of an intentional actor. Our regulative understanding of responsibility, second and third, necessitates there being both an unambiguous normative background consisting of explicit action-guiding normative expectations, and an accountability practice that can uphold these expectations. Our normative understanding of responsibility, finally, required that the actor to which responsibility is attributed satisfies certain necessary conditions in order to justifiably be held responsible. It was claimed that these four requirements are robust over both an *intensional* and the actually followed *extensional* strategy of conceptualising responsibility.

In the remainder of chapter IV it was argued that it is unlikely that these requirements will *actually* be met, not just for responsibility for authority, but for responsibility *in general*. I have discussed several potentially critical complications that stand in the way of the *concept* of responsibility applying. The first complication involved the problem of *ascription*, that is, the problem that the conditions that apply may make it difficult if not impossible to ascribe acts, omissions and their consequences to an intentional actor. A second complication involved the problem of *practical necessity*. This is the problem that the normative background against which an

actor is to be held responsible may be ridden with normative conflict such that the functioning of the mechanism of responsibility is critically undermined. A third complication involved the problem of *constituency*. This problem held that a mismatch might apply between the actor holding another actor responsible and the normative background against which the latter is to be held responsible. A fourth problem involved the problem of *responsiveness*. The gist of this problem is that we can never be sure that the behaviour we want to hold an actor responsible for is not the result of the actor not having what it takes to be responsible. Ironically, this is a greater problem for individual responsibility than for responsibility for authority. The fifth problem involved the problem of the *organizational* constitution of authority. It was argued that attributing responsibility to organizations may ultimately be as problematic as morally blaming sociopaths – those who lack the capacity for empathy and emotional reflection – for their sociopathic behaviour. This chapter ended with the paradoxical conclusion that although it is unlikely that the conditions necessary for the concept of responsibility to apply will *actually* be met simultaneously, the practice of responsibility is ubiquitous in modern society nevertheless.

10
A *general* theory of institutions takes cognitive, regulative and normative approaches in institutional theory as representing alternative *mechanisms* in the constitution of institutional facts.

In chapter V, the air of paradox was resolved by letting go of the presumption that our understanding of responsibility must be *conceptual*. An alternative was explored in some detail. This alternative consisted of a *general* institutional perspective on responsibility. Taking this perspective required that we make sense of the notion of institution first. This proved to be no easy task. Although the notion of institution has had a recent revival in social science and the humanities, a quick glance at the prevailing literature revealed that the meaning of this notion differs significantly along disciplinary lines. Rather than following these often rather arbitrary lines, I set out to develop a systematic and comprehensive perspective on institutions for two separate but related reasons. First, the development of institutional theory has taken place with little regard of disciplinary boundaries. Second, some of the most interesting applications of institutional theory have taken place within the interdisciplinary study of organizations.

I have characterized the notion of institution in contrast to two widespread assumptions of modern social science that are often made somewhat overzealously. The first assumption involves behaviourism. This is the position that the study of man and society should be undertaken in exactly the same way that natural phenomena are studied, that is, by taking a third-person's-perspective on the object of study. The second assumption involves methodological individualism. This is the view that the individual is the natural and exclusive unit of analysis in the study of man and society. With respect to the latter assumption it was observed, first, that

human reality contains chunks of irreducibly 'social matter' and that this distinctively social part of human reality significantly shapes the behaviour of individual actors. The first characteristic of institutions is therefore that human action is *socially constrained*. The second characteristic followed from taking issue with the assumption of behaviourism. It was argued that there is a crucial distinction between institutional and natural constraints to human action, and that the distinguishing feature involves the *intentional* nature of institutional constraints. The fact that a certain chunk of material constitutes a house, for example, is a socially imposed intentional constraint to human action, while the fact that this house is made of stone and wood are natural constraints to it. It is clear that making sense of such institutional facts requires taking the first-person's-perspective of those involved with them. The third characteristic identified holds that there are different ways in which the intentional imposition of the status-functions distinguishing institutional from natural facts can take place. It was argued that three distinctive branches of positive institutional theory identify at least three generic ways in which *status-functions* are imposed.

The *regulative* branch of positive institutional theory, first, involves a consequentialist explanatory approach to the supra-individual constraints in social interaction. The so-called "logic of consequentiality" characteristic of this approach works through the rational anticipation of the consequences of action. In this view, institutional constraints to action arise and work through *ex post* sanctioning of behaviour. The *cognitive* branch of positive institutional theory, second, explains how cognitive economizing, social construction and (*ex post*) sense making, fixate belief in social interaction. The "logic of orthodoxy" that defines this approach to institutions works mainly by *ex ante* screening for beliefs and action. The *normative* branch of positive institutional theory, third, comes to an understanding of institutions by focussing on processes of *ultimate* valuation that are triggered by, and sustained in, social interaction. The dominant logic of this approach to institutions is "the logic of appropriateness" which works both by *ex ante* screening for belief and actions and *non*-intentional *ex post* sanctioning of behaviour. The thesis proposed and argued for in chapter V held that these three distinctive branches of positive institutional theory represent three generic mechanisms of status-function-imposition that can – in principle – operate independently, but that are empirically more often found to work together in the constitution of institutional facts. It was claimed that when operating in conjunction, these three mechanisms over-determine the institutional facts they are constitutive of.

Although not much more was said on this *general* institutional perspective than that (a) the different mechanisms of status-function-imposition may operate disjunctively and (b) that they over-determine institutional reality when operating in conjunction, it has become clear throughout this study that it sheds interesting light on some peculiar phenomena that cannot easily be understood from other theoretical perspectives. One of these phenomena involved the self-fulfilling potential of overly self-regarding behavioural assumptions in institutional theory. It was argued that the purposive construction of institutions on the basis of unrealistic knavish behavioural

assumptions may very well make these behavioural assumptions come true once they become the socially *sanctioned* default assumptions in social interaction. It was explained (albeit in a footnote) that making the regulative mechanism of status-function-imposition work in a particular way may very well make the cognitive mechanism follow its trail. Similarly, this general institutional theory sheds interesting light on the notion of an essentially contested concept; a notion that seemed to apply to the two central concepts at stake in this study; i.e. the notions of authority and responsibility. Seen from the perspective of a general theory of institutions, an essentially contested concept denotes an *institution* for which the proper functioning of the cognitive mechanism of status-function-imposition is hindered by the poor or even counterproductive functioning of the normative mechanism of status-function-imposition. The latter two remarks call attention to the need to come to an understanding of how the different mechanisms of status-function-imposition may relate to each other within the province determined by working disjunctively, on the one hand, an working in conjunction, on the other.

11
The notion of responsibility is better understood in *institutional* than in *conceptual* terms.

It was explained that the three basic understandings of responsibility identified in chapter IV more or less coincide – not coincidentally – with the three generic institutional mechanisms discussed in chapter V of this study. Thus our cognitive understanding of responsibility, first, sheds light how we cognitively make sense of and attribute action and its consequences to actors in everyday social interaction. And our regulative understanding of responsibility, second, calls attention to how normative expectations and sanctioning practices are framed and fitted together in social interaction in the light of social efficacy. Our normative understanding of responsibility, finally, points at how processes of ultimate valuation constrain both our attributions of responsibility and shape our modern self-image as autonomous beings for which being a responsible actor is desirable for its own sake.

In contradistinction to a conceptual understanding of responsibility according to which the requirements pertaining to the three basic understandings of responsibility must be fulfilled *conjunctively* in order for it to apply, an institutional understanding allows these requirements to be fulfilled *disjunctively* because it sees each of these understandings as *alternative* mechanisms of status-function-imposition for the institution of responsibility. It was argued that an institutional understanding of responsibility was both theoretically and pragmatically superior over a conceptual one. It is theoretically superior, first, because an institutional understanding of responsibility is more in touch with modern everyday reality in which the practice of responsibility is ubiquitous and often taken for granted. It is theoretically superior, second, because it can incorporate both responsibility for authority and individual human responsibility within a single conceptual framework – i.e. that of a *general*

institutional theory – while it is not very likely that this is, or can be made true of a conceptual understanding of responsibility. It is pragmatically superior not only because it is more applicable in practice, but also because it sheds interesting light on how the institution of responsibility for authority can be made more resilient in practice by exploiting the complementary nature of the different mechanisms constitutive of it.

