# PRACTICAL REASON AND THE CONDITIONS OF AGENCY

by

Douglas Lavin

B.A., Philosophy, Colgate University, 1991

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Douglas Lavin

It was defended on

December 9, 2004

and approved by

Stephen Engstrom, Associate Professor of Philosophy

John McDowell, University Professor of Philosophy

Jennifer Whiting, Jackman Professor of Philosophy, University of Toronto

Michael Thompson, Associate Professor of Philosophy
(Dissertation Director)

PRACTICAL REASON AND THE CONDITIONS OF AGENCY

Douglas Lavin, Ph.D.

University of Pittsburgh, 2004

How ought one to act? What is action? This dissertation is about how far an answer to the second question can take us towards an answer to the first. Many philosophers think that an answer to the metaphysical question about the nature of action can take us very far towards an answer to the ordinary question about how to act. There are two popular ways of developing this idea. According to *neo-Kantianism*, agency presupposes the capacity to engage in non-instrumental forms of practical thought. According to *neo-Humeanism*, agency is limited to the capacity to engage in instrumental thought. I criticize each and offer a better alternative. I argue that non-instrumental practical thought is not necessary for agency but still possible. Attempts to answer "How ought one to act?" by answering "What is action?" are attempts to explain how something ought to be through an account of what something is. Neo-Kantians and neo-Humeans focus on action because they think this: what practical reason requires of some agent it requires of all conceivable agents. Action seems to be the best place to look to ground norms with the relevant scope. But in order to combine the insights of each—namely that non-instrumental practical thought is not necessary, though still possible—we have to give up on this conception of the scope of practical requirements. And if we do, we are left with the task of finding other features of ourselves, say, our humanity, to ground the robust standards of moral life.

## PREFACE

I have been at this for a long time. It is difficult to let it go. Many debts to many people have been accumulated: what extraordinary luck to owe so much to so many. Steve Engstrom, John McDowell, Michael Thompson, and Jennifer Whiting have guided this project with patience, great insight and concern, and they have guided me by their example. I am deeply grateful.

Even before I asked him to direct my dissertation and he told me instead to "Go luminary!", Michael Thompson has shared without restraint his ideas, imagination, and clarity of mind, his friendship and humor. I have been utterly transformed by all of it. Like children who want to have the moon, some, including myself, come to graduate school wanting to have such an education. These are unreasonable hopes. And I have had wonderful fortune.

I would not enjoy and would not be able to do philosophy without others to talk to. There are great riches in Pittsburgh! I have discussed every bit of this dissertation with Matt Boyle. It would be much worse were it not for the continuous infusion of his good judgment; and I would be much worse without a friend who happily looked at the really awful bits. For countless, alternately crushing and uplifting conversations, my warmest thanks to Donald Ainsle, Alp Aker, Chrisoula Andreou, Eric Brown, Mark Criley, Umut Ergun, Anton Ford, David Gauthier, Matthias Hasse, Ben Laurence, Hans Lottenbach, Eric Marcus, Ram Neta, Doug Patterson, Sebastian Rödl, Kieran Setiya, Lionel Shapiro, Tara Bray Smith, David Sobel, Sergio Tennenbaum, Matt Weiner and Herb Wilson. To the members of the Dead Dissertation Club—carpe dissertation!

This dissertation is dedicated, with love, to my family, which is always there.

## TABLE OF CONTENTS

# 1. ANALYTICAL PHILOSOPHY OF PRACTICAL REASON

It's right to be a minimalist, in a way.

Philippa Foot

## *1.1 Parts of the soul*

We are frequently told, perhaps too frequently, that Plato divides the soul into *parts*. In an extended bit of stage setting, I'd like to traverse this well worn territory in order to point to the concepts framing this dissertation—that of *a kind of practical capacity*, *a form of agency*, *a form of desire*, and *a form of action explanation*. After this is on our horizon, we will be in a position to properly approach the question which provides the dissertation's unity—a question concerning the very conditions of agency, a question about which *parts* of the soul something can and must have if it acts at all.

As Plato has it, a soul is a capacity—in particular a capacity for self-movement (*Phaedrus* 245e).[1] In our philosophical climate the expressions "self-mover" and "moving oneself" are likely to bring to mind only the doers of intentional action and acting intentionally itself. However, in Plato's hands self-mover and self-movement are much more abstract concepts than this rather sublime determination of them. For Plato, "any body that has [only] an external source of motion is soulless, but a body deriving its motion from a source within itself is animate or *besouled*" (*Phaedrus* 245e). So, to be a living thing is to have a soul or to be a self-mover, and what a living thing *does* is an expression of a soul, or self-movement. Of course,

[1] Plato, *Phaedrus*, trans. A. Nehamas and P. Woodruff (Indianapolis: Hackett Publishing, 1995). All references to Plato's works will be given in the text.

living things grow and locomote, but investigation into the nature and explanation of vegetative and animal movement is not these days near the so-called core of philosophy. And anyway these topics are not mine, since what Aristotle called the "vegetative soul" and "sensitive soul" aren't really on Plato's mind when he argues for the partitioning of the human soul in *Republic* Book IV (435c-441c).[2] Neither growth, say, nor mere locomotion is distinctively human movement and the associated capacities are not distinctively human capacities. It is principally an understanding of those capacities and activities which are distinctively human that will help to address what he takes to be the guiding question of practical philosophy—the question of how to live (*Republic* 352d).

Plato's attention is fixed on *intentional self-movement* or *action*, which does seem to be our special talent.[3] He introduces the topic of the structure of the human soul, in large part, by asking, "When we set out after something, do we act with the whole of our soul, in each case?" (*Republic* 436a). And if I understand him, the question whether the human soul has parts is a question about whether we have more than one capacity whose exercise has action as upshot, i.e., whether we have more than one *practical capacity*.[4]

Let me try to bring out the sense of Plato's question by contrasting it with another, one with which it might be easily confused. Plato is *not* asking whether action is the culmination of a complex process, so that we are eventually to see the three parts of the soul (*epithumetikon*, *thumoeides* and *logistikon*) as marking three stages or sub-processes of the origination of action. Someone who reads him this way might take the partitioning in the following way: "Well, first you need appetite to make a proposal, then you need reason to decide whether to act, and

[2] Plato, *Republic*, trans. G. M. A. Grube and C. D. C. Reeve (Indianapolis: Hackett Publishing, 1992).
[3] I use "action" and "intentional action" as synonyms throughout.
[4] As I suggest below, the advantage of putting the point in terms of practical capacities is that it helps us to avoid thinking of a part of the soul as an inner item, and instead to think of it more sensibly as an ability to act for a reason of a certain kind.

then you need spirit to execute the decision, then the thing acts." Inquiring about the parts of the soul as an investigation of the mere ingredients of action would be like asking about the functional organization of an automobile engine: one would try to determine what in the thing makes it go at all. A mechanic might tell us, "Well, first you need the whatsit to shake and the thingamajig to rattle and then the thing goes."[5] Sticking to mechanical analogies, however imperfect, Plato's question is better understood as a question about whether humans have

[5] In "Self-Constitution in the Ethics of Plato and Kant," Christine Korsgaard develops an interpretation of Plato's practical psychology along these lines. She says,

> We can see the three parts of the soul as corresponding to three parts of a deliberative action. Deliberative action begins from the fact that we have certain appetites and desires. We are conscious of these, and they invite us to do certain actions or seek certain ends. Since we are rational, however, we do not act on our appetites and desires automatically, but instead decide whether to satisfy them or not…And then finally there is carrying the decision out—actually doing what we have decided to do. For of course we don't always do what we have decided to do, but are sometimes distracted by pleasure or pain or fear from the course we have set ourselves. So we can identify three parts of a deliberative action corresponding to Plato's three parts of the soul. (*Journal of Ethics* 3 [1999]: 1-29, pp. 6-7)

I won't develop a criticism of Korsgaard's interpretation of the significance of Plato's specification of parts of the soul, since I don't really intend my own discussion to be a responsible bit of textual analysis but rather something like an introduction to the point of view of this dissertation. Nevertheless, I want to mention that the passage Korsgaard offers in support of her interpretation is not to her own point. Korsgaard presents the passage as follows:

> the soul of someone who has an appetite for a thing wants what he has an appetite for and takes to himself what it is his will to have, and…insofar as he wishes something to be given to him, his soul, since it desires this to come about, nods assent to it as if in answer to a question. (*Republic* 437c)

The lesson Korsgaard draws is that "the soul does not act directly from appetite, but from something that endorses the appetite and says yes to it…it is as if what appetite does is put a request to reason, and reason says yes or no" ("Self-Constitution," p. 6). But when we look at the passage in its entirety, we see something quite different.

> Wouldn't you consider all the following, whether they are doings or undergoings, as pairs of opposites: Assent and dissent, wanting to have something and rejecting it, taking something and pushing it away? Yes, they are opposites. What about these? Wouldn't you include thirst, hunger, the appetites as a whole, and wishing and willing somewhere in the class we mentioned? Wouldn't you say that the soul of someone who has an appetite for a thing wants *what he has an appetite for* and takes to himself *what it is his will* to have, and that insofar as he wishes *something to be given to him*, his soul, since it desires *this* to come about, nods assent to it as if in answer to a question? (*Republic* 437b-c, italics mine).

Once we have the whole passage in view it is appears that the point Plato is making is that appetite, wishing and willing are each able to move someone to act, each is a member of "the class we mentioned." Korsgaard's ellipsis distorts this by making it seem as if "this" refers to what he has an appetite for when, it fact, it refers to what he wishes for. If so, the passage is not evidence for the claim that the role of appetite is to "make a proposal" and the role of reason to endorse or reject such proposals. Moreover, treating the parts of the soul as she does requires that Korsgaard treat agency in the abstract, and not humanity, as the whole which unifies the parts. This is a thought out of the spirit of the dialogue (*Republic* 588d), though in the spirit of Korsgaard's own project of teasing subjection to the requirements of morality out of the very idea of agency and action.

more than one *kind* of engine, like a present-day sailboat, or perhaps more to the point, a Mercedes mounted time machine.

That Plato is in fact interested in different practical capacities and not simply in the operation or functional organization of a single one is suggested by the abstract structure of his argument for the division. Famously, he begins by articulating the principle of opposites: "the same thing will not be willing to do or undergo opposites in the same part of itself, in relation to the same thing, at the same time" (*Republic* 436b). This is then used to locate parts of the soul: "if we ever find this happening in the soul, we'll know that we aren't dealing with one thing but many" (*Republic* 436b). Whatever one has to say about Plato's arguments, and much has been said, this seems clear: when he considers cases of internal conflict—cases of a human being undergoing opposites—he tends to consider cases in which someone is both pushed towards and pulled from *doing* something, from performing an action. That is, the oppositions which concern him—"assent and dissent, wanting to have something and rejecting it, taking something and pushing it away" (*Republic* 437b)—are practically or motivationally charged: the opposed states have action types as their contents. For example, when distinguishing the appetitive from the rational part, Plato says that "the soul of the thirsty person, insofar as he's thirsty, doesn't wish anything else but to drink, and it wants this and is impelled towards it" (*Republic* 439a), and when reason is opposed to this it is characterized as something which "draws back" the thirsty person from drinking.

If I'm right about how to understand Plato's inquiry, then the argument for the partitioning is an argument that we humans have multiple kinds of capacity to act or *kinds of practical capacity*. As the vegetative soul is the partner of growth and reproduction and the sensitive soul of mere locomotion, the appetitive, spirited and rational parts of the human soul

are *each* partners of intentional action. Although here we have to do with only one kind of happening, if you like, Plato argues that we have three capacities which have it as upshot.

Were being an agent simply being an individual bearer of such a capacity or power, then we might regard these three capacities as three kinds of agency. I will suggest an alternative. But first consider how in Book IX Plato asks us to put our imagination in the service of such a thought: telling us to see the appetitive part as "a single kind of multicolored beast with a ring of many heads that it can grow and change at will" (*Republic* 588c), the spirited part as a lion or a snake (*Republic* 588d, 590b), and the rational part as a human being (*Republic* 588d). It is a good rule of thumb to exercise caution in the neighborhood of Plato's metaphors: the particular capacities which we have been considering are in the first place parts, and so not obviously independently intelligible, self-standing, or again, to be found in complete isolation from the others, as these images encourage us to regard them. Consider another metaphor. Plato has us imagine the *union* of these capacities in the shape of one, "that of the man, so that to anyone who is unable to look within but who can see only the external sheath it appears to be one living creature, the man" (*Republic*, 588d-e). This last movement of the imagination introduces another way of thinking about a *form of agency*: here we have to do with a combination or unity of particular capacities, with particular capacities sewn together.

So far I've been especially vague about the relation between a part of the soul and action, saying things like action is the expression or upshot of a part of the soul and also that action is partnered with a part. But what does that mean?

In order to bring a bit more clarity to the nature of a practical capacity, let's consider the following question: When I'm intentionally drinking a gin and tonic which practical capacity or part of the soul do we have to do with? Well, that depends. In particular, it is depends on

why I'm drinking or what explains my drinking. I might be drinking because I'm thirsty, in which case my drinking has to do with the appetitive part. But there are other possibilities. When I was twelve my parents told me that drinking gin is for adults and that I couldn't have any, so in their sight I drank some because they said I couldn't. Here, I'm drinking out of anger or pride, in which case my drinking has to do with the spirited part. Or perhaps I've promised Jones that I would drink a gin and tonic, here I'm drinking because I promised, in which case my drinking has to do with the rational part of the soul.

According to Plato, the why of someone's doing something, or again what explains someone's doing something, is a desire, though for him this is a very general category, as general as the category of practical explainer itself. And, perhaps not surprisingly, Plato recognizes *kinds of practical explainer* or *forms of desire*, one corresponding to each part of the soul (*Republic* 580). The thought is that human action can have formally different sources, origins or springs, something made vivid by Plato's dramatic characterizations of the way in which appetite, reason and spirit move one to act: the appetites drive one around like a beast (*Republic* 439b), cool reason orders one to act (*Republic* 442c), while spirit the helper of reason boils, endures and keeps on till it is victorious (*Republic* 440c-d). Now, we might take these as metaphorical depictions of different kinds of hook-up or link between someone's intentionally doing something and why she is doing that. Ideally, we want something on the order of an account of the logical or formal characteristics of these connections and their component parts; still the metaphors can help to persuade us that some such thing is in the offing, or at least help to put into relief what is being sought.[6]

[6] Many have employed decorative language when characterizing the influencing motives of the will, though it would be rash to treat the metaphors as in every case serving such a metaphysically ambitious end. Hume's use of such language clearly does not: "What we commonly understand by *passion* is a violent and sensible emotion of mind,

Talk of the soul and parts of the soul raises worries. Is it a "ghost in the machine," an "immaterial Organ or Ministry," a para-mechanical hypothesis, a spirit substance that exerts itself through a very special sort of medium?[7] We are driven in such a "spooky" and to my mind unacceptable direction when we think that the practical capacity *itself* figures in particular explanations of action. But we need not have anything to do with the thought that a part of the soul *itself* moves its bearer to act. I've suggested that for Plato the source of motivation and what explains someone's doing something is a desire (again broadly construed). And thus we can remain attached to the structure and contents of ordinary action explanation and say that what is offered in response to the reason seeking "Why are you A-ing?" is a formulation of what moves the agent to act. Properly speaking, the expression or upshot of a part of the soul is someone's-doing-something-for-a-reason, as we say nowadays. A practical capacity, whether there be many, only one, or none at all, is a capacity *to-do-A-because-p*.[8] To be the bearer of a certain practical capacity just is to be something whose action might have its origin in a certain kind of item, or again, to be something which is a possible subject of a certain sort of explanation.

### *1.2 Analytical philosophy of practical reason*

Once we have access to our circle of concepts, including a kind of practical capacity and a kind of agency—recall this is a set or unity of practical capacities—we can raise a variety of

---

when any good or evil is presented, or any object, which, by the original formation of our faculties, is fitted to excite an appetite. By *reason* we mean affections of the very same kind with the former; but such as operate more calmly, and cause no disorder in the temper: Which tranquility leads us into a mistake concerning them, and causes us to regard them as conclusions only of our intellectual faculties" (*A Treatise of Human Nature*, ed. L. A. Selby-Bigge and rev. P. H. Nidditch, second edition [Oxford: Oxford University Press, 1978], p. 437).

[7] Gilbert Ryle, *The Concept of Mind* (New York: Barnes and Noble, 1949), p. 62.

[8] I am here ignoring considerable complexity in the constitution of forms of action explanation; in particular, I am ignoring the fact that some, if not all, forms appeal to two elements, what I will call, following Michael Thompson, the ground and the consideration. This is discussed in greater detail in Chapter two, Section four, "Means and Ends."

questions. Plato's interest is primarily in the parts and structure of the *human* soul. But mightn't it be better to bring things a little closer to home? Couldn't I simply wonder about D. L.'s practical nature: Which practical capacities do I have? Which form of agency is mine? Last year I thought perhaps only appetite and reason move me. More recently there has been some evidence for the presence of a spirited part. It's curious. Of course, such an inquiry would be very boring, or anyway boring to all but me and my sentencing judge. It seems likely that if *we* are ever to answer Plato's question of how to live, *we* must expand our horizons and have things to say about *human* nature, about the parts and structure of the *human* soul. Still, even if we set answering this question as our ultimate end, we need not begin with it, and do not address it directly here. In any case, what about the intelligent aliens told of in philosophy and science fiction? Should they care about the parts and structure of the human soul? Or better should their philosophers, as opposed to their military strategists and ad-men? The inquiry we do undertake is sufficiently abstract to be of appeal even to alien philosophers—its starting point is *the very idea of agency*. Agency in this abstract sense is a capacity for operation that is governed by thought, or again, a capacity for operation involving the application of concepts.[9] It is the capacity to act. But what about this exactly?

Our question is this: Which particular practical capacities must something have if it is an agent at all, if it acts at all? For example, might it be that anything which acts must have exactly the three practical capacities that Plato discusses? Or is the very idea of an agent without an appetitive part or without a spirited part simply incoherent? Or this: is the idea of an agent with only an appetitive part or only a spirited part a bit of patent nonsense? We're

[9] For the moment, I am using the expression "operation" as a placeholder; more will be said later, especially in chapter 1, section 2.

asking that sort of thing. My topic, then, is *what can be determined about the possible forms of rational agency on the basis of reflection on the nature of action*.

At the extremities of this kind of investigation, what I'll call *analytical philosophy of practical reason*, are broadly Kantian and Humean positions. It is characteristic of the *analytical Kantian* to hold that if something is an agent then it *must* be able to do more than put means to ends, though, there is great diversity of opinion over what that something more might be. At the other extreme, the *analytical Humean* holds that the capacity to put means to ends is the *only possible* practical capacity.[10] So, where the analytical Kantian attempts to set a robust lower bound of practical sense, the analytical Humean hopes to set a spare upper bound. My own work aims to resist each of these analytical extremes, while nevertheless making a positive contribution to the debate. The next section describes the course of this resistance as well as the character of the positive contribution.

### *1.3 Outline of the argument*

*On the necessity of instrumental reason*. We act: Jones buttered the toast, Smith replenished the house water supply. We put means to ends: Davidson wrote the letter "a" in order to write

[10] Analytical Kantian projects have been pursued recently by, for example, Stephen Darwall, David Gauthier, Christine Korsgaard, Thomas Nagel and David Velleman, though in importantly different ways: Stephen Darwall, *Impartial Reason* (Ithaca: Cornell University Press, 1983), "Kantian Practical Reason Defended," *Ethics* 96 (1985): 89-99; David Gauthier, "The Unity of Reason: A Subversive Reinterpretation of Kant," *Ethics* 96 (1985): 74-88; Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), "The Normativity of Instrumental Reason" [in *Ethics and Practical Reason*, eds. G. Cullity and B. Gaut (New York: Oxford University Press), pp. 215-254], "Self-Constitution in the Ethics of Plato and Kant," and her unpublished Locke Lectures (Oxford University, 2002); Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970); David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000).

Analytical Humean projects have been pursued recently by, for example, Simon Blackburn, James Dreier, Michael Smith, again in importantly different ways: Simon Blackburn "Practical Tortoise Raising," *Mind* 104 (1995): 695-711, *Ruling Passions* (New York: Oxford University Press, 1998), chapter 4; James Dreier, "Humean Doubts about the Practical Justification of Morality," in *Ethics and Practical Reason*, pp. 81-99; Michael Smith, "The Humean Theory of Motivation," *Mind* 96 (1987): 36-61.

"action", Anscombe wrote the letter "f" in order to write "fool". Is it true that necessarily, if something can act, then it can put means to ends?

This question is about the *conditions of agency*, where the relevant condition is of a certain type. The condition is not simply that the agent be placed in a certain environment, for example, a universe not completely governed by deterministic laws, but rather that she possess a *rational capacity*. However, the condition is also not simply that she possess any rational capacity at all, for example, the capacity to aim to represent how the world is. Here the relevant condition is that the agent possess a rational capacity that is *practical*, a capacity to intervene in the world through the application of concepts. Indeed, in chapter one, I raise a question about whether possession of a certain practical capacity—the capacity to do one thing in order to do another—is a condition of agency. To show that it is, would be to set a lower bound of practical sense.

Inquiries of this shape are familiar enough from recent work on practical reason, though the relevant lower bound tends to be set far higher. Consider a remark by Christine Korsgaard:

> The special relation between agent and action, the necessitation that makes that relation different from an event's merely taking place in the agent's body, cannot be established in the absence of at least a claim to law or universality. So I need to will universally in order to see my action as something which *I do*.[11]

Or consider this one by Stephen Darwall:

> Unified rational agency…requires the capacity to adopt a reflective standpoint from which we can consider what reasons there are to prefer one act over another, and from which *we* form an all-things-considered preference. This includes, in our initial account, the capacity to reflect dispassionately and to be motivated by our awareness of different considerations.[12]

However, the project of setting robust lower limits of practical sense is not restricted to those, like Darwall and Korsgaard, who are developing a program that is Kantian in both spirit and

[11] Christine Korsgaard, *The Sources of Normativity*, pp. 228-229.
[12] Stephen Darwall, *Impartial Reason*, p. 111.

substance. For example, in an essay on the foundations of the theory of rational choice, David Gauthier argues that

> action involves a choice among alternatives. For choice to be possible, the desires of the actor must be unified in such a way that they determine a single alternative from those possible actions available to her. The actor's desires must be so related that they determine a preferential ordering of the set of alternative possible actions, from which she may then select a maximal element. The familiar ideas of the theory of rational choice correspond to the pure concepts of the will.[13]

Putting their many differences aside, each of these authors is concerned with what can be determined about the possible forms of rational agency by reflection on the nature of action considered as such—each is engaged in what I am calling *analytical philosophy of practical reason*. More specifically, each is trying to establish that possession of some determinate practical capacity, whether to will universally, to reflect dispassionately, or to maximize, is a necessary feature of anything that can act. And in chapter one, I try to establish that possession of the capacity *to do this in order to do that* is just such a feature. This is *analytical instrumentalism*.[14]

It is widely acknowledged that, as Bernard Williams puts it,

> We cannot simply assume that moral considerations, for instance, or long-term prudential concerns must figure in every agent's S [subjective motivational set]. For many agents, as we well know, they indeed do so, if not altogether securely; but a philosophical claim that they are necessarily part of rational agency needs argument.[15]

Indeed, those like Darwall, Gauthier and Korsgaard who think that justice or prudence are necessarily part of rational agency recognize the burden and address themselves to supplying the relevant argument. But why restrict our curiosity to the status of these admittedly magnificent capacities? As we well know, not just many, but every agent we have ever

---

[13] David Gauthier, "The Unity of Reason: A Subversive Reinterpretation of Kant," pp. 115-116.

[14] It is important to flag the difference between *analytical instrumentalism* and the far more radical *analytical Humeanism*, according to which the capacity to put means to ends is the *only possible* practical capacity. It is also important to flag that there is a real question whether Kant is himself an analytical Kantian. At least three passages suggest that he means to leave open the possibility of a merely instrumental agent: 4:395, 5:58-9, 6:26. For discussion of this question see Stephen Engstrom, "Contradictions in the Will" (unpublished) and Sergio Tenenbaum, "Speculative Mistakes and Ordinary Temptations: Kant on Instrumentalist Conceptions of Practical Reason," *History of Philosophy Quarterly* 20 (2003): 203-223.

[15] Bernard Williams, "Some Further Notes on Internal and External Reasons," in *Practical Reasoning*, ed. Elijah Millgram (Cambridge: MIT Press, 2001), pp. 91-97, pp. 91-92.

encountered has been able to take means to ends. And we do not, I suspect, invite controversy by going on to hold that any agent must. But do we need genuine controversy in philosophy if a demand for explanation is to get hold? Must objects of philosophical investigation be potential battlegrounds? I would have thought that any philosophical claim that something is "necessarily part of rational agency" needs argument. Maybe not to convince another of something that might really be doubted—I can't really doubt that I have knowledge of a world outside of me—but to help us understand what is so plainly the case.

In the next chapter, I attempt to understand why the ability to take means to ends is necessarily possessed by any rational agent. The upshot is that what I call an *atomic agent*—an agent who cannot do one thing in order to do another—is simply impossible. In establishing this, I establish that a certain minimal conceptual structure is necessary for agency. I do this by attending to features of the *form* of the judgments we make when reporting action, especially to the temporal features of tense and aspect.

*Analytical Kantianism and the instrumental ought*. The constructive part of the discussion of the conditions of agency is of value not only because it reveals essential features of agency, but also because it can serve as a scale-model of the type of analytical argument of interest, and a background against which to assess the more far reaching Kantian and Humean versions.

I am not aware of a good argument for *analytical Humeanism*, according to which nothing other than instrumental agency is so much as possible, and I do not take the doctrine to be true. In the absence of such an argument, any defense of the possibility of the purely instrumental agent must proceed slowly, considering what is said against it bit by bit. The chapters three and four take up only one little bit and address precisely the charge that a purely instrumental agent cannot be subject to *standards of correctness* in action. These

chapters address the charge that a purely instrumental agent cannot be a subject of practical *norms*, *requirements* or *oughts*, in particular those given by the so-called *instrumental principle*. As an organizational matter, the discussion orbits Korsgaard's "The Normativity of Instrumental Reason", one of the most exciting recent efforts in the analytical Kantian tradition. There Korsgaard argues along the following lines:

(i) If an agent is under a principle, then it must be able to violate it.
(ii) If something is an agent, then it must be under the instrumental principle.
(iii) So, if something is an agent, then it must be able to violate the instrumental principle.
(iv) But a purely instrumental agent cannot violate the instrumental principle.
(v) So, a purely instrumental agent is not under the instrumental principle.
(vi) So, a purely instrumental agent is not really an agent.[16]

What follows, I am anxious to say, aspires to more than Korsgaard interpretation. Even apart from contributing to a defense of the possibility of a purely instrumental agent, I also mean for the discussion to contribute to our understanding of the metaphysics of practical normativity, and more specifically of the *oughts* internal to *in order to*.

*Practical reason and the possibility of error*. In chapter three, I investigate the premise linking practical reason and the possibility of error. It is what I call the *error constraint*. The error constraint is ambiguous. According to the *logical interpretation*, an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle. The logical interpretation is concerned with what is possible independent of facts about the agent. In contrast, the *imperatival interpretation* is concerned with what is possible given facts about the agent, and says that an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby

[16] The most sustained development of the argument is in "The Normativity of Instrumental Reason" but she advances it many other places as well: *The Sources of Normativity*, "Self-Constitution in the Ethics of Plato and Kant," "Reply to Ginsborg, Schneewind, and Guyer," *Ethics*, 109, October 1998, 49-64 and most recently in her Locke Lectures (Oxford University, 2002).

violate the principle and it is possible for the agent, given her particular capacities, to do it. I argue that the apparent plausibility of the latter derives in part from conflating it with the obviously true logical interpretation, and that we are in any case not forced by argument to accept the error constraint on its imperatival interpretation. Moreover, I show that because the imperatival interpretation entails (a) the liberty of indifference conception of freedom and (b) the impossibility of a perfectly rational agent, it is flatly incompatible with *constitutivism*, the Kantian strategy for elucidating the validity of a principle of practical reason by revealing that it is constitutive of a form of mental activity, and the chief alternative to *platonism* among non-reductive and realist accounts of practical norms. So, the decision about how far to write the possibility of error into practical reason ultimately restricts how we can understand the nature and authority of principles in the first place.

*There is no such thing as the instrumental principle*. In chapter two, while elucidating the non-contingency of instrumental agency, my focus is on the explanatory dimension of practical reason: not on what an agent ought to do, but on what can explain what an agent does do. In chapter three, I adjust my focus towards the normative dimension of practical reason to consider how exactly to understand the thought that error must be possible where practical reason is. But I refrain from saying anything about the content of particular standards of action. In chapter four, I investigate the normative dimension of specifically instrumental agency. As I understand it, the task is to see whether there is an appropriately minimal conception of the norms in play for an agent with such limited practical powers, and also to see whether these norms are independently intelligible or self-standing.

The recent discussion of instrumental rationality has centered on questions about the content and the authority of *the instrumental principle*: *one ought to take the necessary means to*

*one's ends*. This is a mistake. An attempt to provide a minimal theory of practical norms in terms of the instrumental principle confronts a dilemma. On the first horn, the interpretation of the content of the instrumental principle is so spare that it does not articulate a standard that can be genuinely action guiding; that is, it does not articulate a standard which can figure as a content of the practical thinking of an agent. On the second horn, the interpretation is so robust that it is not a plausible specification of a minimal theory of practical norms.

On the robust interpretation, the instrumental principle enjoins resoluteness in the pursuit of ends. This is the interpretation in play in Korsgaard's "The Normativity of Instrumental Reason." And so, the argument of that essay is that to possess resolve in the pursuit of ends, something must be able to engage in non-instrumental reasoning about what to do. This is an interesting claim. By itself, however, it does not pose a challenge to the coherence of instrumentalism. It is too easy to deny the proposition that any agent must be a possible subject of resolution, endurance or strength of mind. Where Korsgaard needs to be considering the conditions of being subject to the norms internal to *in order to*, she is considering something else, namely the conditions of constancy and inconstancy in the pursuit of an end.

Even so, the failure to demonstrate that instrumentalism is incoherent can nevertheless be transformed into an argument that *human* agents are not, in fact, purely instrumental agents. A purely instrumental agent cannot exhibit constancy or inconstancy, while human agents can and do. And so, Korsgaard's argument can be transformed into a challenge to analytical Humeanism, which is committed to the view that only purely instrumental agency is possible. But since we do exhibit these, we are not purely instrumental agents, and so something other than that is possible, indeed actual.

What, then, are the norms in play for a purely instrumental agent? I draw on the discussion of the explanatory dimension of instrumental reason in chapter two and propose that we look to *action-forms* to provide the material for a properly minimal and self-standing account of norms internal to the capacity to do one thing in order to do another. Action-forms satisfy several abstract conditions on anything purporting to be a standard of action: they are potentially explanatory of action, they ground explicitly deontic claims, and there is such a thing as erring in respect of them. Moreover, I suggest that this primitive form of practical normativity can be articulated by reflecting on the form of the judgments we make about action, in particular, by reflecting, once again, on the aspectual opposition found in *X is □-ing* and *X □-ed*.

*Conclusion*. I have tried to write each chapter so that it purses an independent thesis. Nevertheless, what follows is organized as a general defense of a very spare position in analytical philosophy of practical reason, *analytical minimalism*, according to which (i) any agent must possess the capacity to put means to ends and (ii) this exhausts the content of what can be settled analytically.

It is worth asking, even without answering, where analytical minimalism leaves the theory of practical reason quite generally? Here I feel compelled to say, following Philippa Foot's advice, that I am a minimalist only *in a way*. I do not mean for this dissertation to support a minimalist picture of the nature of *human* rational agency. As I just suggested, I think that there are features of the structure of our agency that cannot be captured by an instrumentalist theory of practical reason. Indeed, much of the interest of analytical minimalism, I think, stems from the way it forces us to understand these other features. If they are not part of the structure of any practical outlook whatsoever, how can they possibly be a source of rational

requirements on action, requirements which one can't simply opt out of? I do not address that question here, but mention it to give a sense of the problems that analytical minimalism leaves us with. In the end, analytical philosophy of practical reason is some of the necessary spade work for developing a more complete theory of the structure of specifically human practical thinking.

## 2. ON THE NECESSITY OF INSTRUMENTAL REASON

> Ancient and medieval philosophers—or some of them at any rate—regarded it as evident, demonstrable, that human beings must always act with some end in view, and even with some one end in view. The argument for this strikes us as rather strange. Can't a man just do what he does, a great deal of the time? He may or may not have a reason or a purpose; and if he has a reason or purpose, it in turn may just be what he happens to want; why demand a reason or purpose for *it*? And why must we at last arrive at some *one* purpose that has intrinsic finality about it? The old arguments were designed to show that the chain could not go on for ever; they pass us by, because we are not inclined to think it *must* even begin.
>
> G. E. M. Anscombe, *Intention* §21

### *2.1 The question of the necessity of instrumental reason*

Most philosophers are acquainted with G. E. M. Anscombe's poisoner: "I'm pumping in order to replenish the house water supply," "I'm replenishing the house water supply in order to poison the inhabitants"—or if not with this guy, then with Donald Davidson's scribe: "I'm writing the letter 'a' in order to write the word 'action'," "I'm writing 'action' in order to write the sentence 'action is life'." But one needn't have read Anscombe or Davidson to be acquainted with the subject matter of their philosophical reflections: the sort of explanation of action exhibited here, broadly speaking, explanation of the form "X is doing A in order to do B". That is, we are all acquainted with the explanation of action in terms of an agent's further purpose, aim, goal or end. I will call this form of explanation *instrumental explanation of action*.

We are, I suppose, equally familiar with the form of advice internally related to this: "You ought to pump seeing as you propose to replenish the house water supply," "You really must

write the letter 'a' given that you aim to write the word 'action'". That is, we are all familiar with practical advice of the form "X ought to do A, given that X aims to do B and doing A is part of doing B." I will call the kind of practical requirement in view here *instrumental requirement*.

All parties to the dispute about the nature and scope of practical reason agree that what a rational agent does is sometimes to be explained instrumentally, and also agree that agents are subject to instrumental requirements, norms or standards. Instrumental reason figures as a fixed point in the theory of practical reason. Even so, there are foundational questions to be asked.

Recently a lot of work has been done to address them.[17] It is a striking feature of this recent work that it is singularly focused on questions having to do with instrumental requirement. Actually, the focus is narrower still: much of the discussion is cast in terms of questions about the content and the authority of *the instrumental principle*—the principle enjoining us to take the means that are necessary relative to our ends. According to R. Jay Wallace, a proper investigation of instrumental reason proceeds in two steps. "The first step concerns the kind of requirement represented by the instrumental principle." What is its content? "The second step is to elucidate the normative force of the requirement that is embodied in the instrumental principle…What is it that makes it a

[17] John Broome, "Normative Requirements" *Ratio* (new series) 12 (1999), pp. 398-419, and "Practical Reasoning," in Jose Bermudez, ed. *Reason and Nature: Essays in the Theory of Rationality* (Oxford: Oxford University Press, forthcoming); Stephen Darwall, "Two Dogmas of Empiricism in Ethics" (unpublished); James Dreier, "Humean Doubts about the Justification of Morality," in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (New York: Oxford University Press, 1997), pp. 80-99; Donald Hubin, "The Groundless Normativity of Instrumental Rationality," *Journal of Philosophy* 98 (2001); Christine Korsgaard, "The Normativity of Instrumental Reason" in *Ethics and Practical Reason*, pp. 215-254; Peter Railton, "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," in *Ethics and Practical Reason*, pp. 53-79; T. M. Scanlon, "Structural Irrationality" (unpublished); R. Jay Wallace, "Normativity, Commitment and Instrumental Reason" *Philosophers' Imprint*, vol. 1, no. 3 (2001): http://www.philosophersimprint.org/001003.

rational constraint?"[18] Among those attempting this second step, there is a tendency to argue that being an agent just is, in part, being subject to instrumental requirement, to argue, that is, that subjection to instrumental requirement is partly constitutive of rational agency.

Those developing this sort of *constitutivist* position tend to assume unquestioningly, however, that being an agent also involves being the subject of instrumental *explanation*. If any agent must be subject to instrumental requirement, any agent must be able to act in accord with what is required, and so, in this case, must be able to take means to ends. Indeed, at the bedrock of Peter Railton's own account of the authority of instrumental requirement is the assumption that agents "necessarily possess and act on *ends*."[19] It appears, then, that attempts to validate instrumental standards of action uncritically rest on an assumption about the status of a certain form of explanation of action. And just as we can raise foundational questions about the nature of instrumental requirement, even though everyone will grant that there is such a thing, so too can we raise questions about the status of instrumental explanation. In this chapter, I address a question about its necessity.

We act intentionally: Jones buttered the toast, Smith replenished the house water supply. And we put means to ends: Davidson wrote the letter "a" in order to write "action", Anscombe wrote the letter "f" in order to write "fool". Now, is it true that if

---

[18] R. Jay Wallace, "Normativity, Commitment, and Instrumental Reason," pp. 16-17.

[19] Railton, "On the Hypothetical and Non-Hypothetical" pp. 64-5, cf. 68. The project of elucidating the normative force of the instrumental principle relies at least implicitly on the truth of analytical instrumentalism. While there is disagreement about both the content and ground of the requirement to take the necessary means to one's ends, all sides aim to validate a practical requirement that has *universal* and *necessary* application to agents. Given even the most minimal dependence of the applicability of a norm of action to an agent on the possibility of the agent's acting in accord with the norm, validation of the instrumental principle seems to require the truth of analytical instrumentalism; otherwise, there might be an agent who simply cannot put means to ends, and yet who is required to take the necessary means to its ends.

something can act, then it must be able to do one thing in order to do another? Must some of what an agent does be explained by the fact that it has a further aim? Or again, is the ability to pursue means to ends necessarily possessed by any rational agent? To hold that it is is to accept what I will call the *necessity of instrumental explanation of action* or sometimes *analytical instrumentalism*.[20]

But why should we accept that? Why should we accept that agents must possess ends whose realization involves the taking of means? Couldn't there be an agent which just does what it does? First one thing and then another, but never one with a view to another? After all, there seems to be such a thing as performing an action but not with a view to the realization of some further end—as when I touch a spot on the wall *for no particular reason*, kill Jones *out of anger*, return five dollars to Smith *because I promised*, massage my foot *because I like it*, or help another in need *from duty* or *for its own sake*. There also seems to be such a thing as performing an action but not by performing any other actions. That is, there seems to be such a thing as *instrumentally basic action*—perhaps such actions as starting to walk, blinking, moving my finger or raising my arm are of this type. Now, why couldn't there be an agent whose action must be like that, what we might call an *atomic agent*? An atomic agent acts intentionally and acts for reasons, but can't perform one action in order to perform some other. An atomic agent can't act for the

[20] In section 20 of *Intention*, Anscombe asks "Would intentional actions still have the characteristic 'intentional', if there were no such thing as…further intention in acting?" (Oxford: Blackwell, 1957). My sense of the importance of the question of the necessity of instrumental explanation of action, as well as the spirit of my answer, owes much to Anscombe. Even so, I do not want to rely on her specific treatment of the necessity of instrumental reason except to note again that she is focused on the explanatory dimension of practical reason.

acquisition of some further end. Perhaps there can't be an atomic agent. But why not? What is the ground of this absurdity, if it is one?[21]

What follows falls into four parts. In the first, I draw our attention to the temporality of judgments reporting that something has been done intentionally, in particular to the fact that these judgments have *perfective aspect*. Drawing on this observation, the second part argues for the conceptual priority of what I call *durative telic action*: action that takes time, can be interrupted and remain incomplete. The third part makes some general remarks about instrumental explanation. Finally, in the fourth part, I develop an argument for the claim that performing a durative telic action just is doing one thing in order to do another.

### *2.2 Action and Aspect*

Any argument purporting to show that rational action is possible only where certain types of consideration can figure as explanatory reasons must begin with a conception of intentional action, or, as I will say more simply, action. Many candidates are available and much goes into the initial selection from among these. Within this context, there is a great advantage to

[21] It is worth mentioning that those who want to leave conceptual space for a *divine agent* tend to deny what I am calling the necessity of instrumental explanation. Consider Aquinas: he seems to combine the thesis that the divine intellect is practical (*ST* 1.19.1.), with the denial that God wills "this on account of that" (*ST* 1.19.5) or again that God acts "for the acquisition of some end" (*ST* 1.44.4). St Thomas Aquinas, *Summa Theologiae*, trans. Fathers of the English Dominican Province (London: Burns Oates, 1920; repr. Westminster, Md.: Christian Classics, 1981). It is an interesting question, I think, whether contemporary and medieval theories of practical reason might come apart on the importance accorded to instrumental reason—not because of contemporary skepticism about there being anything else to the forms of practical thought, but because of medieval skepticism about the *sine qua non* status of instrumental calculation itself. Elsewhere I would want to argue that the tension between Aquinas' conception of divine agency and the thesis of analytical instrumentalism is merely apparent. In this I differ with Rowland Stout who is simply willing to deny the coherence of the idea of a divine agent to preserve the thought that instrumental reason is essential to agency: "If it were theoretically possible to have an omnipotent agent, then this would constitute a counter-example to the claim that practical justification must involve means-end justification…But I do not see any philosophical reason to suppose the idea of an omnipotent *agent* to be a coherent one" (*Things That Happen Because They Should*, [Oxford: Oxford University Press, 1996], p.127). I should also say that in taking the thesis of the necessity of instrumental explanation of action to be common ground among theories of practical reason, I differ with Candace Vogler who considers a similar question about the status of instrumental reason but against the background of a supposed dispute. See her *Reasonably Vicious* (Cambridge, Mass.: Harvard University Press, 2002).

operating with a maximally spare conception. The more baroque one's starting point, the more one risks losing one's audience on the ground that the starting point builds in too much: what aspires to the status of transcendental argument will be alleged to be mere stipulation.[22] To avoid this considerable hazard, I operate with an extremely limited conception of action which most will accept is, at the very least, a part of a complete account. The purpose of this section is to say just enough about action to provide a starting point for extracting the capacity to put means to ends. What follows does not, I want to emphasize, purport be something on the order of a theory of action.[23]

It seems like good policy always to begin where Donald Davidson begins and that's what I'll do. In particular, I want to begin with judgments that an action happened, took place or occurred. Here are some familiar examples: Jones buttered the toast; the queen moved her hand; I wrote the word "action"; Smith hit the bull's eye.[24] Quite generally, the class of judgments of interest are those that can be intelligibly given in response to the question "What did X do?". For example, it makes sense to reply "Jones buttered the toast" but does not make sense to reply "Jones was six feet tall"; it makes sense to reply "I wrote the word 'action'" but does not make sense to reply "The queen was angry." I will follow the grammarians in calling a judgment passing this test a *perfective judgment*. I plan to pay considerable attention to perfective judgment in order to see what it might

[22] It is worth recalling that we are in the middle of a larger project meant to raise doubts about *analytical Kantianism*. According to it, practical reason considered as such *must* include more than a merely instrumental reason. I am emphasizing the slightness of the conception of action from which I begin to set up a contrast with the robustness of the typical starting point of analytical Kantian arguments. These arguments are at great risk of presupposing what needs to be established by beginning with conceptions of action that are too robust.

[23] In the background of my discussion, however, is the theory of action developed in Michael Thompson's "Naïve Action Theory," chapter two of his forthcoming *Life and Action* (Cambridge, Mass.: Harvard University Press). I am extremely indebted to that essay.

[24] These are all in the past tense, of course, and to simplify matters I will restrict my attention to the past from here on.

tell us about the nature and structure of action. For this purpose it will be useful to contrast perfective judgments with judgments that something is in a certain state or has a certain property. Consider, for example: Jones was in the bathroom, I had a pen, Smith was the world champion of darts.

Action is reported by perfective judgment. Since being an agent is simply having the capacity to act, part of what it is to be an agent is to be something that figures as the subject of such a judgment. There is, I think, quite a bit of information contained in this and, as surprising as it may seem, excavating that will bring us closer to an understanding of the necessity of instrumental reason. The features of perfective judgments that I am especially interested in are not what Davidson considers in his seminal essay "The Logical Form of Action Sentences," for example, that they admit of adverbial modification and nominalization transcription, or, ultimately, that they involve implicit reference to particular events. I am, instead, interested in what Davidson abstracts from in his discussion—their temporal features, their *tense* and *aspect*.[25]

Tense and aspect are both features of the way a judgment represents the temporality of what it is about, but they must be distinguished. Consider the following judgments:

(Perfective) Jones walked to school.

(Imperfective) Jones was walking to school.

The subject, Jones, and predicate, walk to school, are the same. Each is in the past tense. Where these judgments differ is in their aspect: the first has perfective aspect, the second imperfective. Aspect characterizes the internal temporal shape of what is reported. A judgment with imperfective aspect, Jones was walking to school, represents something as in progress or

[25] Donald Davidson, "The Logical Form of Action Sentences," in *Essays on Actions and Events* (New York: Oxford University Press, 1980).

under way. A judgment with perfective aspect, Jones walked to school, represents something as completed and done.

Tense, on the other hand, relates the time of what is reported to some other time, typically the time the report is made. Consider the following judgments:

(Past) Jones was at school.

(Present) Jones is at school.

(Future) Jones will be at school.

Of course, the first is in past tense, the second the present and the third the future. It is an interesting fact about this trio that they share a common core of meaning which seems best given by the present tense judgment that Jones is at school. This shows up in the fact that the claim that Jones was at school is equivalent to the claim that it was the case that Jones is at school. Indeed, it seems that any past tense judgment reporting that something was in a certain state is equivalent to some claim of the form *it was the case that p* where *p* is a present tense judgment. Many have relied on equivalences of this shape in an argument for the conceptual priority of the present tense, treating judgments of the past and future as built out of it and other conceptual materials.

But is every past tense judgment equivalent to some claim of the form *it was the case that p*? Consider perfective judgments. Can we analyze the claim that Jones walked to school as a claim that it was the case that *p*, for some present tense proposition *p*? We cannot. To appreciate the difficulty, try to find a present tense judgment to substitute for *p* which can do the relevant work. The obvious candidate is "Jones walks to school." But this doesn't work. The problem is that "Jones walks to school" means that Jones is in the habit of walking to school. If that is right, then the past tense of "Jones walks to school" is really "Jones used to

walk to school." And that is not the judgment at issue. The judgment at issue reports a particular occurrence of Jones' walking to school. Are there other candidates? It will not help, I think, to appeal to the imperfective judgment "Jones is walking to school." The past tense of this is simply "Jones was walking to school," which is not equivalent to "Jones walked to school." After all, it might be that Jones was walking to school even if it is not true that Jones walked to school; suppose Jones gets run over by a bus half-way there. But if not either of these, then what? The claim that Jones walked to school is not equivalent to a claim of the form *it was the case that p*, for some *p*.

This suggests quite generally that perfective aspect is incompatible with present tense meaning. This helps, I hope, to put something into relief: actions, we might say, are events or happenings, and our thought has already taken a special turn when it comes to these.[26] [27]

### *2.3 The conceptual priority of durative telic action*

I realize that it is not obvious what these reflections on perfective aspect could have to do with our original question about the necessity of instrumental explanation of action. Let me try to say something about that now. To facilitate the discussion I need to draw two distinctions among actions.

---

[26] Here and elsewhere I am drawing on Antony Galton's discussion of aspect in *The Logic of Aspect* and in "The Logic of Occurrence" in *Temporal Logics and Their Appications* (London: HBJ Publishers, 1987). In a fuller treatment, more would have to be said to make compelling the claim that perfective aspect and the present tense are incompatible. In particular, something would have to be said about the possibility of a reportive reading of "Jones walks to School" playing the relevant role. It is in the present tense, though if one says something in that voice, say, "Seabiscut wins!", it must already be true that Seabiscut won. And so it does not seem that we have something which says of the present what the other says of the past. For work on events and aspect which tries to bring perfective aspect and the present tense meaning together in this way see Terence Parsons in *Events in the Semantics of English: A Study in Subatomic Semantics* (Cambridge: MIT Press, 1990).

[27] Before pressing on, I should note a curious feature of this discussion: everything that I say about action in sections 2.2 and 2.3 can be said about events quite generally, though I do not proceed in such general terms. Nevertheless, to avoid commitment to the unlikely thesis that being the subject of an event essentially involves possessing the capacity to reason instrumentally, I will ultimately have to mark a division with the class of events. Provisional steps are taken in this direction in the final sections of this chapter.

The first is a distinction between *durative* and *instantaneous* actions, between those which take time and those which do not. Every action is either instantaneous or durative and not both. The hallmark of an instantaneous action is that we cannot catch it in the act of happening; at any moment it has either already happened or is yet to come. I have in mind, say, "X started to walk, X stopped walking, X reached the summit, X won the race." In contrast, a durative action is something we can catch in the act of happening; at some moments it is in the process of happening. I have in mind, say, "X walked to school, X climbed Everest or X built a house." To mark this distinction I will rely on the following, somewhat imprecise, test: if "X A-ed" is durative, then at some point X was A-ing, whereas if "X A-ed" is instantaneous, then at no point was X A-ing.

Within the class of durative actions, we can make a distinction between *telic* and *atelic* action. Every durative action is either telic or atelic and not both. The hallmark of a telic action is that the distinction between coming to completion and merely stopping has application. I will assume that "Jones walked to school" reports a telic action. For it to be the case that Jones walked to school, it is not enough that Jones started to walk to school, was walking to school, and stopped walking to school. Jones has to get there. Jones has to bring the process to completion and finish. In contrast, atelic action does not give application to the distinction between completion and merely stopping. With the literature, I will assume that "Jones walked" reports an atelic action. For it to be true that Jones walked, it is enough that Jones started to walk, Jones was walking, and Jones stopped walking. The distinction between telic and atelic action shows up in connection with the imperfective judgments with which each is correlated—in this case, Jones was walking to school and Jones was walking. To mark the distinction I will rely on the

following, somewhat imprecise, test: when A-ing is atelic, "X was A-ing" does imply that X A-ed.[28] On the other hand, when A-ing is telic, "X was A-ing" does *not* imply that X A-ed; indeed, it seems that when A-ing is telic, "X was A-ing" implies that X has not yet A-ed.

To common sense the following claim will be obvious: instantaneous and atelic actions are exceptional cases of action; as a rule actions are durative telic. Statistically speaking this seems plainly true: the representation of durative telic actions in speech and thought prevails. But I want to argue that the priority is not merely statistical but conceptual. And I want to suggest that the conceptual priority of durative telic action is already contained in our reflections on action and perfective aspect.

But before trying to establish this point, let me say why it might be significant for settling the question of the necessity of instrumental reason. An *instrumentally* or *teleologically basic action* is an action not done by doing anything else at all. If "X A-ed" is instrumentally basic, X did not do anything in order to do A, but simply did A directly or just like that. If any actions are instrumentally basic, instantaneous or atelic actions are; but it is precisely these that I am suggesting are derivative cases of action. Durative telic action, however, cannot be instrumentally basic. As I will suggest in the final section, they are necessarily done by doing other things.

If in every context perfective judgments could be systematically replaced by logically equivalent judgments which do not have perfective aspect, we would not need to recognize perfective judgment as a genuine form of judgment. And if it is possible to

[28] These entailments, it should be noted, rest on the contents of the verbs in question—formally, X was A-ing fails to entail X A-ed.

replace only a sub-set of perfective judgments, then those which do not admit of replacement are *fundamental* in the relevant sense.

Now, my thought is that the apparatus of moments, temporal intervals, and states is sufficient to express anything that can be expressed by perfective judgments reporting instantaneous or atelic action. Restricted to these types of action perfective judgment would provide no further expressive power; talking about things having been *done* would then be no more than a manner of speaking. I want to suggest, however, that it is *not* possible to eliminate by paraphrase perfective judgments reporting durative and telic actions. If so, the only thing that requires us to talk in terms of what happened, took place or occurred—of what got done—is durative telic action. It is for this reason fundamental. But let me try to say something to make this plausible.