12
The currently dominant yet highly criticized agency-theoretical approach to authority relations can be theoretically complemented by placing it within the more comprehensive theoretical framework of a general institutional understanding of responsibility for authority.

In order to sketch the rough outlines of what a general institutional theory of responsibility for authority may look like, a more detailed analysis was made of how the regulative (mechanism) of responsibility that forms the heart of an agency theoretical perspective on authority relationships may be complemented by other institutional mechanisms. It was explained that agency relations in general, and authority relations in particular, are often troubled by the problematic functioning of our cognitive understanding of responsibility because there typically is an information asymmetry between agents and their principals to the potential benefit of the former. Under the assumption of a divergence of interests between principals and their agents – and it must be remembered that authoritative governance is often instituted precisely to cope with interest divergence and conflicting ultimate valuation – this may not only lead to the interests of the principals being harmed, but also to a significant deadweight (welfare) loss for all. It was argued that agency theoretical solutions to this so called agency problem were either constrained by the basic relations of the theory itself – i.e. (a) monitoring is constrained by the persistence of information asymmetry while (b) interest alignment is constrained by the divergence of interests given with the relationship itself – or were relying excessively on the rational control of behaviour. Since the latter may be particularly ineffective in a context of significant and persistent information asymmetry – as is commonly the case in authority relationships – such excessive reliance may be found counterproductive.

It was argued that shaping and cultivating a normative understanding of responsibility in authority relationships may help to overcome such problems. It was claimed that the notion of *trust* points at a potentially fruitful way in which a normative understanding of responsibility can be shaped and cultivated in authoritarian governance. It was argued that in the same way that our normative understanding of responsibility is a central constitutive virtue of human agency, trustworthiness is a necessary and valuable constitutive virtue of authoritative agency. Three rough and ready reasons were advanced why trustworthiness may not only be valued by those subject to authority, but also by those exercising it. This is, first,

because trustworthiness may be valued inherently by those seeking authoritative office. Trustworthiness may be valued by those in authority, second, because those subject to authority value the trustworthiness of their leaders. When the former care about the attitudes of their subjects, being a trustworthy leader may be valuable indirectly by those in authority. Trustworthiness may prevail amongst authoritative agents, third, because given general elections or some other competitive screening mechanism it is not unlikely that trustworthy actors have a relatively greater chance of being selected in to positions of authority.

13
Since authority should be made sense of in institutional rather than conceptual terms, the quest for legitimacy that is indissolubly tied up with what authority is about, *ultimately* involves a problem of institutional design.

If anything, this study has made our understanding of the nature of authority more problematic and complex than the ubiquity of authoritarian governance in modern society seems to suggest. The reason for this was that an answer to the question of legitimacy that is indissolubly tied up with our understanding of authority could not be conceptually incorporated within that understanding. This was, first, because there is no solid conceptual demarcation between the normative service conception of authority, on the one hand, and the classic Weberian idea of authority as legitimate power, on the other; and second, because neither could there be drawn a conceptually firm boundary between the latter and power as such. The pragmatically inspired effort to come to an understanding of responsibility for authority went hand in hand with accepting that there is no definite theoretical answer to the question of legitimacy and that we should rather conceive of it as an ongoing quest for legitimacy. The subsequent finding that responsibility for authority is better conceived in institutional than in conceptual terms already foreshadowed the overall conclusion that the notion of authority also denotes an institution rather than a phenomenon that can be captured in unambiguous conceptual terms.

There are two major advantages to understanding authority in institutional rather than conceptual terms. From the perspective of those *in* authority, first, conceiving of authority as an institution makes it easier to understand – and hence to accept – the apparently unbreakable bonds between authority and power in authoritarian governance. This is done in part by making explicit the continuos and complementary features of different obedience generating mechanisms, ranging from natural compliance on the one extreme, to coerced compliance on the other. Understanding the various ways in which compliance to commands can be secured is central to explaining the extraordinary persistence and resilience of authoritarian governance in modern society. From the perspective of those *subject* to authority, second, the finding that there is more than one way in which the question of legitimacy can be 'answered', significantly broadens the scope of legitimate authority in modern society. Thus authoritarian governance may be justified if: (a) complying to a command makes

those to whom it is addressed complying better to reasons they already had to do what the command requires, or (b) if the authority giving the command is sufficiently constrained by the norms that apply at any point in time, or (c) when the authority is made sufficiently responsible for its acts and omissions in case it exceeds these normative constraints. It is therefore *ultimately* the institution of responsibility that answers the quest for legitimacy that authoritarian governance confronts us with. Since this institution is often not sufficiently in place, or may not be optimally fitted to the kind and scope of the kind of authoritative mechanisms involved, there is room for design and purposive intervention in developing and shaping institutions of responsibility in modern society. It was the ambition of this study to provide a viable theoretical perspective for such efforts.

References

ALBANESE, R., M. T. DACIN, and I. C. HARRIS (1997): "Agents as Stewards", *Academy of Management Review*, 183, 609-611.

ALCHIAN, A. A., and H. DEMSETZ (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777-795.

ANSCOMBE, G. E. M. (1958): "On Brute Facts," *Analysis*, 18.

AOKI, M. (2000): "Institutional Evolution as Punctuated Equilibria," in *Institutions, Contracts and Organizations; Perspectives Form New Institutional Economics*, ed. by C. Menard. Cheltenham: Edward Elgar.

ARENDT, H. (1953): "What Was Authority?," in *Authority; Nomos I*, ed. by C. J. Friedrich. Cambridge (MA): Harvard University Press.

— (1969): *On Violence*. Florida: Harcourt Brace.

ARROW, K. J. (1974): *The Limits of Organization*. New York: John Brockman Associates.

— (1985): "The Economics of Agency," in *Principals and Agents: The Structure of Business*, ed. by J. W. Pratt, and R. J. Zeckhauser. Boston: Harvard Business School Press.

AUSTIN, J. (1832 / 1998): *The Province of Jurisprudence Determined*. Aldershot: Ashgate.

AUSTIN, J. L (1962/1975): .*How to do Things with Words*. Cambridge (MA): Harvard University Press.

BARNARD, C. (1968): *The Functions of the Executive*. Cambridge, MA: Harvard University Press.

BARNEY, J. B., and W. HESTERLY (1996): "Organizational Economics; Understanding the Relationship between Organizations and Economic Analysis," in *Handbook of Organization Studies.*, ed. by S. R. Clegg, C. Hardy, and W. Nord. London: Sage.

BARRY, B. (1984): "Tragic Choices," *Ethics*, 94.

BAUM, J. A. C. (2001): *A Companion to Organizations*. Oxford: Blackwell.

BAUM, J. A. C., and T. J. ROWLEY (2001): "Companion to Organizations: An Introduction," in *A Companion to Organizations*, ed. by J. A. C. Baum.

BENDERSKY, J. W. (1983): *Carl Schmitt; Theorist for the Reich*. Princeton: Princeton University Press.

— (1996): "Schmitt and Hobbes," *Telos*, 1996, 122-130.

BENDIX, R. (1956): *Work and Authority in Industry: Ideologies of Management in the Course of Industrialization*. New York: Wiley.

— (1978): *Kings or People: Power and the Mandate to Rule*. Berkeley and Los Angeles: University of California Press.

BEN-NER, A., and L. PUTTERMAN (1998): *Economics, Values and Organization*, Cambridge: Cambridge University Press.

BERLE JR., A. A., and M. G.C. (1932): *The Modern Corporation and Private Property*. New York: MacMillan.

BERLIN, I. (1953/1997): "The Originality of Machiavelli," in *The Proper Study of Mankind: An Anthology of Essays*, ed. by I. Berlin. London: Chatto & Windus.

— (1958): *Two Concepts of Liberty*. Oxford: Clarendon Press.

— (1978): *Concepts and Categories: Philosophical Essays*. London: Hogarth Press.

— (1979): *Against the Current: Essays in the History of Ideas*. London: Hogarth Press.

— (1990): *The Crooked Timber of Humanity: Chapters in the History of Ideas*. London: John Murray.

— (1997): *The Proper Study of Mankind; an Anthology of Essays*. London: Chatto & Windus.

BLACKBURN, S. (1998): "Trust, Cooperation and Human Psychology.," in *Trust and Governance*, ed. by V. Braithwaite, and M. Levi. New York: Russel Sage Foundation.