*Instantaneous action.* Instantaneous action does not have temporal structure or any temporal parts. It thus seems to be the sort of thing an atomic agent might do: nothing else seems to be required to perform one. Take, for example, the action reported by "X stopped walking." There does not seem to be any sense to the question "By doing what did X stop walking?" and so seems to be instrumentally basic. But how exactly are to understand what it is to stop walking without employing perfective judgment and treating this as something one can *do*? What is the relevant paraphrase? For the sake of simplicity, suppose that time is discrete, and so, that there is such a thing as the next moment in time. It is then possible to paraphrase "X stopped walking" as "It was the case that X is walking, and next, X is not walking." That is, for it to be true that X stopped walking is just for it to be true that at some point in time X is walking and then at the next point it is false that X is walking. If we give up on the assumption that time is discrete, it will still

be possible to generate a paraphrase, one much more complicated. Here's a sketch. "X started walking" is paraphrased by the following: either X is walking now but was not walking earlier, or X was walking earlier but was not walking at some time before that.[29] Parallel remarks might be made about "X stopped walking." Even so, to give an analysis of a particular case, is not, of course, to give a general formula or algorithm for producing paraphrases of the relevant type. But I will not attempt that here and will instead rest on the suggestiveness of the particular case. We can always represent an instantaneous action as residing in a difference in which propositions are true before and after a certain point.[30]

If I am right that judgments of instantaneous actions can be paraphrased in this way, then we know that durative actions, or at least some sub-set of them, are fundamental. And if that is so, then we know that perfective judgments must be only one side of an aspectual coin. Recall that in the case of durative action, if X A-ed, then X was A-ing. If Jones walked to school, then Jones was walking to school. The latter, imperfective judgment, "Jones was walking to school", reports an action in progress or under way and has imperfective aspect and is the correlate of the perfective "Jones walked to school." It seems then that action is a form of what occurs and what can be occurring.

*Atelic action*. Let's now consider the case of atelic action. Take, for example, the action reported by "X walked." I'll will follow most authors on the subject by supposing that where

---

[29]From this perspective we can say with Galton that "although the change can only be located within an interval, and not at a moment, it is still an instantaneous change because there is no lower limit to the length of intervals which contain it" (*Logic of Aspect*, p. 34).

[30] The reader will have noticed that my selection of types of instantaneous action has made the task of finding paraphrases too easy. Even from a grammatical point of view starting and stopping are unique among instantaneous actions since with them the following holds: If X stopped then X stopped something else. What's needed is an algorithm for generating paraphrase reports of any instantaneous action. And what's needed for launching such a project is a paraphrase of a instantaneous action like "X won the race." Obviously I have not provided either of these.

A-ing is atelic, "X was A-ing" implies "X A-ed."[31] So, if X was walking, then X walked. If this is right, then a straightforward analysis of perfective judgments concerning atelic actions is available. In making an atelic perfective judgment one is simply saying that something is in a certain state for a while and then stops being in that state. We already know that we do not need perfective judgments to describe stopping. And suppose that we do not need perfective judgments to describe starting. We also do not need perfective judgments to characterize something's being in a state for a while, a present tense proposition being true for a while. Returning to the case of walking: X walked if and only if X was not walking and then for some interval X was walking and then X was not walking. Once again, I have only given an analysis of a particular case, and have not given a general formula or algorithm for producing paraphrases of the relevant type. But I will not attempt that here and will again rest only on the suggestiveness of the particular case.

Even so, if I am right that judgments of atelic actions can be paraphrased in this way, then durative telic actions are fundamental, if any are. And so, perfective judgments must be only one side of an aspectual coin, the other side of which has the following peculiar characteristics: (Incomplete) "X is A-ing" implies X has not yet A-ed, e.g., that Jones is walking to school implies that X has not yet walked to school;[32] (Interrupt) "X is A-ing" does not imply that it will be that X A-ed, e.g., that Jones is walking to school does not imply that it will ever be

[31] I express some misgivings about this test because it obscures the fact that this inference goes through when and only because there is no first moment of A-ing. When this is so, "X was A-ing" is made true by a preceding period of A-ing and it is a different thing that one simultaneously is A-ing and has A-ed. But this is not what sustains the inference from "X was A-ing" to "X has A-ed" in the case of state verbs, for example, seeing red. Here if at a moment X is seeing red, then we can say of that moment X has seen red. I have profited from G. E. M. Anscombe's "Before and After," *Philosophical Review* 73 (1964), and from Terry Penner's "Verbs and the Identity of Actions: A Philosophical Exercise in the Interpretation of Aristotle," in *Ryle: A Collection of Critical Essays*, eds. O. Wood and G. Pitcher (New York: Anchor Books, 1970).

[32] There are some complications with (Incomplete) that I am just going to ignore for now, in particular that it is false when Jones has walked to school on some other occasion.

that Jones walked to school. So, where we can think of things as having happened or occurred, we can think of something as being in the process of happening, and where we can think thoughts of that shape, we can think what it would be for the process to be interrupted and left incomplete.

*Durative telic action*. If perfective aspect is, in a sense, mere window dressing in reports of instantaneous and atelic action, then if anything necessitates perfective judgment, reports of durative telic actions do. It is, of course, easy to raise the question of paraphrase here as well. And if it is answered affirmatively, we should wonder whether perfective judgment is only a highly convenient, maybe indispensable, way of speaking. However, the strategy for paraphrase I considered for the other cases looks hopeless when applied to durative telic case. There I suggested that recourse to moments, intervals and states could capture the phenomenon. But here no such appeal would seem to work. How long must the interval span for the telic perfective judgment to be true? The obvious answer is as long as it takes. But that just means as long as it takes for the process to come to completion. The idea of completion or culmination is ineliminably aspectual.[33]

### *2.4 Means and ends*

The claim is that the capacity to act is internally related to the capacity to put means to ends. We have just described a very spare conception of the former but what is the latter exactly?

---

[33] It's worth mentioning one recent attempt to analyze perfective judgments. In "On the Progressive and the Perfective," (*Nous* 38 [2004]: 29-59), Zoltan Szabo proposes the following: "Jones walked to school" is true if and only if (i) "Jones is walking to school" is true of some event e; (ii) "Jones is at school" is true of some state s; (iii) e causes s; (iv) if e causes e' and e' causes s then "Jones is walking to school" is true of e'. But how are we supposed to understand (iii)? Obviously if it means that e caused s, then Szabo's putative analysis is in trouble because "caused" is perfective. But without that interpretation it is very difficult to hear (iii) as making a singular causal claim at all. This isn't a decisive objection but hopefully it's enough to give a sense of how the development of such an objection might go.

We grasp a capacity through its exercise. When the practical capacity to put means to ends is exercised, a certain sort of explanation of action is in the offing: one in which the action is shown to be towards the realization of an end. There is a lot of specialized vocabulary in this neighborhood and to avoid confusion I just want to use the expression *instrumental* to pick out our topic.

When talking about *instrumental explanation of action* I am talking about a certain kind of answer to Anscombe's reason seeking question "Why?" It is, I think, the real focus of *Intention* as well as Davidson's essays on action and is exhibited in the following exchanges: Why are you replenishing the water supply? To poison the inhabitants; Why are you flipping the switch? To turn on the light. But "to do B" does not, as such, express a complete thought, and we might gain clearer access to this guy's meaning by pressing him in the manner of a second grade teacher: "Answer in complete sentences!"[34] A natural move is for him to abandon the apparatus of question and answer and adopt a straightforwardly explanatory device by employing the explicitly teleological connective "in order to" as follows: I'm replenishing the water supply in order to poison the inhabitants; I'm flipping the switch in order to turn on the light. I rely on this specific formulation, what I call the *criterial presentation* of instrumental

[34] What is given as a response in each is not a complete thought but rather a thought fragment, what Jennifer Hornsby calls a thing done and what I will call an event form. In general, explanation is a relation that holds between propositions or facts and is given basic expression by "p because q." In the special case of action explanation, what falls on the left side is, obviously, an *action*. We need to be careful here because "action" is categorially ambiguous. Hornsby has alerted our attention to the fact that "action" is sometimes used to pick out concrete particulars which are possible subjects of many descriptions: for example, my flipping the switch, my turning on the light, my illuminating the room. She calls actions in this sense *doings*. Hornsby also notes that "action" is sometimes used to pick out things people do; these are universals: for example, to flip the switch, to turn on the light, to illuminate the room. She calls actions in this sense *things done*. I have no objection to any of this, by my lights a useful distinction, though I think that it is also helpful to mark a third sense of "action," one with an entirely different categorial significance. Neither standing for individuals nor for universals, in the third sense an action is a fact and not simply a component of a fact, as each of the others is. When the phenomenon of action is exemplified in this sense, it might have either of the following elementary and familiar shapes: "X is A-ing" and "X A-ed." I mention this to note the terminological policy of this essay, namely to use "action" to pick out facts. For further discussion see Jennifer Hornsby, *Actions* (London: Routledge & Kegan Paul, 1980).

explanation, in the official characterization: an explanation of action is instrumental if and only if it welcomes transposition into final-causal formulation of this specific sort—*X is A-ing in order to B*.

Of course, there are other devices for presenting *explicitly* teleological explanation of action: I'm replenishing the water supply for the sake of poisoning the inhabitants; I'm flipping the switch with the aim (goal, objective, end or purpose) of turning on the light.
Each of these serves to deliver our guy's account of what he's doing, and each is teleological on its face. However, I do not mean to say that every explicitly teleological explanation of action is instrumental. At the very least, there is variety on the surface: I'm leaving for the sake of Smith; I'm fighting for the sake of justice; I'm moving to Pittsburgh with the aim (goal, objective, end or purpose) of knowledge. There is diversity in the grammatical category of the end: proper nouns, abstract nouns and from earlier verbs. And it is a genuine question what significance to attach to these grammatical differences. In particular, we are confronted with a question about whether to take the appearances as indicating differences in the category of the explanans and the structure of the explanation. But however we resolve that, we can still agree that this flexibility of the words "aim", "purpose", "goal", "objective" and "end" as well as the teleological connective "for the sake of" raises such questions. In contrast, the connective "in order to" is unambiguous, taking only verbs as complements. What distinguishes instrumental explanation of action from what appears in other species of teleological explanation of action (if, in fact, there are such) is the character of the end, goal, aim or purpose. As we will understand it, in instrumental explanation "X is doing A for the sake of B" where B is an event form. As I said, these are expressed by verbs of a certain kind, namely those which can be given

in answer to "What did X do?". So, in instrumental explanation the end in the light of which someone's doing something is explained is always something that can itself be done.

To be clear, instrumental explanation of action is not simply what has the surface grammatical structure "X is A-ing in order to B"; rather the idea is that this is the criterial presentation of something which makes other appearances in what we say and think. A *derivative presentation* of instrumental explanation is an explanation of action which welcomes transposition into our final form but which is not itself explicitly teleological.[35] The *philosopher's favorite* is, of course, X is A-ing because X wants to B and believes that A-ing is towards B-ing.[36] While much of the contemporary action theory literature is concerned with questions about which derivative presentations are fundamental, or even legitimate, the initial points I want to make about derivative presentations and their relation to the criterial presentation can be made even if we leave such disputes to the side, and restrict our attention to the philosopher's favorite.

Generally speaking, in the move from the criterial to the derivative two things happen: (i) the explicitly teleological connective "in order to" which joins a proposition and an event-form is replaced by the all-purpose explanatory "because" which joins propositions; (ii) derivative presentations explicitly register structure that is merely implicit in the criterial presentation, namely that instrumental explanation has two components, what I will call a *ground* and a *consideration*.

These two components have been characterized, at one time, or another, as respectively *orectic* and *doxastic*, desiderative and intellectual, conative and cognitive. In this spirit, we

[35] In a complete account, more would need to be said about what it is for an explanation to *welcome* transformation or paraphrase into the criterial presentation.
[36] I use the awkward expression "towards" to leave open the structure of the means-end relation.

might say that the ground is the spring, arche, origin or principle of the action and provides its motivating energy or orientation. The consideration, on the other hand, is the rational link between the ground and action and is what allows us to regard the action as potentially the conclusion of an inference. It is sometimes said that explanation of action by reasons explains why the agent acts as he does by revealing the favorable light in which the agent sees his action. Holding fast to this metaphor we might say that the ground is the source of light while the consideration focuses and directs this onto the action explained.

To see that these components, ground and consideration, are, in fact, implicitly contained in the criterial presentation, first consider that if I say "Jones is going to the Hereford market in order to buy a cow," you could contradict me by saying "Jones does not want to buy a cow." The applicability of instrumental explanation implies that the agent is a subject of the relevant instrumental ground. And so, appreciation of an instrumental consideration, e.g., going to Hereford is towards buying a cow, is not sufficient to instrumentally-move someone to action on its basis. This is also suggested by the fact that the same set of instrumental considerations might lead to different actions: awareness that strong alkaloids are deadly poison to humans, that nicotine is a strong alkaloid, that what's in this bottle is nicotine, might lead to careful avoidance of a lethal dose or suicide by drinking a bottle. It all depends on what you want.[37] Now, also consider that you could contradict my claim "Jones is going to the Hereford market in order to buy a cow" by saying "Ah, yes, Jones wants to buy a cow but does not think that going to Hereford is towards that." The applicability of instrumental explanation implies that the subject appreciates a connection between what she is doing and

[37] This is of course compatible with the fact that mentioning both components is sometimes otiose. I take the point and example from G.E.M. Anscombe, "Von Wright on Practical Inference," in *The Philosophy of Georg Henrik von Wright*, ed. Schlipp (La Salle: Open Court, 1989), pp. 377-404. Also see Anselm Müller, "How Theoretical is Practical Reason?" in *Intention and Intentionality: Essays in Honor of G. E. M. Anscombe*, eds. C. Diamond and J. Teichman. (Ithaca: Cornell University Press, 1979).

what she is going for; where instrumental explanation gets hold, there must be appreciation of a consideration linking what one wants to do and what one is doing. And so, possession of an instrumental ground is also not sufficient to move someone to act on its basis.

It is an unquestionable condition on anything purporting to be an explanation that it be potentially informative. It is plain that "X is A-ing in order to A" is not potentially informative and so not an explanation. To reply to "Why are you A-ing?" by saying "in order to A" is really just to say "for no particular reason," a reply which gives the question "Why?" application but which does not supply a genuine answer to it. So, in instrumental explanation what one is doing and what one wants to do must be distinct types of thing and the consideration linking action and ground must have substantial content, i.e., in "doing A is towards doing B" A and B must be distinct event forms.[38]

Those developing an account of instrumental explanation have tended to fix on derivative presentations, indeed have tended to proceed as if a complete understanding of "X is A-ing in order to B" must travel through those. Why is that? In the first instance, there are those who regard the move away from "in order to" and towards "because" as a necessary step in the elimination of teleology from the theory of action. Typically the strategy is to give a reductive analysis of final causal explanation of action in terms of some concept built out of efficient cause. Although in recent years such attempts at reduction have become very sophisticated,

[38] If "X is A-ing in order to A" is not an explanation, then there can be no derivative presentation of it. For example, "X is buying a Picasso because he wants to buy a Picasso and believes that buying a Picasso is towards buying a Picasso" is not a derivative presentation of a technical explanation. (See Michael Smith's "The Humean Theory of Motivation" for a problematic view of the matter.) Someone might resist by pointing out that, on some occasions at least, that "I want to A" is explanatory of my A-ing, say when I eat some cake because I really really want to. I can't argue for this here but I think that the wanting that figures there is of a different *kind* than that which figures in instrumental explanation. Indeed such wanting, appetitive wanting perhaps, can explain what I am calling "instrumental wanting". Notice that if we shift to other states which seem to be able to figure in instrumental explanation no such ambiguity arises and the candidate explanations are clearly empty: X is A-ing because X is trying to A, intends to A, or is A-ing.

their spirit is nicely expressed by an early effort of Richard Braithwaite: "Teleological explanations of intentional goal-directed activities are always understood as reducible to causal explanations with intentions as causes; to use the Aristotlean terms, the idea of the 'final cause' functions as 'efficient cause'."[39] However, even if one doubts that such an analysis is possible, say, because of the obstacle posed by deviant causal chains, one might still have reason to think that focusing on derivative presentations is necessary for locating the place of mind in nature. As I understand him, Davidson sees the shift to derivative presentations as a necessary step in illuminating the subterranean nomological and mechanistic causal story to be told about the items implicitly appearing in those explanations. Unless we shift our attention to derivative presentations there is nothing that might serve as a candidate for an item appearing in rationalization which is a possible bearer of both mental and physical properties and which is the efficient cause of action. The point of bringing up these two motives for attending to derivative presentations is mostly to put them, and more generally, questions about how to understand the place of mind in nature aside.

Even when we stand back from such problems in the theory of action, there are still reasons for looking to derivative presentations for a deeper understanding of instrumental explanation. First, in instrumental explanation, the *explanans* and *explanandum* stand in a content involving relation and the shift to derivative presentations is necessary for displaying this; moreover having a grasp of these relations is necessary for treating action as the conclusion of an inference.[40] Second, derivative presentations also help us get into view what we might call the explanatory flexibility of instrumental grounds as is exhibited here: Jones is

[39] Richard Braithwaite, "Causal and Teleological Explanation", in *Purpose in Nature*, ed. John Canfield, (Englewood Cliffs: Prentice-Halll, 1966), pp. 27-47, p. 32.
[40] For development of this point see Davidson, "Actions, Reasons, and Causes," p. 9, "Intending," pp. 85-87, and "Hempel on Explaining Action," p. 263, all in *Essays on Actions and Events*.

putting a dime in the soda machine because Jones wants to get a Coke, Jones wants to get a Coke because Jones wants to give a Coke to Smith. The point is that in instrumental explanation what figures as the ground is itself a possible subject of instrumental explanation. In particular, instrumental wanting is suited to figure as either a instrumental explanans or instrumental explanadum: it can supply an answer to "Why?" but it also makes sense to ask "Why?" of it. The explanatory flexibility of instrumental grounds is precisely what permits *chains* of instrumental explanation: X is pumping because X wants to replenish, X wants to replenish because X wants to poison the inhabitants.[41]

In the discussion of derivative presentations thus far, I've been focusing on the philosopher's favorite. However, there are others, or anyway, others have been proposed. In particular, each of the following has been proposed as instrumental ground: wanting, intending, trying, and acting itself. Each proposal has also had its share of critics and much of the recent work on action and its explanation has revolved around the question whether such diversity is genuine or merely superficial. But to the tribunal of common sense, there seems nothing wrong with any of these. In everyday life, we say such things all the time: I'm A-ing because I want to B; I'm A-ing because I intend to B; I'm A-ing because I'm trying to B; I'm A-ing because I'm B-ing.

Each of these grounds is capable of generating *homogeneous* chains. We displayed this above with the philosopher's favorite, but now notice that the same trick can be turned with

[41] A near relative of this point is that instrumental wanting is suited to figure either as the premise of a practical syllogism or as the conclusion. In addition, since usually when someone is in pursuit of a instrumental end, there is a series of things that must be done in order to achieve it, there will also be a chain of practical inferences which can be imputed to the agent, though of course this does not mean that the agent had to have that series of thoughts explicitly occur to her. The distinction I am operating with is different than Thomas Nagel's well known distinction between motivated and unmotivated desires. According to Nagel, a desire is motivated when it is had for a reason and unmotivated when not. But the distinction figuring here is instead between forms of desire for which reasons can be given and those for which they cannot: this leaves it open that one might on a particular occasion have an instrumental want for no reason at all.

each of the others: I'm A-ing because I intend to B and I intend to B because I intend to C. Consider another case: I'm A-ing because I'm trying to B and I'm trying to B because I'm trying to C. Finally, what is least common in the action theory literature and perhaps most common in ordinary life: I'm A-ing because I'm B-ing and I'm B-ing because I'm C-ing. What is most interesting is not just that each of the instrumental grounds is suited to generating homogeneous chains but that the whole set permits the generation of *heterogeneous* chains like the following: I'm A-ing because I want to B, I want to B because I intend to B, I intend to B because I'm trying to C, I'm trying to C because I'm D-ing. Earlier we said that instrumental wanting could be both a instrumental explanans and explandum, so too, it seems can each of the other instrumental grounds.[42] In what follows, I will feign indifference to the question which derivative presentations are legitimate and basic.

### *2.5 Durative telic action and instrumental explanation*

I have been arguing that the distinctive kind of judgment associated with action has its fundamental application in the case of durative telic action. We now need an argument linking being the subject of durative telic action with being the subject of instrumental explanation. I'll say something more suggestive than decisive in the service of that now.[43]

---

[42] Much of this section is devoted to showing how much I mean to include under the heading of technical agency, and also to showing that in the condition of agency argument I do not want to rely on the details of any of the diversity within technical explanation. I am following Michael Thompson in thinking that the diversity is only of limited significance. For others expressing a similar tendency see Anscombe, *Intention* §23, and Paul Churchland "The Logical Character of Action-Explanations," *The Philosophical Review*, 79 (1970): 214-236, p. 231.

[43] Someone will have noticed that most of what I have said about being the subject of perfective judgments and the priority of perfective judgments reporting durative telic action seems to apply to the subjects of events quite generally. Do I plan to argue that if something is the subject of a durative telic event then it must be able to do one thing in order to do another in the sense described in section three? At first glance that would be an extreme view. Consider the following claims: Mars is orbiting the sun for the second time, the ivy is creeping to the top of the gate, and Lassie is running to the police station. Each of these is of the durative telic type. Do I mean to argue that planets, bits of ivy and dogs are instrumental agents? I do not. But how can I avoid the implication?

The thing to observe about durative telic action is this: it is possible for one to be under way without ever coming to completion, and when one does come to completion it will consist of several phases, during each of which we can say that the relevant durative telic action is in progress. It might be that I was walking to school even though, as it turns out, I do not walk to school. And if I do walk to school, then before I make it there, there are other things that I do, say walk across the street, during which we can also say that I am walking to school.

We can now ask two questions. First, what does it consist in for a durative telic action to be under way, when it does not come to completion? What does it consist in for me, say, to have been walking to school, when I do not ultimately walk there? The second question has, instead, to do with durative telic action that does come to completion: what gives the unity to its various phases or stages? In virtue of what do other things that I do come to constitute my walking to school? In each case, the answer must, I think, advert to a truth of the form "X is doing A in order to do B", where B is walking to school.

---

Here's a radical two step approach: first argue on behalf of (Telos) if something is a subject of a durative telic event, then it is the subject of a teleological process; second defend a very restricted conception of the place of teleological explanation which holds that all teleological processes are actions. This would presumably involve giving up on the appearance that "Mars is orbiting the sun for the second time" is true or makes sense. Actually, this is not as difficult to maintain as might first appear. Notice that if the planet does not do what we said it was doing, we have grounds to retract the initial claim. If such grounds are sufficient for retracting the claim, then we are just denying that the durative telic is in play. But how easily will we retract the claim that Mars was orbiting the sun for the second time, when it doesn't make it all the way round? There are other difficulties for the radical approach; in particular, it must manage the appearance that there is, in fact, non-intentional teleological explanation. In any case, I want to adopt a more cautious approach.

I simply ask that we restrict our attention to a sub-class of events which are actions. To mark out this class, I suppose that "X A-ed" reports an action just when it gives Anscombe's sense of the question "Why?" application. In doing so, I am taking two features to be essential to action and not to mere events: they and their explanation are known without observation by their subjects. In the discussion that follows, I rely on these features and so do not generate an argument that entails that any subject of an event must be able to do one thing in order to do another. In an ultimately satisfying account much more would have to be said about what these features are and why they are present where there is intentional action.

To see this consider the conditions under which we are willing to say that X is doing A in order to do B, say, that Jones is walking across the street in order to walk to school. First, Jones must think that she is walking across the street. A second condition is that Jones think that she is walking to school. Finally, Jones must think that walking across the street is a way of making progress towards walking to school. Where it is a question of intentional action, what ties parts to whole together has to do with thought, and all of these thoughts are necessary, for consider a case where one of them is absent.

Imagine some distance from A to B. Now imagine that Jones walks from A to B, which is mid-point on the way to C. If during this initial segment Jones does not think that walking to B is part of walking to C, two things follow. First, if at B Jones is struck by a bus, we will not say that an act of walking to C was interrupted. Second, if in the end Jones is at C, even having occupied every point on a line stretching from A to C, we will not say that Jones intentionally walked from A to C. Where the conditions of instrumental explanation are not met, we withhold the claim that a doing of A was an interrupted doing of B, and also withhold the claim that a doing of A was a phase of an intentional doing of B. If there is something other than the instrumental connection which accounts for the possibility of interruption or the unity of a completed action, I do not know what it could be.

### *2.6 Conclusion*

I began by mentioning that recent work on instrumental reason has been focused on problems about the content and authority of instrumental requirement. I have not touched on any of that. Still, the conviction guiding this chapter is that if we leave unquestioned the thought that I have been investigating, the thought that any agent must have the

capacity to do one thing in order to do another, or as Railton puts it, that agents "necessarily possess and act on ends," we are taking too much for granted, and will not ultimately achieve a satisfactory account of the normativity of instrumental reason. To understand why, and in what sense, agents *ought* to take the means to their ends, a question about practical norms, we need to understand why agents *must be capable* of taking means to ends, a question about the metaphysics of agency. My hope is that by pursuing this metaphysical question, we will ultimately see that performing an action is the same thing as doing one thing in order to do another, that a whole action is composed of parts which are themselves actions, and that the means-end connection is simply the sort of unity that parts and wholes of this kind have. In any case, this chapter has been a provisional and limited investigation of these latter metaphysical topics, although with the hope that it might at some point contribute to a proper grasp of the nature of instrumental requirement. I return to that in chapter four.

# INTERLUDE: ANALYTICAL KANTIANISM AND THE INSTRUMENTAL OUGHT

Perhaps Kant is exactly right when he says,

> That he has reason does not at all raise him in worth above mere animality if reason is to serve him only for the sake of what instinct accomplishes for animals; reason would in that case be only a particular mode nature had used to equip the human being for the same end to which it has destined animals, without destining him to a higher end.[44]

I do not know. I only want to ask whether he makes any sense in employing the idea of an agent whose reason is limited to taking means to ultimate ends themselves immediately given by desire.