BLOM, T. (1998): *Drugs in het Recht; Een Onderzoek naar de Relatie tussen de Strafrechtelijke Handhaving van de Opiumwet, en de daarbij gehanteerde Opsporingsmethoden en het Recht op Privacy van Burgers*. Deventer: Gouda Quint.

BOGHOSSIAN, P. (1989): "The Rule-Following Considerations," *Mind*, 98, 507.

BOVENS, M. A. P. (1998): *The Quest for Responsibility; Accountability and Citizenship in Complex Organisations*. Cambridge: Cambridge University Press.

BRAITHWAITE, V., and M. LEVI (1998): *Trust and Governance*, New York: Russel Sage Foundation.

BRANDT, R. B. (1972): "Utiltarianism and the Rules of War", *Philosophy and Public Affairs*.

BRATMAN, M. E. (1993): "Shared Intention," *Ethics*, 104, 97.

— (1999): *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

BRENNAN, G. (1998): "Democratic Trust; a Rational Choice Approach", in *Trust and Governance*, ed. by V. Braithwaite, and M. Levi. New York: Russel Sage Foundation.

BRINTON, M. C., and V. NEE (1998): *The New Institutionalism in Sociology*. New York: Russel Sage.

BULL, H. (1977): *The Anarchical Society*. London: Macmillan.

CAMBELL, D. (1997): "The Role of Monitoring and Morality in Company Law: A Critism of the Direction of Present Regulation," *Australian Journal of Corporate Law*, 7, 343365.

CHANDLER, A. D. (1977): *The Visible Hand: The Managerial Revolution in American Business*. Cambridge (MA): Belknap Press.

CHANG, R. (1997): "Incommensurability, Incomparability and Practical Reason," Cambridge Mass.: Harvard University Press.

CHAPMAN, B. (1998a): "Law, Incommensurability, and Conceptual Sequenced Argument," *University of Pennsylvania Law Review*, 146, 1487-1528.

— (1998b): "More Easily Said Than Done: Reason and Rational Social Choice," *Oxford Journal of Legal Studies*, 18, 293-329.

CLARK, R. C. (1985): "Agency Costs Versus Fiduciary Duties," in *Principals and Agents: The Structure of Business*, ed. by J. W. Pratt, and R. J. Zeckhauser. Boston: Harvard Business School Press, 55-79.

CONNOLLY, W. E. (1974): *The Terms of Political Discourse*. Lexington (MA): Heath.

— (1984): *Legitimacy and the State*. Oxford: Basil BLackwell.

CONNOR, K. R. (1991): "A Historical Comparison of Resource-Based Theory and Five Schools of Thought within Industrial Organization Economics: Do We Have a New Theory of the Firm?," *Journal of Management*, 17, 121-154.

CONNOR, K. R., and C. K. PRAHALAD (1996): "A Resource-Based Theory of the Firm: Knowledge Versus Opportunism," *Organization Science*, 7.

CULLITY, G., and B. GAUT (1997): "Introduction," in *Ethics and Practical Reason*, ed. by G. Cullity, and B. Gaut. Oxford: Clarendon Press, 1-27.

DAHL, R. A., and E. R. TUFTE (1973): *Size and Democracy*. Stanford, Calif.: Stanford University Press.

D'ANDRADE, R. G. (1984): "Cultural Meaning Systems," in *Culture Theory: Essays on Mind, Self, and Emotion*, ed. by R. A. Schweder, and R. A. LeVine. Cambridge: Cambridge University press, 88-119.

DAVIE, W. (1980): "The Extreme Case in Ethics," *Philosophical Investigations*, 3.

DAVIS, J. H., F. D. SCHOORMAN, and L. DONALDSON (1997): "Toward a Stewardship Theory of Management," *Academy of Management Review*, 22, 20-47.

DEMSETZ (1995): *The Economics of the Business Firm*. Cambridge: Cambridge University Press.

DENNET, D. C. (1981/1987): "Three Kinds of Intentional Psychology," in *The Intentional Stance*, ed. by D. C. Dennet. Cambridge (MA): MIT Press.

DOUGLAS, M. (1986): *How Institutions Think*. Syracuse, NY.: Syracuse University Press.

DUNN, J. (1988): "Trust and Political Agency," in *Trust: Making and Breaking Cooperative Relations*, ed. by D. G. Gambetta. Oxford: Basil Blackwell.

EASTON, D. (1965): *A System Analysis of Political Life*. New York: Wiley.

EISENHARDT, K. M. (1989): "Agency Theory: An Assessment and Review," *Academy of Management Review*, 14, 47-74.

ELSTER, J. (1989): *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

— (1999): *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.

ENZLE, M. E., and S. C. ANDERSON (1993): "Surveillant Intentions and Intrinsic Motivation.," *Journal of Personality and Social Psychology.*, 257-266.

FAMA, E. (1980): "Agency Problems and the Theory of the Firm," *Journal of Political Economy*, 88, 288-307.

FAMA, E. F., and M. C. JENSEN (1983): "Separation of Ownership and Control," *Journal of Law and Economics*, 26, 301-326.

FEINBERG, J. (1970a): "The Nature and Value of Rights", *The Journal of Value Inquiry*, 4, 243-257.

— (1970b): *Doing and Deserving: Essays in the Theory of Responsibility.* Princeton: Princeton University Press.

FEREJOHN, J. (1999): "Accountability and Authority; toward a Theory of Political Accountability," in *Democracy, Accountability and Representation*, ed. by A. Przeworski, S. C. Stokes, and B. Manin. Cambridge: Cambridge University Press.

FRENCH, P. A. (1984): *Collective and Corporate Responsibility.* New York: Columbia University Press.

FRIEDRICH, C. J. (1958): *Authority: Nomos I*, Cambridge, Mass.: Harvard University Press.

— (1960): *Responsibility: Nomos III*, New york: The Liberal Art Press.

FULLER, L. (1964): *The Morality of Law.* New Haven: Yale University Press.

GALLIE, W. B. (1956): "Essentially Contested Concepts," *Proceedings of the Aristotelian Society*, 56.

GAMBETTA, D. G. (1988): "Trust: Making and Breaking Cooperative Relations," Oxford: Basil Blackwell.

GEERTZ, C. (1973): *The Interpretation of Culture.* New York: Basic Books.

GILBERT, M. (1989): *On Social Facts.* London: Routledge.

— (2000): *Sociality and Responsibility: New Essays in Plural Subject Theory.* Oxford: Rowman & Littlefield Publishers.

GLASTRA VAN LOON, J. F. (1956): *Norm en Handeling: een Bijdrage tot een Kentheoretische Fundering der Sociale Wetenschappen.* Haarlem.

— (1958): "Rules and Commands," *Mind*, 67, 514-521.

— (1980): *De Eenheid van het Handelen.* Meppel: Boom.

— (1987): *De Ijdelheid van Beslissingen: Opstellen over Recht, Politiek en Ethiek.* Leiden: Stenfert Kroese.

GOODIN, R. E. (1996): *The Theory of Institutional Design*, Cambridge: Cambridge University Press.

GOODIN, R. E., and P. PETTIT (1993): *A Companion to Contemporary Political Philosophy.* Oxford: Blackwell.

GRANOVETTER, M. (1985): "Economic Action and Social Structure: The Problem of Embeddedness," *American Journal of Sociology*, 91, 481-510.

GRANT, R. M. (1996): "Toward a Knowledge-Based Theory of the Firm.," *Strategic Management Journal*, 17.

GREEN, L. (1988): *The Authority of the State.* Oxford: Clarendon.

GUILLEN, M. F. (1994): *Models of Management: Work, Authority and Organization in a Comparative Perspective.* Chicago: University of Chicago Press.

GUTMANN, A., and D. THOMPSON (1996): *Democracy and Disagreement.* Cambridge (MA): The Belknap Press of Harvard University Press.

HAMPSHIRE, S. (1983): *Morality and Conflict.* Cambridge (MA): Harvard University Press.

HAMPTON, J. (1986): *Hobbes and the Social Contract Tradition.* Cambridge: Cambridge University Press.

HARDIN, R. (1982): *Collective Action.* Baltimore: Johns Hopkins University Press.

— (1987): "Does Might Make Right?," in *Authority Revisited: Nomos XXIX*, ed. by J. R. Penock, and J. W. Chapman. New York: New York University Press.

— (1998): "Trust in Government.," in *Trust and Governance.*, ed. by V. Braithwaite, and M. Levi. New York: Russel Sage Foundation.

HARE, R. M. (1972): "Rules of War and Moral Reasoning.," *Philosophy and Public Affairs*.

HARGREAVES HEAP, S., M. HOLLIS, B. LYONS, R. SUGDEN, and A. WEALE (1992): *The Theory of Choice; a Critical Guide*. Oxford: Basil Blackwell.

HART, H. L. A. (1968): *Punishment and Responsibility*. Oxford: Clarendon Press.

— (1982): *Essays on Bentham: Studies in Jurisprudence and Political Theory*. Oxford: Clarendon Press.

— (1983): *Essays in Jurisprudence and Philosophy*. Oxford: Clarendon Press.

HART, H. L. A., and T. HONORE (1959/1985): *Causation in the Law*. Oxford: Clarendon Press.

HODGSON, G. M. (1998): "The Approach of Institutional Economics," *Journal of Economic Literature*, 36, 166.

HOLLIS, M. (1982): "Dirty Hands", *British Journal of Political Science*, 12, 385-398.

— (1994): *The Philosophy of Social Science; an Introduction*. Cambridge: Cambridge University Press.

HONDERICH, T. (1988): *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes*. Oxford: Oxford University Press.

HOWARD, W. K. (1977): "Must Public Hands Be Dirty?," *Journal of Value Inquiry*, 11, 29-40.

HUME, D. (1739/1978): *A Treatise of Human Nature*. Oxford: Clarendon Press.

JACOBINES, G. M., and D. C. CROSON (2001): "Information Policy: Shaping the Value of Agency Relationships," *Academy of Management Review*, 26, 202-223.

WEMPE, J. F. D. B and, S. P. KAPTEIN, (forthcoming): *The Balanced Company*. Oxford: Oxford University Press.

KELSEN, H. (1945): *General Theory of Law and State*. Cambridge (MA): Harvard University Press.

KLANDERMAN, J. H. (1986): *Ratio, Wetenschap en Recht; een Onderzoek naar de Opvattingen van Wetenschap, Recht en de Grundnorm in de Reine Rechtslehre van Hans Kelsen*. Zwolle: Tjeenk Willink.

KNIGHT, F. H. (1960): "Political Responsibility in a Democracy.," in *Responsibility: Nomos Iii*, ed. by C. J. Friedrich. New York: The Liberal Art Press.

KORNHAUSER, L. A. (1992a): "Modelling Collegial Courts I: Path Dependence," *International Review of Law and Economics*, 12, 169-185.

— (1992b): "Modelling Collegial Courts II: Legal Doctrine," *International Review of Law and Economics*, 8, 441-470.

KORNHAUSER, L. A., and L. G. SAGER (1986): "Unpacking the Court," *Yale Law Journal*, 82.

— (1993): "The One and the Many; Adjudication in Collegial Courts," *California Law Review*, 81, 1-59.

KORSGAARD, C. (1996): *The Sources of Normativity*. Cambridge: Cambridge University Press.

KOSNIK, R. D. (1987): "Greenmail: A Study of Board Performance in Corporate Governance," *Administrative Science Quarterly*, 32, 163-185.

KRIPKE, S. (1982): *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Oxford: Blackwell.

KUHN, T. S. (1962/1970): *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

LADENSON, R. (1970): "Two Kinds of Rights," *Journal of Value Inquiry*.

— (1980/1990): "In Defense of a Hobbesian Conception of Law.," in *Authority*, ed. by J. Raz. New York: New York University Press.

LASSLET, P. (1963): "The Face to Face Society," in *Philosophy, Politics and Society*, ed. by P. Lasslet. Oxford: Basil Blackwell.

— (1963): "Philosophy, Politics and Society," Oxford: Basil Blackwell.

LEWIS, D. K. (1969): *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

LIPSEY, R. E. (1963): *Introduction to Positive Economics*. London & New York: Harper and Row.

LIST, C., and P. PETTIT (2000, September version): "The Aggregation of Reason; an Impossibility Result and Its Implications", Canberra.

LOCKE, J. (1960): *Two Treatises of Government*. Cambridge: University Press.

LUHMANN, N. (1988): "Familiarity, Confidence, Trust; Problems and Alternatives," in *Trust: Making and Breaking Cooperative Relations*, ed. by D. G. Gambetta. Oxford: Basil Blackwell.

LUKES, S. (1974): *Power: A Radical View*. London: Macmillan.

— (1979): "Power and Authority," in *A History of Sociological Analysis*, ed. by T. Bottomore, and R. Nisbet. London: Heinemann Educational.

— (1986): *Power*. Oxford: Blackwell.

— (1987): "Perspectives on Authority," in *Authority Revisited: Nomos XXIX*, ed. by J. R. Penock, and J. W. Chapman. New York: New York University Press.

MACCORMICK, N., and O. WEINBERGER (1986): *An Institutional Theory of Law: New Approaches to Legal Positivism*. Dordrecht: Reidel.

MACHIAVELLI, N. (1514/1995): *The Prince*. London: Everyman.

— (1516/1970): *The Discourses*. London: Penguin Books.

MACMAHON, C. (1994): *Authority and Democracy: A General Theory of Government and Management*. Princeton: Princeton University Press.

— (1995): "The Political Theory of Organizations and Business Ethics.," *Philosophy and Public Affairs*, 24, 292-313.

MACNEIL, I. (1974): "The Many Futures of Contract," *Southern California Law Review*, 47, 854-906.

MANDEVILLE, B. (1714/1989): *The Fable of the Bees*. London: Penguin Books.

MANN, M. (1997): "Has Globaliation Ended the Rise and Rise of the Nation-State?," *Review of International Political Economy*, 4, 472-496.

MARCH, J. G., and J. P. OLSEN (1989): *Rediscovering Institutions*. New York: Free Press.

— (1995): *Democratic Governance*. New York: Free Press.

MASTEN, S. E. (1993): "Transaction Costs, Mistakes, and Performance: Assessing the Importance of Governance," *Managerial and Decision Economics*, 14, 119-129.

MAY, L., and S. HOFFMAN (1991): "Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics," Savage (Maryland): Rowman & Littlefield Publishers.

MEGGINSON, W. L., and J. M. NETTER (2001): "From State to Market: A Survey of Empirical Studies on Privatization," *Journal of economic literature*, 39, 321-389.

MENARD, C. (2000): *Institutions, Contracts and Organizations; Perspectives from New Institutional Economics*. Cheltenham: Edward Elgar.

MINAR, E. H. (1990): *Philosophical Investigations 185-202: Wittgensteins Treatment of Following a Rule*. New York: Graland Pub.

MOE, T. M. (1984): "The New Economics of Organization," *American Journal of Political Science*, 28, 739.

MORAN, P., and S. GHOSHAL (1996): "Theories of Economic Organization: The Case for Realism and Balance," *Academy of Management Review*, 21, 58-72.

MORGENTHAU, H. J. (1948): *Politics among Nations: The Struggle for Power and Peace*. New York: Knopf.

MORTON, P. (1998): *An Institutional Theory of Law: Keeping Law in Place*. Oxford: Clarendon Press.

NAGEL, T. (1978): "Ruthlessness in Public Life," in *Public and Private Morality*, ed. by S. Hampshire. Cambridge: Cambridge University Press, 75-91.

— (1979): *Mortal Questions*. Cambridge: Canbridge University Press.

NOORDERHAVEN, N. (1995): "The Argumentational Texture of Transaction Cost Economics," *Organisation studies*, 16, 605-623.

NOOTEBOOM, B. (2002): "The Foundations of Trust", in *Institutionalised Sociality*, ed. by P. P. M. A. R. Heugens, J. Van Oosterhout, and J. Vromen: Edward Elger.

NORTH, D. C. (1990): *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

— (1991): "Institutions," *Journal of Economic Perspectives*, 5, 97-112.

NORTH, D. C., and R. P. THOMAS (1973): *The Rise of the Western World: A New Economic History*. Cambridge: Cambridge University Press.

NOWELL-SMITH, P. H. (1948): "Freewill and Moral Responsibility," *Mind*.

NUSSBAUM, M. C. (1986): *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. cambridge: Cambridge University Press.

PEEL, M. (1998): "Trusting Disadvantaged Citizens," in *Trust and Governance*, ed. by V. Braithwaite, and M. Levi. New York: Russel Sage Foundation.