And perhaps Hume is exactly right when he says,

> It appears evident that the ultimate ends of human actions can never, in any case, be accounted for by *reason*, but recommend themselves entirely to the sentiments and affections of making, without any dependance on the intellectual faculties. Ask a man *why he uses exercise*, he will answer, *because he desires to keep his health*. If you then enquire, *why he desires health*, he will readily reply, *because sickness is painful*. If you push your enquiries farther, and desire a reason *why he hates pain*, it is impossible he can ever give any. This is an ultimate end, and is never referred to any other object.[45]

I do not know. Again, I only want to ask whether we can make sense of the idea of a merely instrumental form of rational agency: in it desire supplies certain objectives, immediately and without calculation, and instrumental rational calculation moves from there.

---

[44] Immanuel Kant, *Critique of Practical Reason*, 5: 61.

[45] David Hume, *An Enquiry Concerning the Principles of Morals, Appendix I*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Oxford University Press, 1975), 244.

Well, actually, I want to claim that we can make sense of this. Put it this way: *a purely instrumental agent is possible*. That is, it is possible that there is some sort of creature whose practical reason operates in such a way as to be *completely* characterized with the resources of instrumentalism. Maybe this is how it is with the Martians? The Martian would then be a rational agent, of a sort, and yet would be entirely limited to employing reason in the service of achieving further aims. This claim reaches beyond *analytical instrumentalism* and is a potential battleground, largely because it is incompatible with the currently popular position of *analytical Kantianism* according to which something other than instrumental agency is necessary. But why think that there is no place in the realm of possibility for a purely instrumental agent, as I have been arguing there is not for an atomic agent? There are, I suspect, many sources of resistance, though I will mention only two.[46]

The analytical Kantian tends to couple an abstract conception of a principle of practical reason (as possessing *universal validity*) with a substantive view about what these are (often including one governing the determination of so-called ultimate ends). For a principle of practical reason to be universally valid is for it to necessarily bind all beings capable of rational conduct. As Allen Wood puts it "insofar as what we take to be rational really is such (i.e., really is rationally binding), it must be understood in terms of the same *fundamental* principles."[47] Given even a very weak connection between *ought* and *can*, it seems that the limited practical capacities of a purely instrumental agent exclude a place for its being subject

[46] I will not remark on the apparent tension between this passage of Kant's and what I am calling analytical Kantianism, except to say that I have been influenced by Stephen Engstrom's "Contradictions in the Will" and John Rawls' *Lectures on the History of Ethics* (Cambridge: Harvard University Press, 2000) to take those appearances seriously. Also see Sergio Tenenbaum, "Speculative Mistakes and Ordinary Temptations: Kant on Instrumentalist Conceptions of Practical Reason."

[47] Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999) p. 57.

to standards governing the determination of an agent's ultimate ends.[48] So the first source of Kantian resistance comes from thinking that if a purely instrumental agent is possible, then, say, the categorical imperative or the axioms of rational choice theory cannot be principles of practical reason. This is incredibly abrupt, though I hope sufficient to provide some sense of what the Kantian thinks is at stake in the question of the possibility of a purely instrumental agent and why it must be resisted. I do not attempt to address this deep concern in any of what follows, largely because of the great difficulty of doing so, and here put it aside.[49]

There is, in any case, a more immediate source of resistance to our putative possibility and it is the one that I examine. Let me begin with some metaphors. Being a rational agent, everyone will agree, is in part to possess a capacity to *make things happen*, to *intervene* in the flow of events, to *contribute* to what happens. Action is *self-directed* behavior and an agent has *authority* and *ownership* of what it does, it *participates* in what it does—it *does* (in some emphatic sense) what it does. The Kantian intuition, as I understand it, is that *self-governed activity* cannot get hold in a purely instrumental agent; it would be in no more of a position of authority with respect to what it does than a mere brute—in each there is only blind and passive obedience to the promptings of nature and not *self-guided* conduct. It is, thus, not really an agent.

There are many ways in which the intuition might be transformed into an argument. Perhaps the most common strategy is to bring some content to the suite of metaphors, say, by

---

[48] Elsewhere I try to describe a connection between the explanatory and normative dimensions of practical reason which is loose enough to avoid the most common objections to *internalism*, yet strong enough to underwrite the thought that if something is an purely instrumental agent, then it cannot be subject to more than purely instrumental norms. I call the relevant interpretation of internalism the practical capacity requirement.

[49] If there is any hope of ultimately defusing the Kantian worry and not just attacking the position, one must develop a conception of a standard of action that applies to what it applies to by necessity, is known a priori, and yet, does not apply to *all* agents necessarily. The difficulty of the task is, I am sure, obvious though I remain hopeful that it might be done, in part because of the recent work by Philippa Foot (*Natural Goodness* [Oxford: Oxford University Press, 2001]) and Michael Thompson (*Life and Action*).

holding that if something is self-guided, it must possess some specific characteristic, and then to argue that a purely instrumental agent could not possess that. In what follows, I will have in view someone who holds that an agent is both subject to norms and also potentially moved by an awareness of norms.[50] But on reflection, the Kantian argues, we cannot make sense of a purely instrumental agent acting according to a conception of a principle, and even more specifically we cannot make sense of it acting according to a conception of the principle enjoining one to take the necessary means to one's ends, *the instrumental principle*.

Now, I am not aware of a good argument for *analytical Humeanism*, according to which nothing other than instrumental agency is so much as possible, and I do not take the doctrine to be true.[51] In the absence of such an argument, any defense of the possibility of the purely

---

[50] I mean to echo Kant's thought that "Everything in nature works in accordance with laws. Only a rational being has the capacity to act *in accordance with the representation* of laws, that is, in accordance with principles" (*Groundwork*, 4: 412).

[51] Michael Smith's "The Humean Theory of Motivation" is taken by some to have supplied such an argument. But I do not see how this can be. It is an undefended assumption of Smith's essay that "reason explanations are teleological explanations" (*The Moral Problem*, p. 116). He depends on that to argue that all action explanation is of the belief-desire type. But it's the content of the assumption that is really at issue.

Smith is insulted by the intimation that a quasi-hydraulic conception of reason explanation is behind his adherence to the thesis that all action explanation is of the belief-desire sort. Perhaps he is right that the Humean is not committed to it—my own defense of the coherence of instrumentalism makes no such commitment. Still, it is interesting that he is blind to the contentiousness of the thought that nothing other than teleological explanation of action is possible. It might even suggest that he is in the grip of a quasi-hydraulic conception of rational motivation, a suggestion that becomes more plausible when we see that he may be in the grip of a quasi-hydraulic conception of teleological explanation of action in particular. Notice that he takes "X is A-ing in order to A" (the limit case) to be a genuine instance of making an action intelligible in terms of the pursuit of a goal. But it manifestly is not that, as I argued earlier. Smith misses the difficulty because he considers this sort of case in the following guise: X is A-ing because X wants to A and believes that if he does A, he will do A. The crux of his argument connecting teleological explanation to belief-desire explanation is the thought that having the goal of doing A just is having a desire to do A. However, if desire does not play the role of having a goal in the limit case, and is still supposed to be explanatory of action, we might wonder how to understand its explanatory role. It is not implausible to see Smith as implicitly operating with a quasi-hydraulic conception of desire and its influence on action in this case, especially in his thought that it needs to be there if rational motivation can take place. But we might also wonder whether the explanatory role of desire in a typical case is any different.

I have just stipulated that Humeanism has a certain content. But it is a difficult question what exactly unites the many self-styled followers of Hume. I suspect that many who claim that they are Humeans will not recognize themselves in the position whose coherence I am advocating. This is fine and, I think, interesting. If I am right that such a minimal form of agency is possible, then how does the legitimate Humean understand the status of the additional capacities which differentiate their preferred model of agency from what I've been calling the purely instrumental agent?

instrumental agent must proceed slowly, considering what is said against it bit by bit. The next two chapters take up only one little bit and address precisely the charge that a purely instrumental agent cannot be subject to *standards of correctness* in action. That is, the next two chapters address the charge that a purely instrumental agent cannot be a subject of practical *norms*, *requirements* or *oughts*, in particular those given by the so-called *instrumental principle*. As an organizational matter, the discussion orbits Korsgaard's "The Normativity of Instrumental Reason" one of the most exciting recent efforts in the analytical Kantian tradition. There Korsgaard argues along the following lines:

(i) If an agent is under a principle, then it must be able to violate it.
(ii) If something is an agent, then it must be under the instrumental principle.
(iii) So, if something is an agent, then it must be able to violate the instrumental principle.
(iv) But a purely instrumental agent cannot violate the instrumental principle.
(v) So, a purely instrumental agent is not under the instrumental principle.
(vi) So, a purely instrumental agent is not really an agent.[52]

What follows, I am anxious to say, aspires to more than Korsgaard interpretation. Even apart from contributing to a defense of the possibility of a purely instrumental agent, I also mean for the discussion to contribute to our understanding of the metaphysics of practical normativity, and more specifically of the *oughts* internal to *in order to*. Before pressing on, it might help to have in view a sketch of the discussion to come.

In the next chapter, "Practical Reason and the Possibility of Error," I investigate the premise linking practical reason and the possibility of error. It is what I call the *error constraint*. The error constraint is ambiguous. According to the *logical interpretation*, an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby

[52] The most sustained development of the argument is in "The Normativity of Instrumental Reason" but she advances it many other places as well: *The Sources of Normativity*, "Self-Constitution in the Ethics of Plato and Kant," "Reply to Ginsborg, Schneewind, and Guyer," *Ethics*, 109, October 1998, 49-64 and most recently in her Locke Lectures (Oxford University, 2002).

violate the principle. The logical interpretation is concerned with what is possible independent of facts about the agent. In contrast, the *imperatival interpretation* is concerned with what is possible given facts about the agent, and says that an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle and it is possible for the agent, given her particular capacities, to do it. I argue that the apparent plausibility of the latter derives in part from conflating it with the obviously true logical interpretation, and that we are in any case not forced by argument to accept the error constraint on its imperatival interpretation. Moreover, I show that because the imperatival interpretation entails (a) the liberty of indifference conception of freedom and (b) the impossibility of a perfectly rational agent, it is flatly incompatible with *constitutivism*, the Kantian strategy for elucidating the validity of a principle of practical reason by revealing that it is constitutive of a form of mental activity, and the chief alternative to *platonism* among non-reductive and realist accounts of practical norms. So, the decision about how far to write the possibility of error into practical reason ultimately restricts how we can understand the nature and authority of principles in the first place.

In chapter four, "There Is No Such Thing as the Instrumental Principle," I investigate premise (ii) which expresses a substantive interpretation of instrumental reasoning.[53] As I have already said, it is a starting point of much of the theory of practical reason that if something can reason practically, then it can reason instrumentally. As Korsgaard understands it, to reason instrumentally is, in part, to be "motivated and guided" by the instrumental principle. But what is that? I suggest that in Korsgaard's hands the instrumental principle enjoins an

[53] The interpretation is shared by, among others, Thomas Hill, "The Hypothetical Imperative," in his *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca: Cornell University Press, 1992), pp. 17-37, and Allen Wood, *Kant's Ethical Thought*.

agent to have resolve in the pursuit of its ends or to exhibit the virtue Hume calls "strength of mind," and so suggest that her thesis is really that to possess resolve in the pursuit of ends, something must be able to engage in non-instrumental reasoning about what to do. This is an interesting claim but, by itself, does not pose a challenge to the coherence of instrumentalism. Korsgaard must also defend this particular interpretation of the standards involved in calculating how achieve a particular objective. In opposition to Korsgaard's conception, I propose a competing and very minimal interpretation of the standards governing instrumental calculation. In particular, I argue that the only standard internal to instrumental calculation is supplied by an agent's particular objective. While I remove the threat to the coherence of the purely instrumental agent, I nevertheless transform Korsgaard's analytical Kantian argument into a challenge to the analytical Humean by arguing that if an agent can exhibit constancy or inconstancy in the pursuit of an end, it must be more than a purely instrumental agent. Since human agents do, human agents must be more than instrumental agents and so there can be something other than purely instrumental agency.

## 3. PRACTICAL REASON AND THE POSSIBILITY OF ERROR

### *3.1 The error constraint*

That there is a deep connection between reason and the possibility of error is these days a philosophical commonplace:

> For a creature to be correctly said to have a rule, it is necessary that it should be able to break the rule.
>
> The physical or causal possibility of making a mistake, or doing what one is obliged, by what one means, intends, believes, and desires, not to do, is essential to the conception of such states and shows the essentially normative nature of their significance.
>
> An agent may be mistaken about what he has reason to do.…This is essential to preserving the point that statements of what people have reason to do have normative force; no account that excludes this can be adequate.
>
> Reason-giving explanations require a conception of how things ideally would be, sufficiently independent of how any actual individual's psychological economy operates to serve as the basis for critical assessment of it. In particular, there must be a potential gap between the ideal and the specific directions in which a given agent's motivations push him.
>
> There is no normativity if you cannot be wrong.[54]

[54] Jonathan Bennett, *Rationality: An Essay Toward Analysis* (Indianapolis: Hackett Publishing, 1989), p. 17; Robert Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994), p. 14; Bernard Williams, "Some Further Notes on Internal and External Reasons" in *Practical Reasoning*, ed. Elijah Millgram (Cambridge: MIT Press, 2001), pp. 91-97, pp. 92-93; John McDowell, "Might There Be External Reasons?" in his *Mind, Value, and Reality* (Cambridge: Harvard University Press, 1998), p. 105; Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), p. 161.

Of course, many others have explicitly endorsed the connection: Stephen Darwall, "Internalism and Agency," *Philosophical Perspectives* 6 (1992): 155-174; James Dreier, "Humean Doubts about the Practical Justification of Morality," in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (New York: Oxford University Press, 1997), pp. 81-99; Donald Hubin, "The Groundless Normativity of Instrumental Rationality," *Journal of Philosophy* 98 (2001): 445-468; Peter Railton, "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action," in Cullity and Gaut, eds., *Ethics and Practical Reason*, pp. 53-79; John Searle, *Rationality in Action*

Insistence on the link between reason and the possibility of error is just insistence on the normative character of reasoning. On any account, reasoning is activity governed or guided by norms, rules, standards, principles, etc., and so the very idea of this activity must contain the distinction between correct and incorrect application of them. Indeed, to most anyone who has thought about what a principle is and what it is for an individual to be responsible or subject to such a thing, it will seem that: *a reasoner is subject to a principle only if the reasoner can go wrong in respect of it*. And on this basis, most philosophers have accepted as a condition of adequacy on any account of being subject or responsible to a principle that it be able to distinguish between a reasoner's correct and incorrect application of the principle, between her faithfulness and unfaithfulness to it.

Our abstract claim has application to reasoners in their theoretical and practical capacities—as both knowers and doers—and so constrains accounts of both the understanding and the will; it is only the latter, however, which is of interest to me here. In the sphere of action and will, then, we might put the point as follows: *an agent is subject to a principle only if the agent can go wrong in respect of it*. This is the *error constraint* and will be the focus of what follows. While the error constraint certainly has the ring of truth, I'll argue that it is not entirely clear what proposition is expressed by this form of words and that very much hangs on the resolution of the ambiguity and the subsequent determination of which renderings of the error constraint are true. So, let me begin by sketching two ways of understanding how the possibility of error might figure as a condition of being subject or responsible to a principle. I will then raise some questions about the two renderings of the error constraint that will occupy us for the remainder of the essay.

---

(Cambridge: MIT Press, 2001), chap. 3; R. Jay Wallace, "Normativity, Commitment, and Instrumental Reason" *Philosophers' Imprint*, vol. 1, no. 3 (2001): http://www.philosophersimprint.org/001003.

The error constraint is a complex proposition and the crucial ambiguity resides in the *can go wrong* component. But before commenting on that, it is necessary to say something, however sketchy and incomplete, about the others—namely about the concept *principle* and the relation *being subject to*, or, as I will often put it, *being under*. As I'll understand it here, being subject to or being under a principle is simply being positioned within the scope of its authority; loosely speaking, the typical case includes being in a position to act as authorized. And as I'll understand it, a principle is simply what an agent can abide by or follow. I mean to include under this heading such refined objects of practical philosophy as those possessing universal and necessary application to agents; but I also mean to include more ordinary objects like positive law, etiquette, crafts and games, which do not possess universality and necessity. A principle is what I will call a *formal principle* if necessarily, all agents are under it; a principle is what I will call a *substantive principle* if this does not hold.[55] While this gives us a highly antiseptic idea of a *principle* and of the relation *being under*, it does so with a view to allowing the initial scope of this chapter to be as wide as possible and also to avoid settling by fiat many of the questions steering it.

Now, a putative principle formulable as "Do A or don't do A" or "You must either do A or else not do A" isn't something which an agent can violate. Its logical form prevents us from describing or thinking anything at all that would count as a deviation from it: for this reason

[55] There is a tradition of thought about practical reason according to which it is simply a capacity for the mechanical application of general and stateable rules in particular circumstances. And by adopting the expression "principle" to pick out norms of action, and the phrase "being under a principle" to pick out being assessable in the light of a norm of action, I may give the impression that I mean to endorse this model. But that would be a mistaken impression: in speaking of "principles" and "being under principles" I merely assume, with the authors quoted at the very beginning of this section, that agents and actions are assessable in the light of norms. Whether norms have the form of mechanically applicable rules is not settled by this assumption.

There is also a tradition of thought about practical reason according to which principles of reason are formal principles in my sense. I have adopted the expression "formal principle" in order to remain neutral on that question as well.

we want, I think, to say that its violation is logically impossible. And for this reason in turn we want to deny that an agent can be under it. Quite generally, the *logical interpretation* of the error constraint says that *an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle*. It quite reasonably insists, more or less, that an agent is under a principle only if there is something for the agent to go wrong in respect of; really, the logical interpretation is about the principle and demands that this effect a division within the space of action types.

The logical interpretation is concerned with what is possible independent of facts about the agent and this is why we can determine whether a "principle" can be violated by an agent, in this sense, simply by looking at its content, or more accurately, at its aspiration to have content. In contrast, the other interpretation is concerned with what is possible given facts about the agent. It has to do with the coupling of agent-to-principle, or again with the specific nature of an agent's relation to the principle. Here, "can go wrong" means that it is possible for the agent to act out of accord with the relevant principle, perhaps because of *akrasia*, ignorance, vice, or some other feature of the agent. On our second interpretation, *an agent is subject to a principle only if there is some kind of action such that if the agent did it she would thereby violate the principle and it is possible for the agent to do it*. According to this interpretation of the error constraint (itself subject to further refinement in section three) an agent is under a principle only if she is in one way or another imperfectly hooked up with it. For reasons to be explained later, I'll call this the *imperatival interpretation*.

It can now be said that the target of this chapter is the question of the truth of the error constraint, on its imperatival interpretation. Since I'm assuming that being an agent, i.e., having the capacity to employ concepts in the service of action, has an essential normative

dimension which I've expressed as *being subject to principles*, we might also characterize the main thread of this chapter as an inquiry into whether liability to error is a condition of agency or will.

I begin with a suspicion that the seeming plausibility of the imperatival interpretation derives in large part from conflation with the error constraint on its obviously true logical interpretation. Exposing this conflation is the aim of the next section: there I develop the logical interpretation and suggest that the main appearance of the error constraint in the tradition of practical philosophy has been in the guise of the logical interpretation. Because the error constraint is ambiguous, it is not easy to establish commitment to the imperatival interpretation of it, though the thought lurks in the background of much of the recent work on practical reason. And while our real target is imperativalism in all its forms, I will address Christine Korsgaard as its most dynamic and articulate advocate. She is especially clear about endorsing the imperatival interpretation and almost unique in advancing considerations on its behalf. Furthermore, Korsgaard's influential argument for the claim that *instrumentalism* is flatly incoherent employs the imperatival interpretation as a premise: to the extent that imperativalism is impugned, we are in possession of a defense of the *coherence* of instrumentalism.[56] Section three investigates this connection between imperativalism and

[56] The argument is developed at length in Korsgaard's "The Normativity of Instrumental Reason" (in Cullity and Gaut, eds., *Ethics and Practical Reason*, pp. 215-254). Of course, neutralizing Korsgaard's argument does not prove that instrumentalism is coherent—there may be other ways of administering the poison. Even so, it is worth asking why the more general project of vindicating the *coherence* of instrumentalism is of any interest. The most straightforward answer would be that it is a step towards establishing its *truth*. Seems a small step in that direction, though: consider how anemic a defense of the truth of, say, the phlogiston theory of combustion would be on the grounds of its coherence. Leaving aside the question of instrumentalism's truth, for the moment, we can still see that its mere coherence would create trouble for anyone hoping to anchor more-than-instrumental principles in an abstract analysis of rational agency, a strategy typically associated with the Kantian tradition in practical philosophy. Here's the trouble. If instrumentalism is coherent, then it is possible that there is some sort of creature whose practical reason operates in such a way as to be *completely* characterized with the resources of instrumentalism. Maybe this is how it is with the Martians? The Martian would then be a rational agent, of a sort, and yet would be entirely limited to employing reason in the service of achieving further aims. On its face, this mere possibility places a wedge between

instrumentalism, and, in the course of doing so, specifies *weak* and *strong* versions of imperativalism, directing all of our subsequent attention to the latter. Then, to put us in a position to evaluate the imperatival interpretation, section four reminds us of a few abstract features of Kant's own account of the *imperative* and its distinctive role in the rational explanation of action. Section five tries to understand the allure of imperativalism, concluding that it is not forced on us by argument.

Even if the first five sections succeed in specifying several senses of the error constraint and undermining arguments purporting to compel adoption of the imperatival interpretation, the question of imperativalism's truth is still left open. The final three sections make some progress towards closing that, though they do not get it all the way shut. Still, even without proving that imperativalism is false, these sections reveal why it ultimately matters whether it's true or not.

In sections six and seven, I extract a pair of suspect implications of the imperatival interpretation—the conception of freedom as the liberty of indifference and the impossibility of perfectly rational agency. That imperativalism entails such theses is not a *reductio ad absurdum*, though many may thereby be spurred to reject it, perhaps because of prior and deeper commitments, say, to determinism or theism. However, my focus is elsewhere, chiefly on the way these implications, and thus imperativalism, confine our understanding of what it

---

possession of practical reason and being subject to more-than-instrumental standards, and disrupts the attempt to derive the latter from the former.

What, then, of the *truth* of instrumentalism? This is a much longer story and one entirely independent of this chapter. Still, what I would like to say is that the question itself becomes difficult to place if it is fixed that instrumentalism is coherent, and also that we humans beings exhibit reason giving and asking behavior that outstrips anything adequately described with the limited resources of instrumentalism. In that case, instrumentalism would capture how it is with Martian practical reason and yet fail to capture how it is with ours. But now what are we to make of the very *question* of instrumentalism's truth? Hopefully, this compressed line of thought can, when expanded, bring us closer to seeing the place for an analysis of the concept of practical reason on which it admits of *species* or *forms*; the various positions, e.g., instrumentalism, might then be seen as simply purported specifications of forms of agency. The tendency of the current literature is very different; it is to treat the various positions as articulations of competing conceptions of some single concept of practical reason. But this is all off stage here.

is to be under a principle in the first place. Our conception of the metaphysics of practical normativity is confined, I argue in section eight, because imperativalism is incompatible with *constitutivism*: a strategy for elucidating the authority of a principle by revealing it to be internal to what is under it, and the main alternative to *platonism* among non-reductive and realist accounts of practical norms. That constitutivism is also a strategy Korsgaard perspicuously articulates and adopts as her own is of local interest but does not, of course, decide the question of how far to write the possibility of error into agency. About that we must settle for more modest, conditional results: if we go as far as imperativalism would have us go, we are prohibited from a constitutivist understanding of principles, and then forced to choose between improbably writing them into "a world beyond, which exists God knows where," or drearily giving up on the idea of there really being standards of correctness for action.[57]

### *3.2 The logical interpretation*

To see the force of the logical interpretation of the error constraint it will help to consider two occasions of its use. In book one, chapter three of *On the Social Contract*, "On the Right of the Strongest," Rousseau tells us that the principle "Obey those in power," where it has the sense of "yield to superior physical force," is merely apparently meaningful. Its absurdity is revealed by the fact that "it will never be violated," i.e., its violation is inconceivable. But why can't we think its violation? In a characteristically adept turn of phrase Rousseau reveals the

[57] G. W. F. Hegel, *Elements of the Philosophy of Right*, ed. Allen Wood, trans. H. B. Nisbett (Cambridge: Cambridge University Press, 1991), p. 20.

Limited space requires developing the incompatibility claim in the restricted context of formal principles; though I believe that a parallel point can be made in connection with substantive principles, our official results are nevertheless suitably restricted. That so much of the current debate about the content of the principles of practical reason gets played out in terms of a debate about what the formal principles are ensures that this restriction does not sever our discussion from the core of the current discussion.

reason: "As soon as force makes right, the effect changes along with the cause. Any force that overcomes the first one succeeds to its right."[58] His idea here is that the truth of any judgment of relative physical power is determined or fixed by the outcome of actual struggle, so that "X is more powerful than Y" just means "X overcomes Y in a struggle." Given this, it must be that the superior physical power is always obeyed. And if having a right is understood in terms of being strongest or having superior physical power, it must be that the strongest has what is his by right.

In his second sermon at the Rolls Chapel, "Upon Human Nature," Bishop Butler employs a structurally identical argument against the principle "Act as you please," which he interprets as recommending "following that principle or particular instinct which is for the present strongest."[59] Butler develops the argument as follows:

> If by following nature were meant only acting as we please, it would indeed be ridiculous to speak of nature as any guide in morals, nay, the very mention of deviating from nature would be absurd; and the mention of following it, when spoken of by way of distinction, would absolutely have no meaning. For did ever any one act otherwise than as he pleased?[60]

Butler's thought is that if every action arises from desire, and if the truth of any judgment of the strength of desire is determined by the outcome of actual struggle, then it must be that the strongest desire is acted on. Here, it is inconceivable that an agent act, yet not act on the strongest desire, thus it is impossible that an agent be subject to the principle "Act as you please."