PENNOCK, J. R. (1960): "The Problem of Responsibility," in *Responsibility: Nomos III*, ed. by C. J. Friedrich. New York: The Liberal Art Press.

PENNOCK, J. R., and J. W. CHAPMAN (1985): *Criminal Justice: Nomos XVVII*, New York: New York University Press.

— (1987): *Authority Revisited: Nomos XXIX*, New York: New York University Press.

PENNOCK, R. J., and J. W. CHAPMAN (1968): "Representation: Nomos X," New York: Atherton Press.

PERROW, C. (1986): *Complex Organizations: A Critical Essay*. Glenview, IL: Scott, Foresman and Company.

PETERS, B. G. (1999): *Institutional Theory in Political Science: The "New Institutionalism"*. London: Pinter.

PETTIT, P. (1990): "The Reality of Rule-Following.," *Mind*, 99.

— (1993): *The Common Mind: An Essay on Psychology, Society, and Politics*. New York: Oxford University Press.

— (1995a): "The Virtual Reality of Homo Economicus," *Monist*, 78, 308-329.

— (1995b): "The Cunning of Trust," *Philosophy and Public Affairs*, 24, 202-225.

— (1996a): "Institutional Design and Rational Choice," in *The Theory of Institutional Design*, ed. by R. E. Goodin. Cambridge: Cambridge University Press.

— (1996b): "Freedom and Antipower," *Ethics*, 106, 576.

— (1997): *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.

— (1998): "Republican Theory and Political Trust," in *Trust and Governance*, ed. by V. Braithwaite, and M. Levi. New York: Russel Sage Foundation.

— (2001): "Groups with Minds of Their Own (Manuscript, April version)," Canberra.

PIGDEN, C. R. (1991): "Naturalism.," in *Companion to Ethics.*, ed. by P. Singer. Oxford: Blackwell.

PITKIN, H. F. (1967): *The Concept of Representation*. Berkeley, CA.: University of California Press.

— (1968): "Comment: The Paradox of Representation", in: *Representation: Nomos X*, ed. by J. R. Pennock, and J. W. Chapman. New York: Atherton Press.

— (1972): *Wittgenstein and Justice: On the Significance of Ludwig Wittgenstein for Social and Political Thought*. Berkely: University of California Press.

PLATO (1974): *The Republic*. Harmondsworth: Penguin.

POWELL, W. W., and J. P. DIMAGGIO (1991): "The New Institutionalism in Organizational Analysis," Chicago: University of Chicago Press.

PRATT, J. W., and R. J. ZECKHAUSER (1985): "Principals and Agents: An Overview," in *Principals and Agents: The Structure of Business*, ed. by J. W. Pratt, and R. J. Zeckhauser. Boston: Harvard Business Scholl Press, 1-35.

— (1985): *Principals and Agents: The Structure of Business*. Boston: Harvard Business School Press.

PRESTON, L. E. (1998): "Agents, Stewards, and Stakeholders," *Academy of Management Review*, 23, 9-13.

PRZEWORSKI, A., S. C. STOKES, and B. MANIN (1999): "Democracy, Accountability and Representation," Cambridge: Cambridge University Press.

RAWLS (1971): *A Theory of Justice*. Cambridge, MA: Harvard University Press.

RAZ, J. (1975/1999): *Practical Reason and Norms*. Oxford: Oxford University Press.

— (1986): *The Morality of Freedom*. Oxford: Clarendon Press.

— (1990): *Authority*. New York: New York University Press.

— (1994): *Ethics in the Public Domain: Essays in the Morality of Law and Politics*. Oxford / New York: Clarendon press / Oxford University Press.

ROBERTSON, D. H. (1923/1943): *The Control of Industry*. London & Cambridge: Nisbet and Cambridge University Press.

ROBERTSON, P. (1999): "Authority and Control in Industry; Theoretical and Empirical Perspectives," London & New York: Routledge.

ROSS, D. (1930): *The Right and the Good*. Oxford: Clarendon.

SALOMON, L. M. (1994): "The Rise of the Non-Profit Sector," *Foreign Affairs*, 74.

SALOMON, L. M., H. K. ANHEIER, and ASSOCIATES (1999): "The Emerging Sector Revisited: A Summary," Baltimore: Center for Civil Society Studies, Johns Hopkins University.

SARTRE, J. P. (1947): *Les Mains Sales*. Paris: Editions Gallimard.

SCANLON, T. (1999): *What We Owe to Each Other*. Cambridge (Mass.): Harvard University Press.

SCHELLING, T. C. (1960): *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.

SCHLEIFER, A., and R. W. VISHNY (1997): "A Survey of Corporate Governance," *The Journal of Finance*, LII, 737-782.

SCHMITT, C. (1922/1985): *Political Theology: Four Chapters on the Concept of Sovereignty*. Cambridge: MIT Press.

— (1932/1996): *The Concept of the Political*. Chicago: University of Chicago Press.

SCHWEDER, R. A., and R. A. LEVINE (1984): *Culture Theory: Essays on Mind, Self, and Emotion*. Cambridge: Cambridge university Press.

SCOTT, W. R. (1994): "Institutions and Organizations: Towards a Theoretical Synthesis," in *Institutional Environments and Organizations; Structural Complexity and Individualism*, ed. by W. R. Scott, and J. W. Meyer. Thousand Oaks: Sage Publishers.

— (1995): *Institutions and Organisations*. Thousend Oaks, CA: Sage.

— (1998): *Organizations: Rational, Natural, and Open Systems*. New York: Prentice Hall.

SCOTT, W. R., and J. W. MEYER (1994): "Institutional Environments and Organizations; Structural Complexity and Individualism," Thousand Oaks: Sage Publications.

SEARLE, J. R. (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

— (1983): *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

— (1995): *The Construction of Social Reality*. New York: Free Press.

SELLARS, W. (1963): *Science, Perception and Reality*. London: Routledge and Kegan Paul.

SELZNICK, P. (1949): *TVA and the Grass Roots*. Berkeley: University of California Press.

— (1957): *Leadership in Administration: A Sociological Interpretation*. New York: Harper and Row.

SEN, A. (1998): "Introduction," in *Economics, Values and Organization.*, ed. by A. Ben-Ner, and L. Putterman. Cambridge: Cambridge University Press.

SHANKER, S. G. (1984): "Sceptical Confusions About Rule-Following," *Mind*, 93, 423.

SHARMA, A. (1997): "Professional as Agents: Knowledge Asymmetry in Agency Exchange," *Academy of Management Review*, 22, 758-798.

SIE, M. M. S. K. (1999): *Responsibility, Blameworthy Actions & Normative Disagreements: A Defense of Practical Semi-Compatibilism.* Utrecht: Quaestiones Infinitae, Zeno.

SIMON, H. A. (1945/1998): *Administrative Behavior.* New York: Free Press.

— (1991): "Organizations and Markets," *Journal of Economic Perspectives*, 5, 25-44.

SINGER, P. (1991): *A Companion to Ethics*, Oxford: Blackwell.

SOLOMON, R. C. (1992): *Ethics and Excellence: Cooperation and Integrity in Business.* New York: Oxford University Press.

STOCKER, M. (1990): *Plural and Conflicting Values.* Oxford: Clarendon.

STONE, C. D. (1985): "A Comment on "Criminal Responsibility in Government"," New York: New York University Press.

SUGDEN, R. (1998): "Normative Expectations: The Simultaneous Evolution of Institutions and Norms.," in *Economics, Values and Organization*, ed. by A. Ben-Ner, and L. Putterman. Cambridge: Cambridge University Press.

THOMPSON, D. F. (1985): "Criminal Responsibility in Government," in *Criminal Justice: Nomos XXVII*, ed. by J. R. Pennock, and J. W. Chapman. New York: New York University Press.

— (1987): *Political Ethics and Public Office.* Cambridge (Mass.): Harvard University Press.

— (1999): "The Institutional Turn in Professional Ethics," *Ethics & Behavior*, 9, 109-118.

TOLBERT, P. S., and L. G. ZUCKER (1996): "The Institutionalization of Institutional Theory.," in *Handbook of Organization Studies*, ed. by S. R. Clegg, C. Hardy, and W. Nord. London: Sage.

TUOMELA, R. (1995): *The Importance of Us: A Philosophical Study of Basic Social Notions.* Stanford, Calif.: Stanford University Press.

ULLMANN-MARGALIT, E. (1977): *The Emergence of Norms.* Oxford: Clarendon Press.

VAN GUNSTEREN, H. R. (1974): *Denken over Politieke Verantwoordelijkheid.* Alphen aan de Rijn: Samson.

— (1976): *The Quest for Control: A Critique of the Rational-Central-Rule Approach in Public Affairs.* London: Wiley.

VAN INWAGEN, P. (1983): *An Essay on Free Will.* Oxford: Clarendon Press.

VAN OOSTERHOUT, J., and H. KRIBBE (1995): "Tussen Politiek en Digitalisering: een bijdrage aan het Debat over het Nederlands Drugsbeleid.," Dordrecht: Stichting de Hoop.

VAN ROERMUND, B. (1997): *Het Verdwijnpunt van de Wet: een Opstel over Symboolwerking van Wetgeving.* Deventer: Tjeenk Willink.

VAN TRAA, M. (1996): *Inzake Opsporing: Enquete Commissie Opsporingsmethoden.* Den Haag: Sdu Uitgevers.

VAN WILLIGENBURG, T. (2000, manuscript): "Appropriate Emotions; Moral Sentiments and Ethical Objectivity.," Rotterdam: Erasmus University

VROMEN, J. J. (forthcoming): "Social Contracts Evolving," Rotterdam: Erasmus University

WALLACE, M. D., and J. D. SINGER (1970): "Intergovernmental Organisations in the Global System," *International Organization*, 24, 239.

WALLACE, W. (1969): *Sociological Theory.* London: Heinemann.

WALZER, M. (1973): "Political Action: The Problem of Dirty Hands," *Philosophy and Public Affairs*, 2, 160-180.

— (1997): *On Toleration.* New Haven: Yale University Press.

WEBER, M. (1919/1982): *Politik als Beruf.* Berlin: Duncker & Humblot.

— (1947): *The Theory of Social and Economic Organization.* New York: Free Press.

— (1949): *The Methodology of the Social Sciences.* New York: Free Press.

— (1968): *Economy and Society: An Outline of Interpretive Sociology.* New York: Bedminster Press.

WEMPE, J. F. D. B. (1998): *Market and Morality: Business Ethics and the Dirty and Many Hands Dilemma.* Delft: Eburon.

WHITE, H. C. (1985): "Agency as Control," in *Agents and Principals: The Structure of Business*, ed. by J. W. Pratt, and R. J. Zeckhauser. Boston: Harvard Bussines School Press.

WILLIAMS, B. (1973): *Problem of the Self: Philosophical Papers, 1956-1972.* Cambridge: Cambridge University Press.

— (1978): "Introduction," in *Concepts and Categories: Philosophical Essays*, ed. by I. Berlin. London: Hogarth Press.

— (1981): *Moral Luck: Philosophical Papers, 1973-1980.* Cambridge: Cambrifge University Press.

WILLIAMSON, O. E. (1975): *Markets and Hierarchies: Analysis and Antitrust Implications.* New York: Free Press.

— (1979): "Transaction Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 22, 233-261.

— (1981): "The Economics of Organization: The Transaction Cost Approach," *American Journal of Sociology*, 87, 548-577.

— (1985): *The Economic Institutions of Capitalism.* New York: Free Press.

— (1991a): "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives," *Administrative Science Quarterly*, 36, 269-296.

— (1991b): "Strategizing, Economizing and Economic Organization," *Strategic Management Journal*, 12, 75-94.

— (1991c/1996): "Economic Institutions; Spontaneous and Intentional Governance," in *The Mechanisms of Governance*, ed. by O. E. Williamson. Oxford: Oxford University Press.

— (1996): "Economic Organization: The Case for Candor," *Academy of Management Review*, 21, 48-57.

— (2000): "The New Institutional Economics: Taking Stock, Looking Ahead", *Journal of Economic Literature*, 38, 595-613.

WINCH, P. (1958): *The Idea of a Social Science and Its Relations to Philosophy*. London: Routledge and Keegan Paul.

WINSTON, K. (1994): "Necessity and Choice in Political Ethics," in *Professional Ethics and Social Responsibility*, ed. by D. E. Wueste. Lanham: Rowman & Littlefield.

WOLF, S. (1980): "Asymmetrical Freedom," *The Journal of Philosophy*, 77, 151.

— (1985): "The Legal and Moral Responsibility of Organizations," in *Nomos XXVII: Criminal Justice*, ed. by J. R. C. Pennock, J.W. New York: New York University Press.

— (1990): *Freedom within Reason*. New York: Oxford University Press.

— (1992): "Two Levels of Pluralism," *Ethics*, 102, 785.

WOLIN, S. S. (1960): *Politics and Vision: Continuity and Innovation in Western Political Thought*. Boston: Little, Brown.

WUESTE, D. E. (1994): "Professional Ethics and Social Responsibility," Lanham: Rowman & Littlefield.

YOUNG, R. (1991): "The Implications of Determinism.," in *A Companion to Ethics.*, ed. by P. Singer. Oxford: Basil Blackwell.

Summary
(In Dutch)

Hoewel autoriteit in strijd is met de moderne idealen van rationaliteit, autonomie en individuele verantwoordelijkheid, is (bestuurlijke) autoriteit alomtegenwoordig in de moderne samenleving (bijvoorbeeld: de staat, geografische en functionele autoriteiten, internationale organisaties, de onderneming, NGO's, enzovoort). Deze paradox is aanleiding om in deze studie het begrip autoriteit nader te onderzoeken. In hoofdstuk I worden achtereenvolgens de (a) achtergrond, (b) inzet en (c) concrete vraagstellingen van deze studie uiteengezet.

De historisch-theoretische achtergrond waartegen het begrip autoriteit in eerste instantie dient te worden begrepen is die van de premoderne samenleving. Autoriteit was daarin de dominante bestuursvorm in zowel de publieke als de private sfeer. Een verklaring van de dominantie en persistentie van autoriteit in de premoderne samenleving is dat (bestuurlijke) autoriteit was ingebed in een unieke constellatie van historisch-theoretische condities die deze samenleving kenmerkten. De conjunctie van (a) een overwegend metafysisch wereldbeeld, (b) een centrale rol van traditie en (c) grote machtsverschillen maakten autoriteit tot een min of meer natuurlijke en zeer veerkrachtige bestuursvorm in de premoderne samenleving. Deze conjunctie van condities werd echter ontbonden door processen van modernisering. De (a) verwetenschappelijking van het wereldbeeld, (b) ont-tovering van de premoderne normatieve hiërarchische orde en (c) de feitelijke opkomst van spontane coördinatie mechanismen in zijn algemeenheid, en de markt in het bijzonder, ondermijnden de natuurlijke gehoorzaamheid die kenmerkend is voor autoriteit, en leidden tot de geboorte van het vraagstuk van de legitimiteit, dat wil zeggen: de vooronderstelling dat men autoriteit te bevragen alvorens te gehoorzamen.

Het bevragen van autoriteit noopt vooraleerst tot een adequate verklaring ervan. Drie verklaringen van autoriteit staan centraal in de wetenschappelijke literatuur. Ten eerste is autoriteit een beproefd en effectief middel om te kunnen omgaan met de grote mate van arbeidsdeling en specialisatie in de moderne samenleving. Deze verklaring schiet echter tekort omdat hiermee slechts *theoretische* autoriteit verklaard kan worden, en niet de *praktische* autoriteit die kenmerkend is voor moderne bestuursvormen. De focus van deze studie is derhalve op praktische autoriteit. Een tweede verklaring is dat autoriteit een efficiënte oplossing van coördinatieproblemen kan bieden. Ook deze verklaring schiet echter tekort omdat autoriteit ook, en wellicht juist daar aanwezig en functioneel is waar problemen van motivatie het hoofd geboden dienen te worden. Een derde verklaring begrijpt autoriteit derhalve als een antwoord op dergelijke problemen. Ook deze laatste verklaring is echter problematisch omdat hij te sterk steunt op de mogelijkheden van rationele controle, en daarnaast geen verklaring geeft voor het normatieve karakter van autoriteit en de overwegend vrijwillige gehoorzaamheid aan autoriteit waardoor autoriteit gekenmerkt

wordt in zowel de premoderne als de moderne samenleving. Beargumenteerd wordt dat autoriteit derhalve niet uitsluitend vanuit een zogenoemd derde-persoon's perspectief begrepen kan worden. Een fenomenologisch eerste-persoon's perspectief is onontbeerlijk om autoriteit adequaat te kunnen duiden.