The preceding interpretation of Rousseau's objection to the doctrine of the right of the strongest isn't without its difficulties. I'd like to raise one with a view to saying a bit more

---

[58] Jean-Jacques Rousseau, *On The Social Contract*, ed. Roger D. Masters, trans. Judith R. Masters (New York: St. Martin's Press, 1978), p. 48.
[59] Joseph Butler, *Fifteen Sermons Preached at the Rolls Chapel*, reprinted in *Five Sermons Preached at the Rolls Chapel and A Dissertation upon the Nature of Virtue*, ed. Stephen Darwall (Indianapolis: Hackett Publishing, 1983), p. 35.
[60] Ibid., p. 36.

about the structure of the positions that run afoul of the logical interpretation. I have in mind someone who points out that "power" has more than one sense, only one of which is amenable to my account. On the one hand, and perhaps most commonly, we think of a power as a capacity—as something with a kind of generality. Power, in this sense, is something one might have at a time even if not exercising it just then, like Harry Houdini's powers of sight and locomotion when he's blindfolded and stuffed inside a coffin. There is also a sense of "power" as a particularized and occurrent force—the sort of force an individual tsunami rushing towards Honolulu is thought to have. If Rousseau is operating with the former understanding of "power," then the truth of a judgment of relative power is *not* determined by the outcome of conflict and in a conflict between powers the greater power might lose. We could then correctly say "Monsieur R. had the power to defeat Monsieur D." even when he didn't. We might explain, "If he hadn't slipped on that banana, he certainly would've won the duel." Consequently, the strongest might not have what is his by right. So, if Rousseau's objection to the doctrine of right of the strongest is as I've taken it to be, then we ought to find him appealing to the particularized sense of power. And this is precisely what he is doing when he says that "Yielding to force is an act of necessity, not of will."[61] Were "yielding to force" a species of yielding to something general, to a capacity, then yielding would have to be an act of will, what we might regard as an expression of prudence or, perhaps, terror. This is because one must be able to yield to such a thing even when the power or capacity isn't being exercised, and this sort of yielding can only be the expression of will. So, we must take Rousseau to be operating with the particularized sense of power and force in his argument against the very sense of the attempt to explain what it is for X to have a right against Y in

[61] Rousseau, p. 48

terms of facts about the relative strength of X and Y. Now, I take it that behind Butler's gloss of "act as you please" as recommending "following that principle or particular instinct which is for the present strongest" is an acute awareness of this complication about the general and particular senses attaching to "power" and "strength," as well as a commitment to a highly particularized sense.

Our original "Do A or don't do A" wears its emptiness on its sleeve and so only appears in real life as the expression of helplessness or frustration: "Keep drinking or don't. I don't care." Because they are much better dressed, "Yield to the superior physical force" might appear as the basis of a doctrine of political right, and "Act as you please" might appear as the basis of a doctrine of human flourishing—nevertheless, each is as empty.

The emptiness of these purported principles can be brought into somewhat better focus by attending to the following. The execution-conditions of a principle are just the truth-conditions of a certain proposition. And in all of our cases the relevant propositions are necessary truths. In the "Do A or don't do A" case, the necessity of the corresponding proposition derives from its logical form, while in the interesting cases the necessity of the corresponding proposition derives from a relation between the contents of the component expressions. The shift from the transparently absurd to the philosophically tempting cases is something like the shift from "Kill a bachelor or don't kill a bachelor" to "Kill a bachelor or don't kill an unmarried male." The particularized understanding of power, force and strength that Rousseau and Butler employ is really in the service of establishing a meaning connection between "X is more powerful than Y" and "X defeats Y in a struggle" and also one between "Z is the strongest desire" and "Z is the desire acted on." With such connections in place "Yield

to the superior physical force" and "Act as you please" become very much like the second bachelor case.

Rousseau's thought is not simply that it is false that the right of the strongest provides the foundation for political right, but that the idea of the right of the strongest is an "inexplicable confusion" and "meaningless." According to Butler, "Act as you please" is "absurd" and "ridiculous" as a principle of life. The incoherence of these derives from their purporting to be standards of correctness while also being necessarily in accord with every describable state of affairs. This is so because their supposed content is specified in terms of what, in fact, happens. One would like to say: whatever happens is what ought to happen and that only means that here we can't talk about what ought to happen.[62]

---

[62] I allude to Wittgenstein's famous remark at §258 of the *Philosophical Investigations*, trans. G. E. M. Anscombe (New York: Macmillan Publishing, 1958) to suggest that he employs the logical interpretation in the course of attacking attempts to understand the content of "sensation language" on the model of private inner ostensive definition. His point there is that on such a picture "sensation language" could not provide a "criterion of correctness," and so could not be said to mean anything at all.

It might be worth noting that what sits at the foundation of the projects Rousseau and Butler devastate is very much like the organizing impulse of the dispositional account of meaning or content Saul Kripke considers in *Wittgenstein on Rules and Private Language* (Cambridge: Harvard University Press, 1982), pp. 22-37. While not interested in what it is to have a right, or what it is to follow nature, Kripke is nevertheless investigating something normative—in particular, the fact in virtue of which an expression has meaning. The dispositionalist answer is that facts about how a language user is disposed to use an expression determine the meaning of the expression. Kripke argues that the dispositional view is ultimately an equation of performance and correctness, and for this reason can't make sense of the possibility of mistake. This, in turn, reveals that the dispositionalist can't meet a basic condition of adequacy on accounts of meaning—namely, that any candidate for the fact in virtue of which an expression has meaning must be such as to ground the normativity of meaning, i.e., must contain a specification of the correct use of that word. In the end, each of the views attacked aims at grounding and explaining the relevant norms in baldly naturalistic and reductive terms, and each consequently washes up on the logical interpretation of the error constraint.

Although I have not touched on revealed preference theory or its connection with the theory of rational choice, it seems clear that the criticisms this coupling has undergone in, say, the work of Simon Blackburn and David Gauthier are of the Butler-Rousseau type. See Blackburn's "Practical Tortoise Raising," *Mind* 104 (1995): 695-711, as well as his *Ruling Passions* (New York: Oxford University Press, 1998), pp. 161-168. And see Gauthier's *Morals By Agreement* (New York: Oxford University Press, 1986), pp. 26-29.

### *3.3 Imperativalism and instrumentalism*

In this section, I develop the imperatival interpretation of the error constraint, including a specification of weak and strong versions, by considering an occasion of its use. In particular, I consider Korsgaard's employment of imperativalism in her attempt to show that instrumentalism is simply incoherent.

Most philosophers are familiar with G. E. M. Anscombe's poisoner—"I'm pumping in order to replenish the house water supply," "I'm replenishing the house water supply in order to poison the inhabitants"—or if not with this fellow, then with Donald Davidson's scribe: "I'm writing the letter 'a' in order to write 'act'," "I'm writing 'act' in order to write 'action'." But one needn't have read Anscombe or Davidson to be familiar with the subject matter of their philosophical reflections—the sort of explanation of action exhibited here, broadly speaking explanation of the form "X is doing A in order to do B". That is, we are all familiar with teleological explanation of action, or again explanation of action in terms of an agent's further purpose. And we are, I suppose, equally familiar with the form of advice and form of practical inference internally related to this. Let's call an agent whose intentional operations admitted only such explanation, who took only such advice, and who conducted only such practical inferences, a *purely instrumental agent*. Moreover, we might call the view that this is all there is to practical reason *instrumentalism*.

Recently it has become fashionable to speak of the *instrumental principle*, and then to characterize cases of doing A in order to do B as following the instrumental principle, to characterize being suited to receive advice like "you ought to do A seeing as you propose to do B" as being subject to the instrumental principle, and along the same lines, to

characterize the instrumentalist as holding that the instrumental principle is the *only* formal principle of practical reason, i.e., as holding that the principle enjoining one to take the (necessary) means to one's end is the only principle of which it is true that necessarily, if X is an agent, then X is under it.[63] With such ideas and terminology in place, we can make the following claim: *something is an agent only if it is subject to the instrumental principle*.

Now, a starting point of this essay is that being an agent, or having the capacity to employ concepts in the service of action, has an essential normative dimension, which I've formulated as *being subject to a principle*. Thus, when we are asking about the conditions of something's being subject to a principle we are also asking about *the conditions of agency*. Imperativalism is a component of a view about this; it is a specification of an essential or necessary feature of being an agent. And like imperativalism, the claim that something is an agent only if it is subject to the instrumental principle is also a condition of agency claim, though of a slightly different sort; it is a substantive claim about which principles an individual must be under if that individual is to be an agent at all. It is a substantive view about formal principles and not, like imperativalism, an abstract specification of the conditions of being subject to principles simply.

Some, like James Dreier, have argued that the real appeal of instrumentalism derives from the special status of the instrumental principle, a status he argues no other principle has: "The special status of instrumental reason is due to its being the *sine qua non* of having reasons at all."[64] Others, like Korsgaard, have argued that while being under the instrumental principle is

[63] Elsewhere I would complain about this way of describing instrumentalism and its conception of end-oriented agency, but such complaints would be a distraction here, and so I'll follow in suit.
[64] Dreier, p. 99.

indeed a condition of agency, there is only an appearance of uniqueness here and that when we tease out the conditions of instrumental agency itself, we will appreciate that "if there are any instrumental requirements, then there must be unconditional requirements as well."[65] The reason I'm raising such matters in this chapter is that Korsgaard's analysis makes an ineliminable appeal to the imperatival interpretation. I'd like to look into this.

In "The Normativity of Instrumental Reason," Korsgaard argues against the neo-Humean doctrine that "the instrumental principle is the *only* requirement of practical reason." That is, she argues against the doctrine that purely instrumental agents are the only agents there are. More specifically, Korsgaard defends the thesis that "the view that all practical reason is instrumental is incoherent." But why are we to accept that there cannot be purely instrumental agents? According to Korsgaard, a purely instrumental agent "cannot violate the instrumental principle" and for this reason cannot be under it.[66] But how are we to hear her claim that a purely instrumental agent can't go wrong? Along logical or imperatival lines?

Since the logical interpretation is weaker than the imperatival interpretation, it is obviously permissible for an imperativalist to rule out a view on the grounds that it does not meet the weaker logical condition. So, someone might grant that Korsgaard is an imperativalist, which she undoubtedly is, but still recommend reading her attack on the coherence of instrumentalism as structurally similar to the Butler-Rousseau arguments I described earlier.[67] Were this interpretation correct, putting pressure on the imperatival

[65] Korsgaard, "The Normativity of Instrumental Reason," p. 252.
[66] Korsgaard, "The Normativity of Instrumental Reason," p. 220; ibid., p. 251; ibid., p. 231.
[67] Melissa Barry raised the possibility of this reading of Korsgaard's argument in helpful comments on an ancestor of this essay. For Korsgaard's explicit endorsement of imperativalism see "The Normativity of Instrumental Reason," p. 236 and p. 240, as well as, *The Sources of Normativity*, pp. 29-30, pp. 137-138, p. 146, and pp. 161-164.

interpretation would not help to vindicate the *coherence* of instrumentalism, as I'm hoping it can.

But this reading of Korsgaard cannot be correct. Recall Butler's argument: from the fact that X is not acting on a particular desire, it follows that that desire is not strongest. If so, it is impossible that X fail to act on the strongest desire. And thus, an agent can't be under the principle "Act on the strongest desire." The principle under consideration in Korsgaard's essay is the instrumental principle: Take the (necessary) means to your end. It is clear that this principle cannot fall to a Butler-like argument once in the instrumentalist's hands. This is because the instrumentalist does not, and need not, make the relevant meaning connection between "X has an end" and "X is doing what is in fact necessary for achieving an end." According to a reasonable instrumentalist, from the fact that X is not doing what is necessary for doing A, it does *not* follow that X does not aim at doing A. This is enough to enable the instrumentalist to describe a state of affairs which would be out of accord with the instrumental principle. For example, according to the instrumentalist, from the fact that D. D. is not turning left, when turning left is necessary for getting to Katmandu, it does *not* follow that D. D. doesn't have the objective of getting to Katmandu—he might be reading the map upside down. What does this tell us about the place of the error constraint in Korsgaard's argument against instrumentalism?

Since the instrumentalist can satisfy the error constraint on the logical interpretation, charity recommends reading Korsgaard's core argument against instrumentalism as employing another interpretation. But, as the plausibility of our example shows, it is also true that the instrumentalist can satisfy the error constraint on the imperatival interpretation—at least, on the assumption that instrumental agents are susceptible to error in action that derives from

some defect of theoretical reason. Moreover, Korsgaard grants that the instrumentalist is entitled to admit cases of error in action of this character, miscalculations and mistakes. But she also says "the possibility of mistake is not in general very interesting," and does not permit the instrumentalist to locate the requisite place for the possibility of error by appealing to it.[68] Since the instrumentalist can satisfy the error constraint even on the imperatival interpretation, and because Korsgaard seems to see this, charity requires reading her core argument as employing another interpretation of the error constraint. But what is it?

To the extent that exercises of practical rational capacities depend on exercises of theoretical rational capacities, error in the former can be derivative of error in the latter: a faithful man who has nevertheless forgotten a promise made some time ago might, for that reason, fail to keep it; a physically adept woman who is simply ignorant of a relevant means-end connection might thereby fail to achieve her end. However, in such cases the error in action does not lie specifically in the misuse of a practical capacity. Instead, it derives from error that is already present in the conditions of the exercise of the relevant practical capacity. And such error does not express *genuinely practical imperfection*.[69]

Given the distinction between derivative and genuine practical imperfections, we are in a position to distinguish weak and strong versions of the imperatival interpretation. The *weak imperatival interpretation* is what we introduced early in this chapter: an agent is subject to a principle only if there is some kind of action such that if the agent did it she would violate the principle and it is possible for the agent to do it, and it doesn't matter why the agent does it.

---

[68] Korsgaard, "The Normativity of Instrumental Reason," p. 227.

[69] What are the distinctively practical imperfections? An adequate answer to this question would require the specification of the forms of practical thought, the specification of the modes of reason in action. This is because if there is a kind of practical defect then there must be a kind of practical thought or form of practical reason correlated with it. If this is right, then we can't even begin to touch on a substantive characterization of the forms of practical defect. My thinking about the idea of distinctively practical error and defect has benefited from Stephen Engstrom's "Contradictions in the Will" (unpublished).

The *strong imperatival interpretation* adds a further condition, one having to do with the source or origin of the error. It says that *an agent is subject to a principle only if there is some kind of action such that if the agent did it she would violate the principle and it is possible for the agent to do it and her doing it would be a genuinely practical error*. That is, the error in action must be error that does not depend on error in theoretical reason quite generally and is the expression of a genuinely practical defect.

Once the distinction between weak and strong versions of the imperatival interpretation is in play, we must consider where weak imperativalism fits in the scheme of this chapter. In section one, I said that I would *not* take up the question of the way in which the possibility of error figures as a condition of a *theoretical* reasoner's being subject to norms. But the question of the truth of weak as opposed to strong imperativalism will raise exactly that. So, I put it to the side and focus on the truth of strong imperativalism in what follows.

There is direct evidence that Korsgaard endorses the strong imperatival interpretation. The sort of violation, error or going wrong that Korsgaard has in view is exemplified in cases in which "people's terror, idleness, shyness, or depression is making them irrational and weak-willed." Unlike cases of mistake, the actions that result here are "strictly speaking, irrational," they exhibit "true irrationality, by which I mean a failure to respond appropriately to an available reason."[70] All signs point to an interpretation of Korsgaard's main argument against instrumentalism on which it employs the *strong imperatival interpretation* and is really this: a purely instrumental agent can't be distinctively practically defective (truly irrational or irrational strictly speaking) with respect to the instrumental principle, and for that reason

[70] Korsgaard, "The Normativity of Instrumental Reason," p. 229; ibid., p. 228; Christine Korsgaard, "Skepticism about Practical Reason," in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), pp. 311-334, p. 318.

can't be under it.[71] But why is the possibility of distinctively practical error necessary? Why is strong imperativalism true?

### *3.4 Kant on imperatives*

To put us in a better position to assess the imperatival interpretation, I want to remind us of two very general features of Kant's own account of *imperatives* and their distinctive role in the rational explanation of action, features which have largely gone missing in subsequent debates nevertheless conducted in its terms. We might put the first this way: while the imperative is prior in the order of what is known to us, it is not, according to Kant, prior in the order of nature. The second concerns the restricted place of explicitly

[71] Among other things, the second section of "The Normativity of Instrumental Reason" develops a complex argument for the claim that a purely instrumental agent can't be distinctively practically defective, and, as a whole, has prompted much discussion. Two contributions to the discussion are especially relevant to the line of thought I am pursuing here. In "The Groundless Normativity of Instrumental Rationality," Donald Hubin argues that the crux of Korsgaard's argument against instrumentalism cannot be that it fails to meet the error constraint, or in Hubin's terms, that it fails to be "normatively demanding." This is because he sees how easily instrumentalism can meet both the logical interpretation and the weak imperatival interpretation. Nevertheless, Hubin does not see that Korsgaard's interest is in the strong imperatival interpretation, and thus, I think, fails to address her real concern. David Sobel's "Subjective Accounts of Reasons for Action," *Ethics* 111 (2001): 461-492, also records the ease with which the instrumentalist can make sense of derivative error in action. Although Sobel expresses some concern about how the instrumentalist might meet what I'm calling the strong imperatival interpretation, he does not develop an account of that. As I hope is obvious, I am instead recommending that the putative condition of agency be resisted.

Elsewhere I argue that behind the *strong* imperatival interpretation is the thought that action is not the upshot of theoretical reason operating with a special set of contents—"mince pie syllogizing" as Anscombe puts it—but rather, the upshot of a rational capacity of an entirely different kind. When conjoined with the intuition that kinds of rational capacity are in some sense tightly connected with kinds of error, we come close to the strong imperatival interpretation. However, I argue that we can capture the heart of the intuition without making imperativalist commitments by appeal to the much weaker *practical defect constraint*: *for every kind of practical thought there is a unique kind of practical defect connected with it*. It is a further and non-trivial step to claim that each individual bearer of a capacity for thought of the relevant kind must potentially exhibit such a defect, or really be able to mess up in the relevant way.

This raises a further difficulty. In light of the practical defect constraint, someone might claim that Korsgaard's argument against instrumentalism appeals only to this and not at all to the imperatival interpretation. That is, someone might insist that Korsgaard's complaint with instrumentalism is that it can't make sense of the very idea of genuinely practical error. While there is little textual support for this interpretation, and much against it, there is good reason to reconstruct Korsgaard's argument along these lines, if only to see where it might ultimately go wrong. I pursue this project in chapter four, "There Is No Such Thing as the Instrumental Principle".

prescriptive or deontic thoughts, e.g., "I ought to do A", in the explanation of action, as Kant understands it.

Before moving on, a word of caution: we are considering the content and credentials of imperativalism, a thesis attaching to the abstract genus of *principle* which potentially contains, say, rules of a game, craft and positive law as species. Yet most of what Kant has to say about imperatives takes place against the background of a conception of principles as general, of rational principles as formal principles (in my sense), as well as a highly determinate conception of what the formal principles are. In this section, I bracket any divergence over such matters and restrict my attention in a similar fashion. This is not because I am here giving up on the generality of the claim I'm investigating; rather, ease of exposition of Kant's conception of an imperative recommends such a route.

It is no mere coincidence that in each of Kant's most elaborate discussions of the nature of an imperative in general he first introduces the concepts of a practical principle and a practical law: "Practical *principles* are propositions that contain a general determination of the will…they are objective, or practical *laws*, when the condition is cognized as objective, that is, as holding for the will of every rational being."[72] Such laws contain specifications of what is practically necessary. By "necessary" Kant means "in accord with a rule" and by "practically necessary" he means "in accord with the rule given by objective laws of the good" or again, the rule given by the concept of the good.[73] A practically rational being is, for Kant, simply something with the capacity to act

[72] Immanuel Kant, *Critique of Practical Reason*, 5:19. Also see *Groundwork of the Metaphysics of Morals*, 4:412-414, and *The Metaphysics of Morals*, 6:221-222. All translations are from *Practical Philosophy*, ed. and trans. Mary Gregor (Cambridge: Cambridge University Press, 1996). All references to Kant's writings are given in the notes by the volume and page number of Kant's *Gesammelte Schriften*, ed. the Royal Prussian Academy of Sciences (Berlin: de Gruyter, 1902–).

[73] Kant, *Groundwork*, 4:412-414.

according to a conception or representation of a law. Or again, a practically rational being is something with the capacity to employ the concept *good*, or to act in the light of the rule given by that concept. As with other abilities, capacities or powers, this can be possessed either perfectly or imperfectly. In the first case, where reason infallibly determines the will and what is done is necessarily in accord with the relevant law, "no imperatives hold…the 'ought' is out of place."[74] It is only with the introduction of the idea of an imperfectly rational being, one which has the capacity to act on the representation of laws, but which might also misuse that capacity, that the idea of an *imperative* has any room to appear:

> An imperative differs from a practical law in that a law indeed represents an action as necessary but takes no account of whether this action already inheres by an *inner* necessity in the acting subject (as in a holy being) or whether it is contingent (as in the human being); for where the former is the case there is no imperative.[75]

So, it is no accident that Kant's discussion of the nature of an imperative is preceded by discussion of practical principles and the subset of these which are laws, because an imperative is, in some sense, derivative of these. An imperative is, in essence, a law considered in relation to a certain sort of being, namely an imperfectly rational one: "Imperatives are only formulae expressing the *relation* of objective laws of volition in general to the subjective imperfection of the will of this or that rational being, for example, of the human will."[76]

---

[74] Ibid., 4:414.

[75] Kant, *Metaphysics of Morals*, 6:222.

[76] Kant, *Groundwork*, 4:414, italics mine. Of course, Kant distinguishes between hypothetical and categorical imperatives and some might read his remarks about the good as having application only to categorical imperatives and the sublime rational capacities which underlie them. I'm not reading him this way and think that the thought experiment described in the *Critique of Practical Reason* at (5:58-59) gives me some license to do so.

But when, according to Kant, is something imperfect in such a way as to receive principles as imperatives? Not just any imperfection will do. Were, for example, a creature to lose its capacity for growth prematurely, it would have an imperfection to be sure, though this need not have any bearing on the facility and grace of its reason. But does Kant think that any rational imperfection suffices for turning laws into imperatives, as it were?

Given the distinction between derivative and genuine practical imperfections sketched in the previous section, we are in a position to claim that an agent is related to a principle as an imperative just when the agent has the relevant distinctively practical imperfection. Sometimes this seems to be Kant's view: "[Imperatives] say that to do or omit something would be good, but they say it to a will that does not always do something just because it is represented to it that it would be good to do that thing." However, there are other passages which lend support to a different interpretation—in particular, one on which the possibility of, say, mere ignorance of fact is sufficient for turning principles into imperatives. "The imperative thus says which action possible by me would be good and represents a practical rule in relation to a will that does not straightaway do an action just because it is good, *partly because the subject does not always know that it is good*, partly because, even if he knows this, his maxims could still be opposed to the objective principles of a practical reason."[77] Any attempt at a resolution of this interpretative tension is another essay; nevertheless, it is interesting to describe these conceptions of the conditions under which an agent is related to a principle as an

[77] Kant, *Groundwork*, 4:413. Ibid., 4:414, italics mine.

imperative, if only to emphasize the parallel with weak and strong versions of the imperatival interpretation, each of which Kant would reject.

Now that we've said something about the conditions under which an agent is related to a principle as an imperative, we can ask, "What, if anything, is distinctive about the *relation*?" Since the principles we have in view here are practical, i.e., they figure as the determining grounds of action, we should expect to find something distinctive about the way in which they figure in the determination of an imperfectly rational *will*. The concept which Kant introduces to locate what is special here is *necessitation* (*Nötigung*) or *constraint* (*Zwang*) and is to be distinguished from practical necessity, which he glosses as *goodness*. What he says is that an imperfectly rational being is *necessitated* or *constrained* to act in a certain way by the representation or conception of objective practical principles, or again an imperfectly rational being is *necessitated* to do A by the appreciation that doing A is practically necessary or good.[78]

According to Kant, practical and deontic thoughts, e.g., I must do A, I ought to do A, and I should do A, are expressions of rational *constraint* or *necessitation*. Thus they figure in the practical thinking or among the grounds of action only of imperfectly rational beings.[79] This restricted understanding of the place of deontic thoughts in deliberation and action explanation, if it is in fact Kant's, would not be idiosyncratic. Hume's view on the matter seems to be much the same. Although Hume famously refuses to give "the motive of duty" a foundational, or even significant, place in his account of moral motivation, he nevertheless wants to provide some account of action on its basis:

[78] See, for example, *Groundwork* 4:412-414 and *Metaphysics of Morals*, 6:222.
[79] Kant, *Metaphysics of Morals*, 6:379.

> But may not the sense of morality or duty produce an action, without any other motive? I answer, it may: But this is no objection to the present doctrine. When any virtuous motive or principle is common in human nature, a person, who feels his heart devoid of that principle, may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty, in order to acquire by practice, that virtuous principle, or at least, to disguise to himself, as much as possible, his want of it.[80]

Whatever else we have to say about the passage, this much is certainly contained in it: only what is imperfect, indeed only what appreciates its own imperfection, acts on the motive of duty, and, I would say more generally, on explicitly prescriptive or deontic thoughts. It is a striking fact that these are here a secondary case of acting for a reason. And this should make us pause. In much contemporary work, curiosity about the motivational powers of reason is focused in the first instance on the question of the truth of *judgment internalism*—the thesis that it is a necessary condition of a genuine instance of a judgment that I ought to do A, that I am disposed to act in a way appropriate to it. But the tradition pulls us in a very different direction in holding that acting on explicitly deontic thoughts is not the grassroots level of acting for a reason, and moreover that it has a place only where the agent is imperfect.[81] This is not, I suggest in section eight, a mere historical artifact of Kant's conception of practical reason but crucial to the explanatory strategy he is pursuing.