Het openen van de "black box" van autoriteit behelst allereerst dat autoriteit begrepen wordt in termen van *redenen* in plaats van uitsluitend oorzaken. Het gaat hier, ten tweede, om een bijzonder soort redenen, dat wil zeggen, om zogenoemde tweede orde '*uitsluitende* redenen', die de pretentie hebben wat voor eerste orde redenen een actor ook mag hebben, te overtroeven. Dus het bevel "zwijg!" dient in deze visie begrepen te worden als een hogere orde reden om te zwijgen, die de eerste orde redenen die ik heb om te spreken overtroeft. Uit dit voorbeeld blijkt tevens dat de actor zo'n tweede orde 'uitsluitende reden' alleen zal accepteren als hij erkent dat degene die het bevel geeft het recht heeft om dat te doen. Dit laatste maakt duidelijk dat de vraag naar de legitimatie van autoriteit onlosmakelijk verbonden is met de vraag hoe autoriteit werkt, en dus uiteindelijk met de vraag wat autoriteit is. De inzet van deze studie is de legitimatievraag, die besloten ligt in de betekenis van autoriteit, van een antwoord te voorzien.

Joseph Raz heeft een interessant antwoord gegeven op deze legitimatievraag. Kort gezegd houdt dit antwoord in dat de autoriteit van een persoon X gerechtvaardigd is als de bevelen van X zijn ondergeschikten helpen om *beter* te doen waar ze, los van wat X wil, zelf al reden toe hebben. Omdat hier autoriteit uiteindelijk in dienst staat van degenen die er aan ondergeschikt zijn, wordt deze conceptie van autoriteit ook wel de "serviceconceptie" van autoriteit genoemd. Het is belangrijk om op te merken dat deze serviceconceptie van autoriteit een *normatief* begrip van autoriteit behelst. Autoriteit wordt hier begrepen in termen van zijn normatieve consequenties: als X het recht heeft om mij bevelen te geven dan impliceert dat voor mij de plicht om te gehoorzamen. Op deze manier kan ook een scherp onderscheid worden gemaakt tussen autoriteit en macht. Ik heb immers op geen enkele manier de plicht te gehoorzamen als X macht over mij uitoefent. Deze studie behelst in belangrijke mate een kritiek op deze (normatieve) serviceconceptie van autoriteit. Deze kritiek betreft in essentie drie problemen die in vier afzonderlijke hoofdstukken aan de orde worden gesteld.

Het eerste probleem staat centraal in hoofdstuk II, en betreft het door Raz als essentieel aangemerkte onderscheid tussen zijn *normatieve* serviceconceptie van autoriteit, enerzijds, en de klassieke Weberiaanse opvatting van autoriteit als legitieme macht, anderzijds. Zoals gezegd ligt het verschil tussen deze twee in de vermeende afwezigheid van normatieve consequenties in het geval van de laatste. Een bevel kan men immers als norm begrijpen, terwijl dat niet het geval is voor de uitoefening van macht, los van de vraag of deze machtsuitoefening gelegitimeerd is of niet. De conclusie van dit hoofdstuk is echter dat dit door Raz essentieel geachte onderscheid

onhoudbaar is. In grote lijnen is de argumentatie achter deze conclusie dat normen pas normatief, dat wil zeggen: (a) *richtinggevend* voor het handelen, en (b) *bindend* zijn, als ze *effectief* zijn, dat wil zeggen: als ze in het algemeen worden gevolgd en nageleefd. Omdat sommige normen niet uit zichzelf effectief zijn, en er derhalve interventie (machtsuitoefening) nodig is om deze effectiviteit af te dwingen, berust de normativiteit van sommige normen in laatste instantie op de macht die ze 'verwerkelijkt'. Het gevolg is dat macht in bepaalde gevallen constitutief is voor (normatieve) autoriteit, en ondergeschikten dus redenen voor het handelen kan geven die ze zonder deze machtsuitoefening niet zouden hebben. Het onderscheid tussen autoriteit en (legitieme) macht is derhalve onhoudbaar, evenals Raz' antwoord op de legitimatievraag die ten grondslag ligt aan zijn begrip van autoriteit. De legitimatievraag dient dus geherformuleerd te worden in termen van het verschil tussen legitieme macht en macht op zich.

Dit tweede onderscheid is aan de orde in hoofdstuk III. Zoals gezegd behelst de geherformuleerde legitimatievraag het onderscheid tussen macht op zich en legitieme macht. Autoriteit, nu in de klassiek Weberiaanse betekenis van legitieme macht, kan in de moderne samenleving begrepen worden als macht begrensd door geldige en effectieve normen. De conclusie van hoofdstuk III is echter dat ook deze interpretatie van autoriteit onhoudbaar is. Het probleem zit in de notie dat legitieme macht begrensd wordt door geldige en effectieve normen. Dit uitgangspunt is in strijd met het eeuwen oude politiek-theoretische inzicht dat personen in een positie van autoriteit soms slecht moeten zijn om goed te kunnen doen, dat wil zeggen, dat er soms gehandeld moet worden in strijd met geldige en effectieve normen omwille van een groter collectief of maatschappelijk goed. Dit probleem, dat in de literatuur bekend staat als het 'vuile handen probleem', is een evidente anomalie met betrekking tot een klassiek Weberiaanse autoriteitsopvatting. Na het vuile handen probleem nader conceptueel te hebben geduid in termen van (a) normatief conflict (noodzakelijke voorwaarde), en (b) representatief handelen (voldoende voorwaarde tegen de achtergrond van normatief conflict), wordt beargumenteerd dat autoriteit vaak juist daar aanwezig en functioneel is waar vuile handen gemaakt moeten worden, dat wil zeggen, waar (mogelijke) normatieve conflicten moeten worden beheerst, en zonodig doorbroken, middels sociale representatie constructies (bijvoorbeeld: adjudicatieve, legislatieve en executieve autoriteit). Naast het oplossen van coördinatieproblemen en het bestrijden van motivatieproblemen is het beheersen en doorbreken van normatieve conflicten onmiskenbaar een centrale functie van autoriteit in de moderne samenleving. De overkoepelende conclusie van dit hoofdstuk is dat er geen *theoretisch* antwoord is op de legitimatievraag die voorafgaat aan elk antwoord op de vraag wat autoriteit nu precies is. Geclaimd wordt dat de legitimatievraag derhalve dient te worden begrepen als een *praktisch* probleem. Dat laatste impliceert niet alleen dat de vraag naar legitimiteit geen enkelvoudig en eenduidig antwoord heeft, maar ook dat de legitimiteit van autoriteit eigenlijk bij voortduring moet worden veilig gesteld.

In de hoofdstukken IV en V wordt vanuit dit *pragmatische* perspectief een meer praktische oplossing voor de legitimiteitsvraag verkend. Onderzocht wordt of een notie van verantwoordelijkheid voor autoriteit wellicht aanknopingspunten biedt om genoemde problematiek tegemoet te treden. Deze zoektocht wordt ingezet met de suggestie in het achterhoofd dat autoriteit en verantwoordelijkheid binnen een hiërarchie zich wellicht kunnen verhouden zoals vraag een aanbod dat doen binnen een markt. Omdat een *concept* van autoriteit voor verantwoordelijkheid niet in de literatuur voorhanden is, worden in hoofdstuk IV vooraleerst de mogelijkheidsvoorwaarden voor een dergelijk begrip onderzocht. Na een inventarisatie van de belangrijkste feitelijke betekenissen van verantwoordelijkheid zoals deze besloten liggen in de belangrijkste instituties van de moderne Westerse samenleving, worden drie fundamentele concepties van verantwoordelijkheid geïdentificeerd. Op basis van deze drie concepties worden vervolgens vier noodzakelijke voorwaarden geformuleerd die moeten zijn vervuld om het concept verantwoordelijkheid van toepassing te laten zijn. In het resterende deel van hoofdstuk IV wordt beargumenteerd dat het niet waarschijnlijk is dat aan deze vier noodzakelijke voorwaarden in de moderne samenleving wordt voldaan. Dit geldt niet alleen voor een concept van verantwoordelijkheid voor autoriteit, maar veeleer voor het concept verantwoordelijkheid op zich. Betoogd wordt dat een aantal fundamentele problemen de eenduidige toepassing van een concept van verantwoordelijkheid in de weg staan. De problemen die aan de orde komen zijn onder andere: (a) het probleem van toeschrijving, (b) het probleem van de praktische noodzakelijkheid en (c) het probleem van 'constituency'. Dit hoofdstuk eindigt met de paradoxale conclusie dat hoewel het niet waarschijnlijk is dat het concept verantwoordelijkheid van toepassing is in de moderne samenleving, praktijken verantwoordelijkheid daarin feitelijk alomtegenwoordig zijn.