With so much said about the imperative and its Kantian origins, let's return to the main thread of the chapter. It is, we now know, plainly and trivially true that *an agent is*

---

[80] David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and rev. P. H. Nidditch, second edition (Oxford: Oxford University Press, 1978), p. 479.

[81] For a helpful classification of substantive theses in the theory of practical reason going by the names of "internalism" and "externalism" see Stephen Darwall, "Internalism and Agency" and *The British Moralists and the Internal 'Ought': 1640-1740* (Cambridge: Cambridge University Press, 1995), pp. 9-12. Darwall is right that judgment internalism has figured prominently in contemporary arguments for practical non-cognitivism. Still, as Darwall recognizes, the general orientation is not limited to non-cognitivists. See for example John Broome's "Reason and Motivation," *Proceedings of the Aristotelian Society*, suppl. vol. 71 (1997): 131-146. I have not argued for this restricted conception of action on deontic thoughts. Still, if it is correct, and if imperativalism is false, then we do not address reason's practicality at the fundamental level by addressing the question of judgment internalism.

*subject to an imperative only if it is possible for the agent to violate it*. This much is simply contained in the definition of an imperative. But why think that this is a condition that applies to being under a principle quite generally? Why think that liability to error is a condition of reasoning or principled agency? Why think that principles are imperatives? Or in our terms, why think that the imperatival interpretation is true?

### *3.5 The allure of imperativalism*

Thus far the focus has been the description of three, progressively stronger interpretations of the error constraint: the logical interpretation, and the weak and strong imperatival interpretations. In the course of presenting these, the soundness of Korsgaard's influential argument against instrumentalism was linked to the truth of strong imperativalism. Now the focus will be an investigation of its truth. Fair enough. But is our interest in the error constraint limited to resolving ambiguity and defending the intelligibility of instrumentalism? It is not. When we reach for the error constraint in the first place, even before the various renderings are set out, we do so as a way of bringing some substance to the distinction between occupying a position in the space of reasons and occupying a position in the realm of law, to borrow a suggestive remark of John McDowell's. The several renderings present us with alternatives, and, I will argue, the subsequent decision about which to endorse ultimately restricts how we can understand the nature and authority of principles in the first place. Our reflection on the possibility of error will, thus, turn out to be a partial meditation on the metaphysics of practical normativity. As a first step, in this section we try to understand the allure of imperativalism and ultimately argue that it is not forced on us by argument.

Let's begin by describing a couple of obstacles to supplying a proper *defense* of imperativalism. We have already encountered one, the ambiguity of the error constraint itself. Inasmuch as it is easy to conflate the logical and imperatival interpretations, it is easy to transfer illicitly the obviousness of the former to the latter.[82] There is another, more elusive danger—conflating a discussion of what an *imperative* is with an argument on behalf of the imperatival interpretation. Consider Korsgaard's most robust formulation of imperativalism:

> [Kant] does not explicitly give up the view that the will's imperfection is what makes us subject to an *ought*, but it seems to me that he should have, for imperfection is a red herring here. Even a perfectly rational will cannot be conceived as *guided* by reason unless it is conceived as capable of resisting reason. It may be true, as Kant insists, that a divine will is not subject to temptation and so just would do what reason requires, but it is not true, as he seems to infer, that no *ought* applies to the divine will….Obviously, one of the central ideas of this essay is that we can be subject to normative principles only if we can resist them, because without that possibility they cannot function as guides. But I do not agree with Kant that the absence of any specific temptation to resist them removes the possibility of resistance in the sense needed for normativity. It is not imperfection which places us under rational norms, but rather freedom, which brings with it the needed possibility of resistance to as well as of compliance with those norms.[83]

Earlier I mentioned that it is plainly and trivially true that an agent is under an *imperative* only if it is possible for that agent to violate it. If an agent under an imperative is the topic of this passage—if Korsgaard has in mind *imperatives* and the kind of direction they provide when using expressions like "normative principles," "rational norms," "norms," "guides" and "normativity"—then all of what is said would be plainly and trivially true. Clearly, it is a condition of adequacy on any *defense* of the imperatival interpretation that it not hang on a simple and straightforward appeal to concepts such as these, i.e., concepts which might

[82] John Searle employs strong imperativalism as a premise in an argument for a certain conception of freedom, what he calls "the gap," not worrying that it might itself be subject to challenge. See Searle, *Rationality in Action*, pp. 16-17 and pp. 66-67.
[83] Korsgaard, "The Normativity of Instrumental Reason," p. 240.

equally be used to supply the definition of an *imperative*. This remark on terminology is not intended to be a criticism of imperativalism, but only to show how little work can be done on its behalf by a straightforward appeal to norms, guidance et cetera.[84]

The allure of imperativalism does not vanish altogether when the relevant ambiguities of the error constraint are resolved, or when the threat of mere stipulation is revealed. And we should not expect it to—the attraction of the thought runs deep. But where does it run to? What is the substantive consideration underwriting the imperatival interpretation? It is, I think, something like the following. No mere mechanism can be under a principle or follow a rule; such a thing might only be a locus of mere regularity. But a would-be perfectly rational will—something whose will is in a state of perfection and thus which can't really go wrong—must amount to no more than a strange sort of mechanism or automaton.

Korsgaard expresses this thought when she says that a "perfectly rational will" is something "whose own conduct is not guided by normative principles at all, but instead describable in a set of logical truths."[85] The claim is that principles can be no more than "merely descriptive" of a perfectly rational will, just as physical laws are no more than "merely descriptive" of the trajectories of asteroids and planets. This chain of ideas also finds expression in John Searle's recent monograph, *Rationality in Action*: "In order to behave rationally I can do so only if I am free to make any of a number of possible choices and have open the possibility of behaving irrationally. Paradoxically, the alleged ideal of a perfectly rational machine, the computer, is

---

[84] Korsgaard's own writings do not indicate that she is entirely clear about the risk. There are many passages that suggest that she does not distinguish principles (rules, norms, standards, etc.) from *imperatives* ("The Normativity of Instrumental Reason," p. 217 and p. 236). She takes normativity to be a species of *necessitation* and suggests that *necessitation* is a component of any adequate account of how "reasons direct, guide, or obligate us to act or judge in certain way" (*The Sources of Normativity*, p. 226). Recall that for Kant, *necessitation* is a feature of an *imperfectly* rational being's following a rational principle, and yet, she says that Kant exhibits confusion in holding that a perfectly rational will is both under laws of reason and not *necessitated* to follow them ("The Normativity of Instrumental Reason," p. 239).

[85] Korsgaard, "The Normativity of Instrumental Reason," p. 240.

not an example of rationality at all, because a computer is outside the scope of rationality altogether."[86] One might deny that a mechanism can be under a principle for any number of reasons, though what matters here is that Searle seems to deny that there might be a "perfectly rational machine" on grounds that equally rule out the possibility of a "perfectly rational animal" or a "perfectly rational spirit" or whatever, namely the absence of "the possibility of behaving irrationally." The threat seems to be that at the limit of perfection the agent goes out of view, and the purported solution seems to be that only with the imperatival interpretation can we distinguish the agent-principle connection and the mere object-physical law connection. Let's investigate this.

All sides agree that the way in which agents are directed by principles is radically different from the way in which mere bits of stuff are directed by physical laws. The distinction between these two sorts of "being directed" is often marked as that between active direction and passive direction, activity and passivity. Korsgaard puts the point this way: "It is in the nature of activities, as opposed to mechanical processes, that one who engages in them is self-guided"; "The rationality of action depends on the way in which the person's own mental activity is involved in its production, not just on its accidental conformity to some external standard."[87] Let's call the thought expressed in these passages the *participation requirement* and formulate it as follows: *the efficaciousness of a principle must be mediated by the thought of what follows it*. According to the participation requirement, the relevance of a principle to what happens in the world is mediated by the conceptual activity of that which is under it. The contrast is, of course, with the rather unmediated grip that a physical law has on what is, in some other sense, subject to it. With the advocates of imperativalism, let's grant that the participation

[86] Searle, p. 66.
[87] Korsgaard, *The Sources of Normativity*, p. 236; Korsgaard, "The Normativity of Instrumental Reason," p. 236.

requirement must be met by any adequate substantive account of the way in which agents are related to principles.

The imperativalist seems to picture matters this way: there is not enough space or enough of a gap between a would-be perfectly rational agent and a principle for her to be said to act on a conception of it. But if the connection between a principle and what an individual actually does is not mediated by a conception of the principle, then what the individual does cannot be put down to an agent's activity or participation. What it does must instead be merely the result of the play of mechanical forces. The imperativalist might continue: even if there is some space for self-consciousness or reflection in such a thing, it would be practically impotent. The relevant powers of reflection could place such beings merely in the position of "some leaves blown about by the wind and saying 'Now I'll go this way…now I'll go that way' as the wind blew them."[88]

This chain of ideas gives us a somewhat clearer picture of what it is about would-be perfectly rational wills that purportedly renders them unfit for genuine practically rational activity, though a link is missing. We still want to understand why there isn't enough of a gap between a would-be perfectly rational agent and a principle for her to act on a conception of it. The thought linking unswerving accord with a principle and mere mechanism seems to be something like the following: if X always does A in C, then the explanation of why X did A in C can't be an explanation by reasons. But this goes undefended. And I am now simply asking

[88] The quotation is from Ludwig Wittgenstein, "Lectures on Freedom of the Will," as cited in G. E. M. Anscombe, *Intention* (Cambridge: Harvard University Press, 2000), p. 6. Kant describes a similar case in the teleological argument of the opening pages of the *Groundwork*: "And if reason should have been given, over and above, to this favored creature, it must have served it only to contemplate the fortunate constitution of its nature, to admire this, to delight in it, and to be grateful for it to the beneficent cause, but not to submit its faculty of desire to that weak and deceptive guidance and meddle with nature's purpose. In a word, nature would have taken care that reason should not break forth into *practical use* and have the presumption, with its weak insight, to think out for itself a plan for happiness and for the means of attaining it. Nature would have taken upon itself the choice not only of ends but also of means and, with wise foresight, would have entrusted them both simply to instinct" (*Groundwork*, 4:395).

why we should expect the *that* of unswerving accord to travel together with the *because* of mere mechanism—that connection is not baldly contained in the participation requirement.[89,90]

At this point, the imperativalist might challenge my formulation of the participation requirement. What is needed to capture the special sort of contribution that agents make to what they do, the imperativalist might insist, is a strengthening, perhaps along the following lines: only if an agent can determine for herself whether to comply with such principles as she grasps will her behavior be guided by her and be a result of her activity as an agent. Other strengthenings are no doubt possible.[91] For our purposes, the question is whether any such strengthening can remedy the imperativalist's argument. I doubt that such a move will help.

Either the liability to error is packed into a proposed strengthening or it is not. If it is not, then *prima facie* it will be possible to place a wedge between "participation" and potential

---

[89] The slide is easy to make. In the course of elaborating his earlier remark about the impossibility of a "perfectly rational machine" Searle says, "A computer is neither rational nor irrational because its behavior is entirely determined by its program and the structure of its hardware. The only sense in which a computer can be said to be rational is observer-relative" (*Rationality in Action*, pp. 66-67). Here Searle shifts from complaining about the absence of "the possibility of behaving irrationally" to complaining that the behavior is entirely determined by non-rational elements—very different grounds for holding that something is "outside the scope of rationality altogether."

[90] It might help to forestall misunderstanding by begging the question for a moment. As I have it, a perfectly rational agent does what it does through the exercise of practical judgment. One way to start to spell this out is to say that a perfectly rational agent can ask and answer "Why?" questions as, say, these are characterized by Anscombe in *Intention*. In querying whether the possibility of error is essential to agency, I do not mean to query whether the possibility of reflection on one's reasons is essential to agency as this is exhibited in the asking and answering of "Why?".

The ability to ask and answer such questions is sometimes linked to the ability to construct philosophical theories of practical reason. According to Korsgaard, ordinary reasoning and philosophical reasoning are absolutely continuous: "a person who starts out reasoning in some perfectly ordinary way…finds himself on a route that has no natural stopping place short of the unconditioned Ideas of Reason and the metaphysical perplexities to which they sometimes lead" ("Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind," *Ethics* 109 [1998]: 49-66, p. 61). Kant seems to see matters along similar lines: the principle of morality "is really an obscurely thought *metaphysics* that is inherent in every human being because of his rational predisposition" (*Metaphysics of Morals*, 6:376). Of course, there are other, very different ideas about the degree of continuity of ordinary and philosophical reflection. I raise this immensely difficult and important topic merely to say that in urging us away from imperativalism, I do not mean to urge us away from or towards any particular view about the degree of continuity of ordinary deliberation and philosophical reflection. If such reflection is simply a continuation of ordinary deliberation, then the capacity for such reflection will be a non-contingent feature of being under a principle. If it is not, then it is much harder to see how a case for its non-contingency might be made out. In any case, these questions are simply left open by my investigation of practical reason and the possibility of error. I am grateful to an anonymous referee for *Ethics* for helping me to see the risk of conflating these issues.

[91] I owe the instructive challenge as well as the specific suggestion to an anonymous referee for *Ethics*.

error, as I did above. The subsequent burden would then fall on the imperativalist to defend the connection between unswerving accord and mere mechanism, as I suggested above. Of course this would have to be done on a case by case basis. If the liability to error is packed into a proposed strengthening, then obviously the participation requirement, so construed, entails imperativalism. However, it would then no longer be possible to take it as common ground from which to argue for imperativalism. We have not then, I think, been forced by argument to adopt the imperatival interpretation, though no compelling reason has been given to reject it either.

### *3.6 Imperativalism and the liberty of indifference*

How, then, can we make any progress towards resolving the question of imperativalism? As I've indicated, my plan is to investigate the way in which imperativalism constrains our understanding of the nature and authority of principles. To do that it will help to first spell out a couple of implications of imperativalism: in this section I argue that imperativalism entails the liberty of indifference conception of freedom, and in the next that it entails the impossibility of a perfectly rational agent.

At the end of the long passage quoted at the beginning of the previous section, Korsgaard says that it is by appeal to the concept of freedom that we understand what it is to be under a rational principle:

> I do not agree with Kant that the absence of any specific temptation to resist them [normative principles] removes the possibility of resistance in the sense needed for normativity. It is not imperfection which places us under rational norms, but rather freedom, which brings with it the needed possibility of resistance to as well as of compliance with those norms.[92]

[92] Korsgaard, "The Normativity of Instrumental Reason," p. 240.

Korsgaard is careful to deny that being subject to temptation or felt desire can be the source of the possibility of resistance in the sense she requires; this is because being a subject of felt desire is *not*, while being able to resist *is*, part of what it is to be an agent under a principle, the former can't account for the latter. This complaint with Kant's view of the matter reveals the way in which Korsgaard writes the possibility of resistance into the very idea of an agent under a principle. This is also revealed by her treatment of the source of resistance as itself a capacity or ability: "Even a perfectly rational will cannot be conceived as *guided* by reason unless it is conceived as capable of resisting reason."[93] What, then, are we to make of this freedom which "places us under rational norms"?

Korsgaard is sympathetic to Kant's notion of negative freedom: "Nothing in human life is more real than the fact that we must make our decisions and choices 'under the idea of freedom.' When desire bids, we can indeed take it or leave it."[94] However, the sort of freedom involved in accepting the imperatival interpretation must be quite different; it must be far more radical than freedom from determination by felt desire. Indeed, to be free in this more radical sense seems to involve both the capacity to make a choice for any principle and the capacity to make a choice against any principle. That is, accepting imperativalism for the reason that the freedom of a rational agent brings with it the needed possibility of resistance to and compliance with rational standards commits one to such a radical conception of freedom,

[93] Ibid., p. 240. At certain points Korsgaard takes herself to be involved in a dispute with Kant about how to understand the capacity to resist: "Kant apparently identified our capacity to resist the dictates of reason with the imperfection of the human will" ("The Normativity of Instrumental Reason," p. 239). I think it is unlikely, however, that Kant would accept that there is such a capacity, ability or power.
[94] Korsgaard, *The Sources of Normativity*, p. 97.

which, following the tradition, we might call *libertas indifferentiae* or the liberty of indifference.[95]

It might be useful to locate the imperativalist in an exchange between Kant and Carl Leonhard Reinhold over closely related, though somewhat narrower, matters.[96] Reinhold criticizes the core of Kant's project—the inseparability of freedom and practical reason—on the grounds that it makes it inconceivable how one could freely violate the moral law. Then with a view to accommodating the fact of free and immoral action, Reinhold adopts the position that freedom is the capacity to determine oneself either in accordance with, or contrary to, reason. Kant responds:

> But freedom of choice cannot be defined—as some have tried to define it—as the ability to make a choice for or against the law (*libertas indifferentiae*), even though choice as a *phenomenon* provides frequent examples of this in experience. For we know freedom (as it first becomes manifest to us though the moral law) only as a *negative* property in us, namely that of not being *necessitated* to act through any sensible determining grounds. But we cannot present *theoretically* freedom as a *noumenon*, that is, freedom regarded as the ability of the human being merely as an intelligence, and show how it can exercise constraint upon his sensible choice; we cannot therefore present

[95] In the context of a distinct but closely related inquiry—one about the nature of "moral responsibility" rather than practical rule-following—Susan Wolf describes a kind of agent that is free in this very sense:
"The autonomous agent must be one who is able to act in accordance with Reason *or not*. That is, she must be able to regard the rational course of action, insofar as there is one, as just one alternative among others…this ability to choose among the rational, irrational, and nonrational alternatives alike is not an ability to choose on some higher-than-rational basis. Rather, it is an ability to choose on no basis whatsoever, an ability, if you will, to choose whether to use any basis for (subsequent) choice at all" (*Freedom within Reason* [New York: Oxford University Press, 1990], p. 54). While I object to Wolf's attribution of this position to Kant—*Metaphysics of Morals* 6:226-227 is decisive evidence that he does not hold it— I accept her characterization of just how radical it is. In the course of considering various arguments for the claim that something must be radically "autonomous" in order for it to be a possible locus of "moral responsibility," there is one which Wolf raises only to leave to the side: "A second possibility is that those who continue to insist that radical autonomy is necessary for responsibility do so not because they disagree with my view that the ability to act in accordance with Reason is sufficient for responsibility but because they think that this ability itself requires at least a kind of radical autonomy" (*Freedom within Reason*, pp. 61-62). The imperativalist, I think, might then be seen as arguing for "autonomism" on similar grounds and this chapter might be seen as a contribution to Wolf's defense of what she calls the Reason View.

[96] The matters are somewhat narrower because Reinhold and Kant reserve the concept of freedom for characterizing a being with distinctively moral capacities, while in our discussion a being with any practically rational capacity is free or active in some sense. The matters are closely related because in every case we are concerned with the conditions of being under rational norms. It is just that they reserve the concept of freedom for characterizing the species or form of activity characteristic of beings with moral capacities.

> freedom as a positive property. But we can indeed see that, although experience shows that the human being as a sensible being is able to choose in opposition to as well as in conformity with the law, his freedom as an intelligible being cannot be defined by this, since appearances cannot make any supersensible object (such as free choice) understandable.[97]

Since freedom is not knowable through experience, and since error is knowable only through experience, the latter cannot be a component of the definition of the former. The problem of imputable errors of reason must be solved in another way than by prying apart freedom and practical reason.[98] However, accommodating the mere fact of error, explaining what we encounter in experience, *cannot* be what moves the imperativalist to hold that "the possibility of self-government essentially involves the possibility of its failure."[99] Indeed, we can treat imperativalism as a later stage in the development of the Kant-Reinhold dispute. According to imperativalism, even if nobody has ever, in fact, acted out of accord with what reason prescribes, we would still be in possession of the same materials for insisting on the imperatival interpretation. This is because, as the imperativalist thinks, the liability to error is contained in the very idea of a free and rational agent.

### *3.7 Imperativalism and perfectly rational agency*

According to the imperatival interpretation, the idea of a Kantian *divine* or *holy will* under, say, the principle of promise keeping is as absurd as the idea of an *arbitrium brutum* or mere animal will under the same. Neither sort of being can really go wrong with respect to this

---

[97] Kant, *Metaphysics of Morals*, 6:226.
[98] For discussion of the central importance of this passage in Kant's practical philosophy see Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), pp. 129-136, and also see Allen Wood's essay "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy*, ed. A. Wood (Ithaca: Cornell University Press, 1984), pp. 73-101, pp. 79-83. Notably the passage also figures in Korsgaard's own sympathetic interpretation of Kant in "Morality as Freedom" in *Creating the Kingdom of Ends*, pp. 159-187. This suggests that she will resist the attribution of the liberty of indifference conception of freedom which, of course, is fine, though I do not see how she can distance herself while continuing to maintain imperativalism. I return to this tension in section eight.
[99] Korsgaard, "The Normativity of Instrumental Reason," p. 248.

principle, and so, neither can be conceived as genuinely being under it as a principle. More generally, imperativalism implies that it is impossible for an agent to be *essentially infallible*, or perfectly rational *by nature*. (Thus, imperativalism offers a novel proof of atheism, on the assumption that God is an agent and essentially infallible, certainly the position of classical theology.)

However, even though the imperatival interpretation implies the incoherence of the very idea of an essentially infallible agent, Korsgaard sometimes seems averse to losing access to this concept, no doubt because she is averse to losing contact with the tradition of practical philosophy she has done much to help us understand. She insists, against Kant, that a "divine will" and a "holy will" are also capable of resisting reason and therefore subject to "oughts."[100] But this is very confusing. On the face of it, we can't make any sense of the idea of an essentially infallible agent that is even capable of resistance.[101] If X is capable of doing A, then it is possible for X to do A. So, if something is capable of resisting reason, then it is possible for it to resist reason. But it is simply not possible for an essentially infallible agent to resist reason. Thus, it is not true that an essentially infallible agent is capable of resisting reason. Korsgaard admits that "a divine will is not subject to temptation and so just would do what reason requires," but this isn't an adequate gloss of the excellence characteristic of a *divine* or *holy will*.[102] Instead, the sort of thing Korsgaard is describing here might be found in an Oliver

[100] Ibid., p. 240. See the passage quoted at length at the beginning of section five above.
[101] Although I'm using the expression "capacity to resist," I think we are in some sense already on the wrong track when doing so. A capacity, ability or power is, I would like to say, always something, in some sense, good. Indeed, one way of putting the underlying sentiment of this chapter is that the idea of a capacity or power to resist reason is a confusion on the order of the idea of a capacity not to see, or the treatment of blindness as itself a capacity. See Wood's "Kant's Compatibilism," pp. 81-82, for an interesting discussion of this sense of "capacity." For the parallel between irrationality and blindness, see Saint Anselm's "The Fall of Satan," in *Truth, Freedom, and Evil: Three Philosophical Dialogues*, ed. and trans. Jasper Hopkins and Herbert Richardson (New York: Harper and Row, 1967), sec. 11.
[102] Korsgaard, "The Normativity of Instrumental Reason," p. 240.

Sacks anecdote: a human being who has lost any capacity to experience pleasure and felt desire. There is a significant difference between the idea of an agent that, *as it happens*, is not going to go wrong and the idea of an agent that cannot, under any circumstances, go wrong. Only the second is *essentially infallible* and it cannot also be capable of resistance.

Now, in Kant's system the *holy will* functions as "a practical *idea*, which must necessarily serve as a *model* to which all finite rational beings can only approximate without end and which the pure moral law, itself called holy because of this, constantly and rightly holds before their eyes."[103] Although the imperativalist cannot consistently let a perfectly rational *being* serve as an ideal, perhaps she has the resources to articulate another, more plausible, more terrestrial, conception of a perfectly rational agent; and perhaps this might serve as an ideal actually attainable in the life of that which operates with it. In the tradition, this is done by shifting attention away from characterizing a kind of being—what is perfectly rational by nature, and what I've been calling *essentially infallible*—and towards characterizing a state of will, a disposition, *habitus* or *hexis*. I will now argue that even the more modest conception of perfect rationality—as a state of will non-accidentally issuing only in correct action, though only contingently possessed by its bearer—is unavailable to the imperativalist.

Any characterization of such a practical disposition must permit the distinction between a perfectly rationally ordered state of will and other states of will, some of which, nevertheless, often give rise to similar actions in similar circumstances. In other words, any characterization must permit the distinction between perfect rationality and its imitation. But, I'll now argue, this can't be done without the use of materials inaccessible to imperativalism. If X is in a certain state and has not, in fact, gone wrong it does not follow that X is in an ideal state—X

[103] Kant, *Critique of Practical Reason*, 5: 32.

might have the unblemished record by accident. If X is in a certain state and will not, in fact, go wrong it does not follow that X is in an ideal state—X might acquire the unblemished record by accident. How, then, are we to distinguish the agent in an ideal state from the merely lucky agent, if not according to the totality of past, present and future performance? Actual success doesn't guarantee counterfactual success and it is the latter which is needed to distinguish the ideal from the lucky. In order to distinguish perfect rationality from its imitation, it seems that one has to say something like this: if X is in an ideal state of will, then it must be *no accident* that when X acts, X acts correctly. But can the imperativalist consistently say this? The imperativalist must accept that an agent is liable to error even when in possession of the ideal state of will. That is, within the imperativalist framework, given some agent X which is in the "ideal" state, there must be some circumstance in which X's power to err would be exercised. Were this denied it would be quite dubious whether the capacity for resistance had any explanatory role whatsoever. If there is some circumstance in which X, such as he is, would go wrong, and if it is only an accident that X does not find himself there, then for any state of will that X is in, it must be an accident that X does not go wrong. And so, the imperativalist is not entitled to a conception of an ideal agent as what has a perfectly rational state of will.[104]

Suppose I'm right that the imperativalist is not entitled to either conception of a perfectly rational agent—as essentially infallible or only as possessing a flawless state of will. Should she care? Were we of a different era we might press her in something like the manner of Saint Anselm: "I don't think that freedom of choice is the ability either to sin or not to sin. Indeed,

[104] I have not shown that these are the only terms in which to develop a conception of an ideal agent. So, strictly speaking, I have not proved that no conception of the ideal agent is available to the imperativalist. But what are the alternatives?

if this were its definition, then neither God nor the angels who are not able to sin would have free choice. But to say that they have no free choice is blasphemous."[105] Since we cannot, these days, get any purchase on our adversaries through accusations of blasphemy, we must resort to more subtle and less decisive maneuvers.

### *3.8 The incompatibility of imperativalism and constitutivism*

Sections six and seven articulated implications of imperativalism, respectively, the liberty of indifference conception of freedom and the impossibility of perfectly rational agency. Drawing on these results, the final section, section eight, establishes the incompatibility of imperativalism and *constitutivism*—a strategy for understanding that in virtue of which an agent is under a principle—thereby bringing the question of imperativalism within the scope of problems about the metaphysics of practical normativity quite generally.

The problem of the normative or binding force of a principle is familiar enough, I suppose; still it might be worth trying to say a little bit more in order to provide some context for a description of the constitutivist position. On any account, if an agent is under a principle, then its performance of some action is recommended or sided with. This is just to make the benign point that principles are standards of correctness, something we can read right off the dictionary definition of "principle", whatever that's worth. Now, when an action is sided with we can say so explicitly by employing deontic vocabulary as in "X ought to do A," "X may do B" or "Everyone must do B in C." So, if an agent is under a principle, then some deontic claim with it as subject is true, though I'll concentrate on "oughts". Constitutivism is, then, a reply to "In what does the binding

[105] Anselm, p. 122.

force of a principle consist?" or again to "What it is for an action to be something an agent ought to do?"

Let's narrow this question a bit: In what does the binding force of a *formal* principle consist? As I'll understand it here, the constitutivist finds a way in through reflection on the nature of agency considered as such. Here is Korsgaard's helpful formulation of the position:

> There are in our tradition two things which philosophers have meant by 'reason'. Reason refers to the active as opposed to the passive capacities of the human mind, and 'reason' also refers to certain sets of principles—logical principles, moral and other practical principles, and the principles that Kant associates with the pure concepts of the understanding. What Kant did…was to try to bring these two conceptions of reason together: to explain the normative force of the principles by showing that they are constitutive of reflective mental activity itself. To choose *is* to follow the hypothetical and categorical imperatives; to understand *is* to employ the concepts and principles of the understanding, and so on. And in the same way, my own aim is to portray moral principles [and principles of practical reason generally] as constitutive of, and so as essential to, making human choices, and leading a human life.[106]

Let me try to rephrase this. The constitutivist provides an answer to the metaphysical question "What is agency?" with a view to elucidating the binding force of formal principles. Very generally, the constitutivist's objective is to demonstrate that any particular exercise of rational agency, or action, involves a commitment to comply with certain principles. The objective is to be achieved by showing that the relevant principle is simply a partial description of agency itself; that is, the aim is to be achieved by showing that the capacity to act just is, in part, the capacity to follow the relevant principle. If so, then when one does exercise that capacity one is committed to abiding by the relevant principle—it is a standard internal to the capacity. Obviously the greatest difficulty for any constitutivist account is in showing that

---

[106] Korsgaard, *The Sources of Normativity*, p. 236. Also see "The Normativity of Instrumental Reason," p. 243, "Motivation, Metaphysics, and the Value of the Self," p. 65, "Self-Constitution in the Ethics of Plato and Kant," *The Journal of Ethics* 3 (1999): 1-29, esp. pp. 12-15, and her unpublished "Self-Constitution: Action, Identity and Integrity," delivered as the John Locke Lectures (Oxford University, 2002), which develops this theme in great detail.

there are, in fact, standards essentially contained in such spare materials as the bare capacity for action. But skepticism about that aside, we can grant that were such an analysis correct, there would be no way for an agent to get far enough outside the relevant standards to doubt that it is under them: indeed, no such way for any agent, necessarily.[107]

So, constitutivism is an attempt to extract binding principles from the nature of agency; imperativalism is a thesis about the nature of agency itself. The question is whether an explanatory strategy like constitutivism can be employed while conceiving of agency in such terms. Intuitively the answer is no: in trying to understand what agents ought to do, constitutivism draws us towards what agency is, whereas imperativalism forces us away by insisting that agents cannot be as they ought.

Constitutivism begins from a conception of the capacity to act and a conception of action. Imperativalism insists that to be an agent is, in part, to possess a capacity to act out of accord with any principle whatsoever; and so to be an agent is, in part, to possess the capacity to act out of accord with any formal principle. Let's hold that fixed for the moment and imagine someone trying to apply the constitutivist's argument schema: the capacity to act just is, in part, the capacity to follow some principle, say, the instrumental principle. And so, when one does exercise that capacity one is thereby committed to abiding by the relevant principle, say, the instrumental principle. The difficulty is, I'm sure, obvious. Given imperativalism, it is also

---

[107] What I am calling constitutivism has been explored recently, though in very different ways and under different headings, by others: Stephen Darwall, *Impartial Reason* (Ithaca: Cornell University Press, 1983); David Gauthier, "The Unity of Reason: A Subversive Reinterpretation of Kant," *Ethics* 96 (1985): 74-88; Thomas Nagel, *The Possibility of Altruism* (Princeton: Princeton University Press, 1970); Peter Railton, "On the Hypothetical and Non-hypothetical in Reasoning About Belief and Action"; David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000).

For the constitutivist, the class of formal principles contains those which are fundamental and those which are derivative, i.e., arrived at through the application of fundamental principles. The constitutivist is fixed on the fundamental principles, but with an interest in understanding the basis of those derived from it. It is important to note here that if the fundamental principles are not internal to agency, then neither are any of the principles derived from it.

true that the capacity to act just is, in part, the capacity to violate some principle, say, the instrumental principle. But if it is constitutive of agency both to follow and to violate a principle, then we can no longer derive an intelligible commitment simply to follow from the nature of agency itself.

The problem is that a constitutivist account of the authority of formal principles must make use of a conception of a perfectly rational agent. It must begin from there, using the ideal to fix the standard constitutive of the activity in question—treating deviations from the relevant standard as we might treat, say, blindness in a human being, an accident, and not as an expression of what rational agency *is*. Yet the imperativalist cannot help herself to that very starting point. Trying to combine imperativalism and constitutivism is like trying to ground one's moral philosophy in a conception of human nature while maintaining the doctrine of original sin.

Here's another perspective on the tension. Recall the way reason, freedom and the possibility of error are linked in the defense of imperativalism canvassed in section five. There we were told that it is by appeal to the concept of freedom that we understand what it is to be under a principle, reminding us that freedom here includes the capacity to resist the guidance of these. But, as I argued above, given the requirement of the possibility of resistance, this freedom amounts to the liberty of indifference. Given this conception of freedom, one exercises one's will in accord with its nature as free or active no matter whether one complies with or violates any principle. It seems that we lose all hope of explaining an agent's subjection to a principle by appeal to the active nature of the capacity whose exercise is action.

So, for this reason too, an imperativalist is prohibited from elucidating the authority of a formal principle, or explaining the commitment to comply with it, by revealing it to be

constitutive of reflective mental activity, or again, internal to agency or acting considered as such. Trying to combine imperativalism and constitutivism is like trying to maintain a voluntarism according to which a fiat of God's unconstrained will creates morality, while also holding that God's will drafts him into the moral order.

What is the relevance of the incompatibility to the truth of imperativalism? How far does the incompatibility take us towards deciding the question of whether part of being under a *formal* principle is being able to err with respect to it? Obviously this is not settled by the incompatibility argument itself. It is important to the assessment of the incompatibility that the chief non-reductive and realist alternative to constitutivism is *platonism*.[108] So it might be of use to say something very brief about that first.

According to platonism, we reach the explanatory bedrock in our understanding of being under a principle with explicitly normative facts themselves. The idea is that these make up a practical normative order that is prior to and independent of the subjectivity of what is under it. On this sort of view, a principle sides with action because a principle is simply a fact like "Everyone ought to do A" or "X should do B in C." For platonists, ultimately there is no explanation to be given of facts of this kind, even if there is much to say about their content. Now, to account for the practical efficacy of the grasp of a principle, platonists have sometimes maintained that unlike other such objects of knowledge, the relevant facts have intrinsic to-be-pursuedness and are such as to elicit movements of the will when grasped; but

[108] A non-reductive realist, in my sense, holds that normative claims are truth-apt (against non-cognitivism), sometimes true (against an error theory) and denies that such truths are reducible to non-normative truths. That is, a non-reductive realist in my sense holds that there are facts about what an agent ought to do and that these are not reducible to non-normative facts. I have not argued that constitutivism is the only non-reductive and realist alternative to platonism, but am following a plausible and well developed tradition of carving up the possibilities, at least so far as formal principles go. Both Korsgaard and Derek Parfit seem to work with this picture of the options, though Parfit leans in the direction of platonism. See his "Reasons and Motivation," *Proceedings of the Aristotelian Society*, suppl. vol. 71 (1997): 99-130, esp. pp. 107-109.

more recently, advocates of platonism have tended to postulate a background disposition to act in accord with judgments about what one ought to do. This depiction is very spare but still portrays something familiar, and, I think, something connected with great difficulties.

Now on the assumption that platonism and constitutivism are exhaustive of the non-reductive and realist alternatives for understanding the authority of such principles, the question of imperativalism poses a stark choice. If one opts for imperativalism, one is forced either to adopt a mysterious platonism about the nature and authority of formal principles, or else to give up on the idea that there really is such a thing as being under one. Whether imperativalism particularized to formal principles is false will largely depend on whether platonism is as hopeless, and constitutivism as promising, as Korsgaard and others have argued, but that must be pursued elsewhere.[109]

---

[109] But is constitutivism so promising? I should mention that many think that it inevitably runs into a terrible difficulty, one recently raised by R. Jay Wallace for Korsgaard ("Normativity, Commitment, and Instrumental Reason," section 1) and also raised by Philip Clark for David Velleman ("Velleman's Autonomism," *Ethics* 111 [2001]: 580-593). To illustrate the problem, let's look at Clark's essay. There he criticizes Velleman's attempt to argue from a conception of the nature of action as an activity whose constitutive aim is autonomy, to a conception of the standards governing action, on the following grounds: "Autonomy, as Velleman conceives it, is a goal that is achieved in every fully intentional action and so cannot serve as the standard of rational assessment for action" ("Velleman's Autonomism," p. 593). Velleman has subsequently expressed concern that there is a real problem here (*The Possibility of Practical Reason*, p. 30). And one might think that the difficulty is so forbidding that one should prefer imperativalism to having to resolve it.

But what exactly is the source of the difficulty? Even with so much interesting work being done in the constitutivist tradition, we still lack, I think, a correct conception of the schema for such a position. In particular, we lack a correct conception of the logical form of the claims describing the essence or nature of agency, the claims in virtue of which we are supposed to understand the force of "oughts" applying to particular agents. The default tendency is to treat the description of essences as universal generalizations in which all members of the relevant kind are said to possess a property. Indeed, the difficulty arises for Velleman just when his description of the nature of action, i.e., his characterization of action's constitutive aim, is taken to have the underlying form of a universal generalization in which anything that is an action has the property of being autonomous.

My thought is that the difficulty is not for the general constitutivist attempt to, as it were, understand norms in terms of natures, but only for a certain conception of the form of descriptions of natures or essences. A promising direction for the constitutivist to go, I think, is to resist the urge to assimilate such descriptions to universal generalizations and instead look towards generics to describe "the what it is" which is to serve to underwrite standards of assessment. For recent work in this direction see Julius Moravcsik, "Essences, Powers, and Generic Propositions," in *Unity, Identity, and Explanation in Aristotle's Metaphysics*, ed. T. Scaltsas, D. Charles and M. L. Gill (Oxford: Clarendon Press, 2000), pp. 229-244, and also see Michael Thompson, "The Representation of Life," in *Virtues and Reasons: Philippa Foot and Moral Theory*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Clarendon Press, Oxford, 1995), pp. 247-296.

I want to conclude this chapter by noting another limitation of our results and advertising a direction for further work. Even if imperativalism is incompatible with constitutivism particularized to formal principles, we are still far from having forged a lever with which to unsettle imperativalism *in general*. Recall that when raising the question of imperativalism, I asked not only whether the liability to error is a condition of being under a formal principle, but also, whether the liability to error is a condition of being under more ordinary, *substantive* principles like those of baseball, etiquette and the laws of Pennsylvania, which do not have the same jurisdiction. But understanding what, if anything, it is to be under one of the latter is not within the reach of constitutivism, as we've considered it. And so the incompatibility argument just given is not directly relevant to the question of imperativalism in these more ordinary cases. I come to this topic with a conviction that our dialectic can be developed even when particularized to substantive principles. But whether an attempt to extract substantive principles from the nature of what is subject to them—an attempt of the kind John Rawls makes for the practice of baseball in "Two Concepts of Rules" and Philippa Foot makes in her recent *Natural Goodness* for the human life-form—is also obstructed by imperativalism must be pursued elsewhere.

# 4. THERE IS NO SUCH THING AS THE INSTRUMENTAL PRINCIPLE

> Here, too, we should guard against projecting categories that can appear only at more developed stages back into this original form. This can only lead, as happened particularly in the case of Kant's philosophy, to a fetishized distortion of the original 'ought', and one which would also have a negative effect on our comprehension of the more developed forms. The fact of the matter, as regards the initial appearance of the 'ought', is simple enough.
>
> Georg Lukacs, *The Ontology of Social Being*, volume 3, Labour

## *4.1 Introduction*

In chapter two, when elucidating the non-contingency of instrumental agency, my focus was entirely on the explanatory dimension of practical reason: not on what an agent ought to do, but on what an agent must be able to do. I then conjectured that the idea of a purely instrumental agent is coherent—that there might be an agent whose practical capacities are entirely limited to doing one thing in order to do another. On behalf of the analytical Kantian, I described a line of argument purporting to show that there is, in fact, only the appearance of a possibility here, as there is only the appearance of a possibility of an atomic agent. The Kantian line of argument, however, attends to the normative dimension of practical reason: we must rule out the possibility of an agent with such limited practical capacities because we cannot make sense of its being under a principle. We see this, the argument goes, when we see that a purely instrumental agent could not go wrong in the relevant way. In chapter three, "Practical Reason and the Possibility of Error," I accordingly adjusted my focus towards the

normative dimension of practical reason and considered how exactly to understand the thought that error must be possible where practical reason is, ultimately rejecting (strong) imperativalism, the conception in play in the Kantian argument. But I refrained from saying anything about particular standards of action. In this chapter, I would like to investigate in some detail the normative dimension of specifically instrumental agency. As I understand it, the task is twofold: (i) to see whether there is an appropriately minimal conception of the norms in play for an agent with such limited practical powers; (ii) to see whether such norms are independently intelligible or self-standing.

Recent discussion of instrumental rationality has centered on questions about the content and the authority of *the instrumental principle—one ought to take the necessary means to one's ends*. In section two, I suggest that this is a mistake. An attempt to provide a minimal theory of practical norms in terms of the instrumental principle confronts a dilemma. On the first horn, the interpretation of the instrumental principle is so spare that it does not articulate a standard that can be genuinely action guiding; that is, it does not articulate a standard that can figure as a content of the practical thinking of an agent. On the second horn, the interpretation is so robust that it is not a plausible specification of a minimal theory of practical norms. On the first horn, the instrumental principle isn't a principle and on the second it isn't internal to instrumental explanation of action.

In section three, I suggest, in particular, that in Korsgaard's hands the instrumental principle enjoins resoluteness in the pursuit of ends. If this is correct, then the argument of "The Normativity of Instrumental Reason" is that to possess resolve in the pursuit of ends, something must be able to engage in non-instrumental reasoning about what to do. This is an interesting claim. By itself, however, it does not pose a challenge to the possibility of a purely

instrumental agent. It is too easy to deny that any genuine agent must be a possible subject of resolution, endurance or strength of mind. Korsgaard needs to consider the conditions of being subject to the norms internal to *in order to*; however, she is, instead, considering something else—namely the conditions of constancy and inconstancy in the pursuit of an end, the conditions of resolution and its failure.

Even so, the failure to demonstrate that instrumentalism is incoherent can be transformed into an argument that *human* agents are not, in fact, purely instrumental agents. A purely instrumental agent cannot exhibit constancy or inconstancy, while human agents can and do. And so, Korsgaard's argument can be transformed into a challenge to analytical Humeanism, which is committed to the view that only purely instrumental agency is possible. Since we do exhibit these, we are not purely instrumental agents, and so something other than that is possible, indeed actual. This is the focus of section four.

What, then, are the norms in play for a purely instrumental agent? Section five draws on the discussion of the explanatory dimension of instrumental reason in chapter two; the proposal is that we look to *action-forms* to provide the material for a properly minimal and self-standing account of norms internal to the capacity to-do-one-thing-in-order-to-do-another. I suggest that this primitive form of practical normativity can be articulated by reflecting on form of the judgments we make about action, in particular, by reflecting, once again, on the aspectual opposition found in *X is □-ing* and *X □-ed*. If this is right, then, as Lukacs says, "The fact of the matter, as regards the initial appearance of the 'ought' is simple enough" and there is no such thing as the instrumental principle.

### *4.2 The content of the instrumental principle*

Recall the second premise of Korsgaard's condition of agency argument: *if something is an agent, then it must be under the instrumental principle*. For Korsgaard, and many others who also endorse this claim, it is taken to be a fixed point in the theory of practical reason: "Most philosophers think it is both uncontroversial and unproblematic that practical reason requires us to take the means to our ends…The interesting question, almost everyone agrees, is whether practical reason requires anything *more* of us than this."[110] I've already mentioned that in trying to elucidate how practical reason requires this much, Korsgaard argues that practical reason must require more. But before saying anything else about that, I want simply to ask what the instrumental principle requires in the first place. What is its content?

Where there might seem to be consensus and a fixed point, there is, in fact, dispute, in particular over the proper specification of the content of the instrumental principle. On the one hand there are those who, like R. Jay Wallace, advocate a spare conception, taking it only to require *consistency* relative to ends one is pursuing. While on the other hand there are those, like Korsgaard, who advocate a more robust conception, taking it to require *constancy* relative to ends one is pursuing for good reason. I'll say a bit about each of these proposals.

*Instrumental consistency*. Wallace is singularly impressed by cases in which great skill is exercised in the pursuit of a goal even when the agent does not regard the goal as worth pursuing. There are several possibilities here: an agent might be pursuing a goal for no reason at all, or simply be immediately determined by appetite to pursue a goal, or an agent might positively think he should not be doing what he is doing. Only the akratic case interests Wallace. Whatever his limitations, the akratic agent may nevertheless exhibit *cleverness*, say, in

[110] Korsgaard, "The Normativity of Instrumental Reason," p. 215.

finding the one place in Provo to buy a bottle he has decided, against his good sense, to drink up right then. Wallace insists that any interpretation of the instrumental principle which neglects this phenomenon should be rejected:

> This kind of practical intelligence seems correctly characterized as a matter of rationality, relative to the akratic agent's ends. Given their determination to achieve the chosen ends, it seems a requirement and not merely an option that such agents should take the means that are necessary to bring their ends about, and those who fail to do this exhibit a characteristic breakdown of rationality. Indeed, they display a lapse precisely in regard to the instrumental principle. This strongly suggests the need for an account of the normative requirement expressed in the instrumental principle that will apply both to cases in which agents take their ends to be well-grounded, and to cases in which they do not do so.[111]

There is presumably more than one way to formulate a requirement involving means and ends so that it is indifferent to the normative status of the agent's pursuit of the end. Wallace's specific suggestion is to treat the instrumental principle as placing a requirement of consistency on combinations of attitudes but not licensing detachment of normative conclusions. Suppose, for example, that you are running the marathon and know that you can do this only if you run the twenty first mile. Instrumental consistency requires that you either run the twenty first mile or give up on running the marathon. It does not, Wallace is at pains to emphasize, require that you run the twenty first mile. And on this interpretation, the instrumental principle has application whether or not the agent's pursuit of the end is well-grounded: it has application even when the agent is running the marathon for no reason, or running it, against his best judgment, to humiliate his clumsy brother-in-law. To sum up, we might put the proposal this way: if X is A-ing and B-ing is necessary for A-ing, then X ought either to give up doing A or do B.[112]

---

[111] Wallace, "Normativity, Commitment and Instrumental Reason," pp. 15-16.

[112] Wallace is following Broome in treating the form of these normative judgments as O(if X wills to do A, then X wills to do B).

*Instrumental constancy*. Before saying anything about the merits of Wallace's proposal, I want to get another in view, one which is *not* indifferent to the normative status of the agent's pursuit of the end. On Korsgaard's account, the instrumental principle has application only if the agent ought to pursue the end in the first place. Here the instrumental principle tells us to adopt those means that are necessary with respect to ends that one is pursuing for a good reason. As a preliminary formulation, we might put the proposal this way: if X ought to do A and B-ing is necessary for A-ing, then X ought to do B.[113]

The feature of the principle of instrumental constancy that is most important to Korsgaard is best seen when we take up the point of view of someone acting on it. Let's consider, as Korsgaard puts it, "when this law must be enforced."

> I have determined upon my end, but now I am reluctant to take the means. The imperative is conditional upon my willing the end, so if I just gave up the end, I could escape its force. Sometimes when we see what the achievement of an end will require of us, we give it up as not worth the bother, or consign it wistfully to the realm of mere wish. But I'm not talking about that kind of case; this end is one I *do* will, one I can't or won't give up. It is only that the means are difficult, or scary, or dull, and I am having trouble screwing myself to the task. That's when I am guided by the imperative—that's when I say to myself—'since you will this end, you must take these means'.[114]

For the moment, I just want to remark that whereas the cognitive excellence Wallace is interested in is cleverness, the excellence that seems to be correlated with the tendency to abide by the principle of instrumental constancy is instead something like what Hume calls strength of mind, what Kant calls perseverance and resolution, and what my Midwestern aunt, in her own special way, calls sticktoittiveness.[115]

---

[113] Though she does not treat deontic operators as taking propositions as objects, Korsgaard's construal of the instrumental principle seems to be the following: If X ought to do A and O(X wills to do A, then X wills to do B), then X ought to do B.

[114] Korsgaard, *The Sources of Normativity*, p. 230.

[115] This conception of instrumental rationality is shared by others: Allen Wood says "We need only to emphasize that instrumental reason is *normative*. It is not merely the effect knowledge has on desire. The threat *to rationality* in

*Consistency or constancy?* Each is an interesting proposal but which is the proper specification of the content of the instrumental principle? Wallace criticizes Korsgaard for construing the instrumental principle too narrowly: it leaves out cases in which an akratic agent is acting in pursuit of a goal. And I have been following Wallace in proceeding as if there is an important topic about which there is a dispute: the content of the instrumental principle. But I now want to suggest that it is doubtful whether there is really anything more than a mere difference in, or at best struggle over, some technical terms. One reason for doubting that there is a genuine dispute here is that there does not seem to be anything prohibiting Korsgaard from endorsing Wallace's principle of *instrumental consistency*. Indeed, at one point she says,

> Even if our ends lack such normativity, so long as they continue to be the ends we have in view, or the ones we effectively want most, we may certainly be inspired by instrumental thoughts to take the means to them.[116]

So, Korsgaard accepts that we can make sense of the intelligent pursuit of a goal in the absence of a conception of the normative standing of the goal. And if we leave room for that, then we also leave room for a distinction between proper and improper exercises of what she elsewhere calls *instrumental intelligence*: one can be better or worse at determining how to get something done.[117]

---

pursuing an end is *not* that our desires will not be informed by the right beliefs about how to achieve the end, but rather that when the time comes to perform the necessary action, our desires may no longer conform to the norms of conduct we established in setting the end. The function of instrumental reason is not to inform desire regarding means but to constrain the will to hold to its rational plan to pursue an end, perhaps even in the face of distracting or contrary desires that tempt the will to abandon the plan" (*Kant's Ethical Thought*, p. 64); Thomas Hill says "The Hypothetical Imperative can function as a general proscription of a kind of duplicity that we associate with neurotic behavior. What it condemns is the irrational failure to follow through on our own morally permissible projects. What it prescribes is that we decide to take the requisite steps to achieve goals that we have already decided to pursue" ("The Hypothetical Imperative," p. 20).

[116] Korsgaard, "The Normativity of Instrumental Reason," p. 252.

[117] Korsgaard, Locke Lectures (Oxford University, 2002), lecture 4. Indeed, it seems that Korsgaard's interpretation of the instrumental principle contains Wallace's interpretation as a part.

In the next breath, however, Korsgaard denies that possessing instrumental intelligence and being evaluable in the light of its standards, amounts to possessing *instrumental rationality* and being subject to *instrumental requirements*. And she insists that for these other, richer phenomena to get hold an agent must have a reason for pursuing the relevant end.

> But no account of a *requirement* of taking the means to our ends can be derived from the mere fact that we possess this kind of intelligence. If there is a principle of practical reason which *requires* us to take the means to our ends, then those ends must be, not merely ones that we happen to have in view, but ones that we have some reason to keep in view.[118]

What's going on?

To see what Korsgaard thinks is missing, it will help to return to the details of Wallace's interpretation. Indeed, a detail of Wallace's own discussion provides a clue. At one point he reflects on "how automatically the core requirement of instrumental rationality tends to be complied with," and suggests that there are necessary limits to the very conceivability of "a willful violation of the core requirement of instrumental reason." He says,

> When people believe that an available means is necessary relative to one of their alleged ends, but fail to adopt that means, we tend to question whether they are really committed to the end after all. The core requirement of instrumental reason thus functions more like a constraint on interpretation than do other principles of practical reason.[119]

The principle of instrumental consistency does not, it seems, leave room for the possibility of "deliberate irrationality." I will say something to explain this later. For the moment, I want to explore the thought that precisely this absence warrants Korsgaard's refusal to find genuine rationality and requirement in the principle of instrumental consistency.

This might seem odd since I already tried to pry apart genuine rationality and potentially exhibiting deliberate irrationality. But there is a further and benign connection between these

[118] Korsgaard, "The Normativity of Instrumental Reason," p. 252.
[119] Wallace, "Normativity, Commitment and Instrumental Reason," p. 26.

that might be made, and might be operative here. Recall that in chapter four I made a distinction between imperativalism and the *practical defect constraint*. According to the latter, an agent is subject to a principle only if there is a describable state of affairs which is someone's being genuinely practically defective with respect to it. It is, I said, a further and significant step to claim that each individual bearer of a capacity for thought of the relevant kind must potentially exhibit such a defect and I militated against taking this further step. But I did not say anything against the weaker thesis. Indeed, I think something like it is true.