In hoofdstuk V wordt deze paradox onschadelijk gemaakt door de presumptie te verlaten dat verantwoordelijkheid begripsmatig of conceptueel begrepen dient te worden. Een alternatief wordt verkend in de vorm van een *institutioneel* begrip van verantwoordelijkheid. Na de notie van institutie nader te hebben gepreciseerd, worden drie afzonderlijke stromingen geïdentificeerd in de (positieve) institutionele theorie: een (a) regulatieve, een (b) normatieve en een (c) cognitieve stroming. Deze drie stromingen worden vervolgens samengebracht in een algemene institutionele theorie, volgens welke deze drie stromingen drie afzonderlijke mechanismen beschrijven die in principe afzonderlijk, maar in de realiteit vaker gezamenlijk, functioneren in de constitutie van instituties. De drie concepties van verantwoordelijkheid die in het voorgaande hoofdstuk zijn geïdentificeerd corresponderen grofweg met deze drie mechanismen. Beargumenteerd wordt waarom een institutioneel begrip van verantwoordelijkheid zowel theoretisch als praktisch superieur is aan conceptueel begrip van verantwoordelijkheid. Door het te contrasteren met het in de literatuur dominante, maar tegelijkertijd fel bekritiseerde

'agency model' van autoriteitsrelaties, geef ik een grove schets van hoe een institutioneel begrip van verantwoordelijkheid voor autoriteit er uit zou kunnen zien. Dit hoofdstuk eindigt met de conclusie dat een algemene institutionele theorie van verantwoordelijkheid voor autoriteit een omvattender analyse van autoriteitsrelaties biedt dan het 'agency model', omdat deze laatste uitsluitend gebaseerd is op het functioneren van een enkel constitutief mechanisme (het regulatieve).

In hoofdstuk VI worden dertien centrale thesen geformuleerd, die niet alleen samenvatten en concluderen wat in voorgaande hoofdstukken is behandeld en beargumenteerd, maar gezamenlijk ook een argumentatie constitueren tegen een enkelvoudige en eenduidige *theorie* van autoriteit. De hoofdconclusie van deze studie is dat er geen eenduidig begrip van autoriteit bestaat, omdat de legitimatievraag niet op theoretisch niveau kan worden beantwoord. In lijn met de voorgaande analyse van verantwoordelijkheid wordt betoogd dat autoriteit veeleer als institutie moet worden begrepen. Een institutioneel perspectief laat niet alleen zien dat de legitimatievraag in de praktijk op verschillende manieren beantwoord kan worden en dat legitimiteit steeds opnieuw moet worden veilig gesteld, maar ook dat gehoorzaamheid aan autoriteit verschillende oorzaken kan hebben (waaronder macht). Geclaimd wordt dat verantwoordelijkheid voor autoriteit in laatste instantie de legitimiteit ervan garandeert. Omdat autoriteit en verantwoordelijkheid aan elkaar gekoppelde instituties zijn, en verantwoordelijkheid in de context van bestuurlijke autoriteit *feitelijk* niet altijd afdoende geïnstitutionaliseerd is, zijn institutioneel ontwerp en interventie in laatste instantie de peilers waarop de claim van legitimiteit, welke besloten ligt in elk begrip van autoriteit, gewaarborgd kan worden.

Curriculum Vitae

Johannes (Hans) van Oosterhout (1965) was born in Kampen and raised in Voorburg, where he received both primary and secondary education. After having served in the military, he studied at Leiden University where he received his MA in political science (Cum Laude) in august 1993. Since October 1993 he has been an assistant professor at the Faculteit Bedrijfskunde/Rotterdam School of Management, Erasmus University Rotterdam, where he teaches courses in philosophy of science, business–society management and the theory of institutions. His research interests include: (1) the institutions of governance, (2) institutional theory and design, (3) and the normative and ontological theory of institutions and organizations.

ERIM PH.D. SERIES

RESEARCH IN MANAGEMENT

Title:	**Operational Control of Internal Transport**
Author:	J. Robert van der Meer
Promotor(es):	Prof.dr. M.B.M. de Koster, Prof.dr.ir. R. Dekker
Defended:	September 28, 2000
Series number:	1
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-5892-004-6

Title:	**Quantitative Models for Reverse Logistics**
Author:	Moritz Fleischmann
Promotor(es):	Prof.dr.ir. J.A.E.E. van Nunen, Prof.dr.ir. R. Dekker, dr. R. Kuik
Defended:	October 5, 2000
Series number:	2
Published:	Lecture Notes in Economics and Mathematical Systems, Volume 501, 2001, Springer Verlag, Berlin,
ISBN:	3540 417 117

Title:	**Optimization Problems in Supply Chain Management**
Author:	Dolores Romero Morales
Promotor(es):	Prof.dr.ir. J.A.E.E. van Nunen, dr. H.E. Romeijn
Defended:	October 12, 2000
Series number:	3
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-9014078-6

Title:	**Layout and Routing Methods for Warehouses**
Author:	Kees Jan Roodbergen
Promotor(es):	Prof.dr. M.B.M. de Koster, Prof.dr.ir. J.A.E.E. van Nunen
Defended:	May 10, 2001
Series number:	4
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-5892-005-4

Title:	**Rethinking Risk in International Financial Markets**
Author:	Rachel Campbell
Promotor(es):	Prof.dr. C.G. Koedijk
Defended:	September 7, 2001
Series number:	5
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-5892-008-9

Title:	**Labour flexibility in China's companies: an empirical study**
Author:	Yongping Chen
Promotor(es):	Prof.dr. A. Buitendam, Prof.dr. B. Krug
Defended:	October 4, 2001
Series number:	6
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-5892-012-7

Title:	**Strategic Issues Management: Implications for Corporate Performance**
Author:	Pursey P. M. A. R. Heugens
Promotor(es):	Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. C.B.M. van Riel
Defended:	October 19, 2001
Series number:	7
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-5892-009-7

Title:	**Beyond Generics: A closer look at Hybrid and Hierarchical Governance**
Author:	Roland F. Speklé
Promotor(es):	Prof.dr. M.A. van Hoepen RA
Defended:	October 25, 2001
Series number:	8
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-5892-011-9

Title:	**Interorganizational Trust in Business to Business E-Commerce**
Author:	Pauline Puvanasvari Ratnasingam
Promotor(es):	Prof.dr. K. Kumar, Prof.dr. H.G. van Dissel
Defended:	November 22, 2001
Series number:	9
Published:	ERIM Ph.D. series Research in Management
ISBN:	90-5892-017-8

Title: **Outsourcing, Supplier-relations and Internationalisation: Global Source Strategy as a Chinese puzzle**
Author: Michael M. Mol
Promotor(es):   Prof.dr. R.J.M. van Tulder
Defended: December 13, 2001
Series number: 10
Published: ERIM Ph.D. series Research in Management
ISBN: 90-5892-014-3

Title: **The Business of Modularity and the Modularity of Business**
Author: Matthijs J.J. Wolters
Promotor(es):   Prof. mr. dr. P.H.M. Vervest, Prof. dr. ir. H.W.G.M. van Heck
Defended: February 8, 2002
Series number: 11
Published: ERIM Ph.D. series Research in Management
ISBN: 90-5892-020-8

The Quest for Legitimacy:
On Authority and Responsibility in Governance

Although authoritative governance is ubiquitous in modern society, the nature of authority is one of the most neglected and understudied topics in economic, political and organizational theory today. This study aims to correct for this lacuna. Its main conclusion is that there is no such thing as an unambiguous concept of authority. Its is argued that both authority and responsibility in governance should be conceived in institutional rather than conceptual terms, and that the quest for legitimacy that is indissolubly tied up with any understanding of authority, ultimately involves a problem of institutional design.

ERIM

ERIM — ERASMUS RESEARCH INSTITUTE OF MANAGEMENT