The practical defect constraint is really nearer to the logical interpretation than the imperatival interpretation, since it does not have to do with whether some particular agent such as he is might really make a mistake, but rather with whether the content of the principle is such that there can be such a thing as *someone's knowingly* or *self-consciously* violating it. The describability of deliberate irrationality is significant because it seems to make room for a principle to recommend a course of action to an agent, and hence for it to play a role in the content of an agent's thought about what to do. One way to put the point is that, if a principle can't meet the practical defect constraint, then it is impossible for it to be genuinely action-guiding. The refusal to treat a proposal like Wallace's as articulating a rational requirement can rest on this weaker conception of the place of the possibility of error in practical reason. In the rest of this chapter, I will suppose that Korsgaard has only this weaker condition in view.

When one self-consciously fails to take the necessary means to one's ends, one exhibits lack of resolve, inconstancy or weakness of mind. I've already suggested that a purely instrumental agent can't fail in that way. But why should we accept that? Why think that the self-conscious violation of the principle of instrumental consistency is impossible?

### *4.3 The conditions of constancy: a challenge to analytical Humeanism*

When Wallace articulates the principle of instrumental consistency, he has us abstract from the fact that an agent does anything more than have ends and take means to ends; in particular, he has us abstract from an agent's having further reasons for his ends. As I said, he is especially interested in the particular case of an agent who on a particular occasion is doing A akratically, and knows that doing B is necessary for doing A. He is interested in a subject with a capacity to set ultimate ends for a reason, but whose ultimate end on a particular occasion need not have the relevant normative standing. But I want to cast this discussion, at least initially, in terms of a kind of agent who simply can't have further reasons for his ends. Indeed, as I have characterized it, the purely instrumental agent is precisely such a thing: it puts means to ends but does not ultimately *endorse* its pursuits; it does not ultimately have a reason for doing what it is doing. Rather its ends are immediately determined by the strongest desire (in Butler's sense) and instrumental calculation moves from there.

A weak claim: a purely instrumental agent is not a possible subject of resolve. It cannot exhibit constancy in the pursuit of ends. Why not? If something can be characterized as exhibiting resolve in the pursuit of an end, then it must also be possible to characterize it as exhibiting lack of resolve in the pursuit of an end. But what it would be for a purely instrumental agent to exhibit lack of resolve? We cannot, I think, make any sense of this possibility. Wallace is of some help in explaining this. As Wallace sometimes suggests, there is no such thing as a deliberate flouting of the instrumental principle. The explanation of this absence appeals to the belief component of intending to do A—if X intends to do A, then X believes that it is possible that he do A—and the thought that deliberate theoretical

irrationality is not itself possible. For it seems that if intending to do A must involve believing that it is possible that one will succeed, then a subject who knowingly fails to do what is necessary for doing A cannot intend to do A, since he must think that it is possible that he do A.[120]

A stronger claim: for any particular case of action, there is no room for resolve and lack of resolve, if the agent's end is not had for a good reason. That is, having a reason for doing something is a condition of the possibility of the failure of resolve and its contrary. To see this let's consider three cases in which someone intends to do A and then comes to learn that B-ing is necessary for A-ing and then does not come to intend to do B. The difference between the cases in entirely a function of *why* the agent intends to do A. Let's imagine the situation like this. Three agents W, Y, Z stand at the corners of a room, each looking towards the center, each seeing a glass filled with wine resting on a table, and each with the intention of drinking what is in that glass. But there are important differences between them. W intends to drink what's in the glass on a whim and for no particular reason; we can suppose that if we asked "Why do you intend to drink what's in that glass?" he'd say "No reason really, just thought I would." The second case: having had hepatitis C for years, Y, a wine lover, nevertheless intends to drink but against his best judgment; we can suppose that if we asked "Why do you intend to drink that wine?" he'd say "Well, I just love wine, but I know I shouldn't do it." Finally, let's suppose that Z has not done anything but work for many months, and knowing that his friend was right to urge him to let up a bit and have some fun, but still in the grip of anxiety prohibiting him from any pleasure, promised to have a drink. So

[120] In a proper treatment of the topic, it would be important to show that the same point can be made for any of what I called instrumental grounds, i.e., wanting to do A, trying to do A, doing A. When properly understood, each of these carries the implication that it's possible that I succeed in doing A.

here's what we've got: if we ask "Why do you intend to drink what's in that glass?" the responses will be respectively "No particular reason," "I love drinking wine," and "I promised my friend Jones that I would." If we ask "Do you think that you should be drinking what's in that glass?" the responses will be respectively, indifference, no and yes. Now let's suppose that as each steps quite near to the glass, each sees a huge hand surrounding the bottom of the glass. Each then realizes that the only way to drink the wine is to take the goblet from the violent Mr. Smith. Suppose also that none of our agents subsequently adopts the intermediate intention of taking the glass from Mr. Smith. What can we say of them?

The question I want to raise is whether we can make sense of the relevant agents exhibiting lack of resolve: in the background is the thought that if something can't exhibit lack of resolve then it can't exhibit resolve either. So, can one have lack of resolve when doing something on a whim or against one's best judgment? I will suggest that this is not possible and that it is only in the last case that resolution and its failure have room to get hold. In the first and second cases, there is no way to distinguish a change of mind (whether mere alteration or for a reason) and lack of resolve, which I think means that there is no room to talk of a failure of resolution in them. No doubt we can attribute a failure to complete what was started. But no practical irrationality of the relevant sort is involved. Only in the last case is there an all things considered judgment which is contradicted in practical terms. In the last case, since we can attribute a thought like "I ought to drink the wine" even after seeing what is required for doing it, we still have some grounds on which to attribute the end to the agent, in spite of the lack its expression in action. [121]

[121] This is, at best, only a sketch of how the relevant argument might go.

*Wallace's inconsistency?* I've been suggesting that it is impossible to have resolve without being a subject of more than instrumental reasons. Yet sometimes, Wallace suggests that an agent can. Let's examine this.

There is potentially a serious problem for Wallace's more general remarks about the nature of the will as a capacity for self-determination, the exercise of which yields resolutions to do A, but whose operations are not "essentially normative." Indeed, my suggestion about the conditions of constancy raises a problem quite generally for conceptions of the will which treat it simply as a capacity for issuing in resolutions, commitments, or determinations, whether reason based or not.

Wallace wants to maintain both that the will is a capacity for self-determination and also that willings are not essentially normative. And so he must resist the thought that normative endorsement is a condition of the possibility of determination, resolve, perseverance or the lack of this. And indeed, he thinks that there is no difficulty in prying apart determining oneself to do A and having a reason to do A. Here's what he says:

> In a spiteful and nasty mood I might resolve to burn all my roommate's books, without really supposing that what I am doing is best, on the whole; indeed, I might not really believe that it is good or justified in any way at all. In this case, the information supplied about the content of my resolution is enough to undergird the attribution of the resolution to me, as agent; it specifies a goal that I might in principle fail to reach—by, say, neglecting to burn the roommate's cookbooks in the kitchen.[122]

By contrast, Korsgaard insists against a position like Wallace's that "If I am to will an end, to be and to remain committed to it even in the face of desires that would distract and weaknesses that would dissuade me, it looks as if I must have something to *say to myself* about

[122] Wallace, "Normativity, Commitment and Instrumental Reason," pp. 7-8.

why I am doing that."[123] Her thought seems to be that commitment to an end requires that the end be endorsed and pursued for a reason. Wallace replies:

> The reasonable point to which Korsgaard is calling attention is that intentions that diverge from one's normative judgments will not form a reliable basis for long-term planning about the future…But this good point does not rule out the possibility of short-term intentions to act—still less intentions in acting—that diverge from our normative commitments. Granted, an agent who encounters large obstacles on the way to executing an akratic intention of this kind will find it hard to follow through on their intention, and will probably give up. But not necessarily: thinking that I really shouldn't do so, I might nevertheless choose to go out and buy a bottle of rum—and persist, despite discovering that the first shop I drive to is closed, and the second out of stock. In any case, there are plenty of situations in which we don't encounter any unusual additional obstacles on the way to carrying out our short-term akratic intentions.[124]

But has Wallace really hit on the relevant worry? As he understands it, Korsgaard is denying that it is possible to execute a temporally extended plan of action unless one has a reason for doing what one is doing. But this does not seem to capture her thought. The proper reading of Korsgaard's challenge is instead that there is no normatively assessable failure of resolve if the agent doesn't follow through. Maybe it is more accurate to say that her challenge is that without rationally determined ends, an agent cannot follow through because it cannot fail to follow through, where this means that it cannot exhibit inconstancy.

Can Wallace meet this more refined challenge? I don't think so. Again, something he says provides a clue.

> We should equally be able to carry out such activities when the choice to engage in them was not one that we initially endorsed—provided, perhaps, that we do not encounter too much resistance along the way. One would need a powerful philosophical argument to establish that appearances must be deceptive in this area, that we can follow through on our immediate intentions in the face of potential resistance only if we initially viewed them as justified.[125]

---

[123] Korsgaard, "The Normativity of Instrumental Reason," p. 250.
[124] Wallace, "Normativity, Commitment and Instrumental Reason," p. 8.
[125] Wallace, "Normativity, Commitment and Instrumental Reason," p. 9.

Another way of putting Korsgaard's point, I think, is that when the pursuit of an end is not one that an agent endorses, the concept of *too much resistance* is just the concept of whatever amount of resistance would lead to the agent's giving up on the relevant pursuit. I will ask without answering a question about what we are now to make of Wallace's rhetoric of agency as a distinctive capacity for *determination* and *commitment*? That will have to be developed elsewhere. For now I want to get us back on the track of the coherence of instrumentalism.

### *4.4 The conditions of agency: a challenge to instrumentalism*

Perhaps it is already obvious that I think that the real interest of Korsgaard's discussion in "The Normativity of Instrumental Reason" is that it tells us something interesting about the conditions of constancy. Indeed, as it turns out, it can be used to pose a challenge to analytical Humeanism; insofar as we do in fact exhibit resolve and its lack, it would seem that we are more than instrumental agents, and thus that there can be more to agency than instrumental agency.[126]

But this lesson is modest in comparison to the lesson Korsgaard draws from her discussion of the conditions of being subject to the instrumental principle—namely, that the idea of a purely instrumental agent is incoherent. To be entitled to that much stronger thesis, Korsgaard would also need to argue that her substantive version of the instrumental principle must govern the rational activity of putting means to ends. That is, she would have to argue for premise (ii) of the argument sketched in chapter three above: if something is an agent, then it must be under the instrumental principle. As it happens, she just takes this to be common

[126] Many philosophers who call themselves "Humean" might not be bothered by this claim because they might not recognize themselves in the position I have described. In fact, I suspect that this is true of most self-styled "Humeans". The most probable point of resistance is the conception of desire I have been operating with.

ground among the various theories of practical reason. But, I think, the advocate of the coherence of the purely instrumental agent should just deny it.

It may seem naïve to do this. Perhaps one will want to insist that it's undeniable that one ought to take the necessary means to one's ends. And do I really mean to deny that? What I want to suggest is the following: one can admit that there is a sense in which this claim is true, while denying that it expresses an action guiding *principle*, in particular one that supplies the core of a minimal theory of practical norms.

Here's what such a denial might look like. Suppose that I am right that there is no such thing as a purely instrumental agent's exhibiting deliberate instrumental irrationality, i.e., aiming at A-ing, knowing that B-ing is necessary for A-ing and failing to aim at B-ing. It is then very difficult to see how the principle that "One ought to take the necessary means to one's ends" might be able to figure in the thought of what is subject to it, or again that it might be normative for an agent. Indeed, I do not think that it can. For that matter, it is also very difficult to see how particular applications of the principle might figure in the thought of what is subject to it: "when I say to myself—'since you will this end, you must take these means.'"[127] For what can the relevant *ought* and *must* mean?

I am granting that there are such limitations. And basically asking "So what?" The abruptness of the challenge is more tolerable when we see that we can still maintain that "One ought to take the necessary means to one's ends" is true. But here the significance of the truth is only that it is a generalization of particular normative truths of the form "X ought to do B with regard to his aiming at doing A" but not itself a principle of action.

[127] Korsgaard, *The Sources of Normativity*, p. 230.

This amounts to denying that the instrumental principle is genuinely practical. But if one denies this, one has to answer the following question: What then are the practical principles internal to purely instrumental agency? For one might demand that the purely instrumental agent be possibly practical defective in some way or other, with respect to whatever the minimal norms turn out to be. The real challenge to the instrumentalist is to make sense of some practical requirement at all, not to make sense of "You ought to take the means to your ends" as itself expressing a genuine practical requirement.

I think that we can extract from Korsgaard's discussion a workable and abstract specification of conditions of adequacy on any account of practical normativity: (i) a practical principle must be potentially explanatory of action, (ii) it must be able to underwrite explicitly deontic judgments, and finally, (iii) there must be such a thing as being genuinely practically defective with respect to it, i.e., it must meet the *practical defect constraint* described earlier. I want to show that an instrumentalist of the sort I'm imagining can meet these conditions. Perhaps some philosophers will deny that an account of practical normativity must meet such strong conditions; but even a philosopher who denies this should be interested to see that a purely instrumentalist account of agency can meet them.

### *4.5 A "simple enough" conception of instrumental oughts*

Let me try to say where we are. If what I've said is correct, Korsgaard owes us an explanation of why the instrumental principle must have the status she attributes to it, whereas Wallace, or better, an advocate of the purely instrumental agent, still owes us an account of a minimal theory of practical norms. And if a minimal theory can be articulated it will be that much harder for someone to argue for the special status of the so-called instrumental principle. I

want to end by saying a little something, however thin, about what a properly minimal theory of practical norms might look like. This will be, in effect, to argue for the possibility of what Korsgaard denies.

I have been using the word "principle" to mean standard of action—principles side with doing things. When an action is sided with, we can say so explicitly by employing deontic vocabulary as in "X ought to do A," "X may do A in C" or "Everyone must do B in C." Indeed, claims such as these are a starting point of philosophical reflection on practical normativity. What, if anything, could it be for such a claim to be true? The difficulty of answering this question is increased, if one also thinks that such truths are intimately related to the wills of the relevant agents. There has been much controversy over the proper specification of this "intimate relation," but let's just work with the following: a practical deontic claim is true only if, necessarily, the relevant agent is potentially moved to act as recommended.[128] In accepting such a constraint, one displays a sensitivity to another meaning of "principle" as origin, source or root of someone's doing something. If the verbal clue points to something deep, something captured by the constraint, then principles both *side with actions* and *are potentially explanatory of an agent's acting as recommended*—they possess so-called normative force and action-guidingness. So, if an agent is under a principle, its performance of an action is recommended and it has the potential to act accordingly.

We have left ourselves mostly in the dark about how to understand a principle's *siding with* and *potentially explaining* action. Still, as I said earlier, a constitutivist hopes to give a non-reductive and realist account of being under a principle, but, unlike the platonist, without

[128] Anyone familiar with Darwall's taxonomy of "internalisms" and "externalisms" in the theory of practical reason will see in this a near relative of his "existence internalism." It usually formulated with the "unqualified," "unsubscripted" or "flat, flavorless" ought in mind, though I think the point can be extended even to the truth of qualified or subscripted oughts—including those internal to baseball, the laws of Pennsylvania, or even cloak making.

appeal to a normative order completely independent of the subjectivity of what is under it. I would like to end with a few remarks about the abstract shape of any constitutivist proposal and then try to bring things down to earth by considering the particular case of a constitutivist account of being under an *instrumental* principle.[129]

As I understand it, the starting point of any constitutivist account is a description of something that is potentially explanatory of action. It is important to note that the constitutivist assumes—not always explicitly—that *what is potentially explanatory of action exists through the subjects whose action it potentially explains*. It is a much vexed question exactly how to understand this and what *kinds* of thing might play this role, though to illustrate the basic idea it is enough to note that an intention meets the condition, while the platonist's brute normative facts do not. Still, a further qualification is necessary to capture the class of practical explainers of interest and use to the constitutivist. Here's why. Suppose, for the sake of argument, that the attitude I express by saying "I ought to do A" is potentially explanatory of action and meets the above condition. Now an appeal to this sort of explicitly normative attitude could not help us to explain or elucidate what it is for an act to be something that an agent ought to do. If the expression is truth-apt and the attitude expressed doxastic, then appealing to such attitudes simply raises the question what would make that true, clearly not something the attitude itself can help us to understand. If the expression is not truth-apt and the attitude expressed not doxastic, on the other hand, then there is no need to explain what denotic facts come to, for "apparent normative facts will come out, strictly, as no real facts at all."[130] So the constitutivist must leave to the side the attitudes expressed by explicitly deontic

[129] I am taking it that a technical principle is substantive as opposed to formal in the sense described in chapter four.
[130] Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge: Harvard University Press, 1990), p. 23.

claims when specifying the class of potential explainers of interest.[131] To sum up: according to a constitutivist, what is potentially explanatory of action is internal to the subjects whose action is potentially explained (in the sense that the former depends for its existence on the latter) and standards are internal to what is potentially explanatory of action (in the sense that the latter are partly constituted by the former).

Elsewhere I would want to develop in detail a particular application of the strategy to the case of *instrumental* principles which might make these remarks on constitutivism more concrete. For the moment, let me end with a few brief remarks on what that might look like. I begin with an example of Peter Railton's:

> Suppose that I have written you a letter and have spelled 'correspondence' correctly, rather than as the often-seen 'correspondance'. You…remark upon my unexpected success to a colleague…Suppose your friend replies, "No, there simply is no question of why Railton spelled 'correspondence' with an 'e'. *Spelling* is a normative concept—acts of spelling constitutively involve satisfying the norms of spelling. So he couldn't have spelled the word with an 'a'— to have written 'correspondance' wouldn't have counted as a spelling of 'correspondence' at all." Now there certainly is a 'normative sense' of spelling, according to which 'correspondance' cannot count as a spelling of 'correspondence'. In this sense, it is analytic that spelling is correct, and even losers in spelling bees never spell incorrectly. That's why, though it may sound odd to say so, when we ask why or how someone spelled correctly we typically are *not* using the term in this 'normative sense'. As you intended your question to your colleague, my spelling 'correspondence' with an 'e' was either a happy accident or a pleasant surprise, not an analytic truth.[132]

Railton is mocking the position that spelling is a genuinely normative concept. It seems right to me that one can't spell 'correspondence' with an 'a' even though I do not see that maintaining such a thing leaves no room for treating spelling as a genuinely normative concept, in some sense. Here's a way of putting the point: if Jones *spelled* "correspondance" then Jones

[131] I am not saying that a constitutivist is forced to deny that these attitudes are potentially explanatory of action, but only that they are not a possible starting point for a constitutivist account of being under a principle.
[132] Peter Railton, "Normative Force and Normative Freedom," *Ratio* (new series) XII (1999): 320-353.

simply did not spell "correspondence". But if Jones is *spelling* "correspondence" and uses "a" rather than "e", then Jones is misspelling "correspondence". My thought is just that in the case of action the opposition between perfective and imperfective judgments is the primitive expression of instrumental normativity.

In discussions like Railton's, the word "act" is sometimes used to pick out what "X is A-ing" reports and sometimes used to pick out what "X A-ed" reports. If it is restricted to the latter then, with some special exceptions, it is true that "acts of A-ing constitutively involve satisfying the norms of A-ing." But this does not mean that if X is A-ing, X can't fall short of precisely those norms, or that it might be true that X didn't A. This is only the seed of an idea but hopefully it provides a sense of how a simple enough, constitutivist conception of instrumental oughts might be developed.

In any case, let me sum things up: there is no such thing as the instrumental principle in the sense that it is not an action guiding principle that is a part of the minimal theory of practical norms. Even so, when the words "one ought to take the necessary means to one's ends'"are rendered so that they can express an action guiding principle, then Korsgaard is right that a purely instrumental agent could not be under it. As I put it, a purely instrumental agent cannot have resolve in the pursuit of an end. But while this does not show that instrumentalism is incoherent, it does suggest that analytical Humeanism is false; an agent that can have resolve in its pursuits is more than an instrumental agent, and we are such a thing.

# 5. CONCLUSION

I will not attempt to sum up the argument of this dissertation. By way of conclusion, it may be useful to say something about what might be done after questions in analytical philosophy of practical reason are settled. In particular, I want to describe what seems to me a very difficult methodological problem facing anyone trying to describe *kinds* of practical capacity that are not essential to agency in our sense.

In the beginning, I said that we would attend to this question: Which particular practical capacities must something have if it is an agent at all? I hoped to raise a question like the one P. F. Strawson raises when he asks about the modal status of certain conceptual capacities quite generally. But I wanted to ask only about the connection between the capacity to act and certain parts of the soul, as it were. The reason for spending so much time dwelling on the notion of a part of the soul, I said, was that we needed to familiarize ourselves us with that and, more generally, a certain point of view, if we were going to be in a position to appreciate the organizing theme of this dissertation. We can now see that that was somewhat misleading. In the first place, we did not get far enough in the articulation of parts of the soul to have to rely on the relevant sort of division. In a sense, I have simply been discussing the nature and status of a single part, though that way of talking is superfluous given the lack of variety. In the second place, I did not say anything about how to individuate the parts, and in that sense never said anything about what one *is*. Having a clear grasp of this way of carving things up—a principle for individuating the parts and comprehension of the sort of division we are making

when reaching for that metaphor—is indispensable to progress in work on practical reason. I do not plan to fill in that gap now, largely because I have no idea how to do that. Instead, and as a second or third best, I want to end by trying to make the methodological problem, and the difficulty of solving it, more vivid.

*Form and content in practical psychology*. "I had prayed to you for chastity and said 'Give me chastity and continence, but not yet.' For I was afraid that you would answer my prayer at once and cure me too soon of the disease of lust, which I wanted satisfied, not quelled….My inner self was a house divided against itself."[133] The circumstances and expression of Augustine's inner turmoil are no doubt unusual, though I'm confident, reasoning from my own case anyway, that the phenomenon of inner conflict, chaos and contradiction, of being in knots and tangles, all mixed up and pulled in different directions, is rather too usual. We should be careful here. When we are like of a blizzard of one, as Mark Strand says, we are not, or not always, simply a locus of opposed forces. We see this in Augustine. There is a sense in which he is on the side of, or has taken a stand or identified himself with, one of the contending parties to the dispute. Stark, though less beautiful, expressions of this particular species of inner conflict also appear in ordinary life: "I got carried away and swept off," "That wasn't me talking," "I did it despite myself," "I'm at war with myself," "I didn't *really* want to do it," and "I don't want to do what I want to do." And in the everyday, when we try to understand how this might be, when we try to understand what we would have to be like to be the subject of such an experience, we say things like "Human beings are complicated" (overheard just the other day). But what do we have in mind when we say such things as

[133] Augustine, *Confessions*, trans. H. Chadwick (Oxford, Oxford University Press, 1998), ch. 8, secs. 7-8.

this—that human psychology is *structured*, *compound*, *complex*, *composite* or *heterogeneous*? Or that the soul has *parts*?

Such turmoil has been the frequent object of philosophers' attention in large part because acquiring an understanding of it promises an understanding of "the structure of a person's will" more generally.[134] I'd like to turn our attention away from Augustine, and away from ordinary life, and toward a recent implementation of such a strategy. In an early and influential essay, Harry Frankfurt tries to *explain* this discord in terms of the distinction between first and second-order desires, and, ultimately, to see the structure of a person's will as articulated by a hierarchy of desire. As Frankfurt tells it, desires are paradigmatically ascribed by "X wants to A" and the distinction between desires of the first-order and those of the second-order is a function of the referent of "to A." If an action type like "to walk to school," "to shoot some heroin" or "to spend the night with my neighbor's spouse" is substituted for "to A," then the desire is of the first-order. If instead something like "to want to walk to school" or "to want to shoot some heroin" is substituted, then the desire is of the second-order. Frankfurt insists that this distinction, suitably refined, can do the bulk of the relevant explanatory work. The principal refinement is the introduction of the notion of a second-order volition: a person's will is "an *effective* desire—one that moves…a person all the way to action," and a person has a second-order volition "when he wants a certain desire to be his will."[135] Frankfurt then construes our species of inner conflict as conflict between desires of different orders, in particular between first-order desires and second-order volitions, and the privileged standpoint of the person is said to be the second-order volition.

[134] Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* vol. 68 (1971), p. 12.
[135] Frankfurt, "Freedom of the Will and the Concept of a Person," pp. 14 and 16.

In a well known response, Gary Watson claims that Frankfurt's conceptual resources are plainly insufficient to meet his explanatory aims: "Since second-order volitions are themselves simply desires, to add them to the context of conflict is just to increase the number of contenders; it is not to give a special place to any of those in contention."[136] And so, Watson reasons, a hierarchy of desire cannot account for the special *authority* that attaches to one of the elements of the conflict; it cannot account for the sense in which an agent "takes sides" or "identifies" with some desires and not others. Watson has an argument but I'm more interested in the spirit of the objection.

The distinction between first and second-order desires, indeed between any nodes on the hierarchy of desire, as well as the distinction between second-order desires which are volitions and those which are not, are classifications of practical explainers according to their *content*. Consider another. If we classify "to want to buy a house in Celebration," "to want to educate my children in Celebration" and "to want to make my mark in Celebration" as Celebration-desires and classify "to want to play for the Steelers in Pittsburgh," "to want to assassinate Frick in Pittsburgh" and "to want to drink Iron City in Pittsburgh" as Pittsburgh-desires, we also engage in classification of elements of practical psychology according to their content. It is obvious that the distinction between Celebration-desires and Pittsburgh-desires offers no prospect of a philosophical payoff. Frankfurt's contents are far more exquisite and for that reason, I guess, have given many great hope that profit can be squeezed from them. Still, it's worth bearing in mind that the mode of classification is the same in each case.

We might characterize the spirit of Watson's objection like this: Frankfurt fails to locate structure, complexity or heterogeneity in the "soul" because he hasn't introduced distinctions

[136] Gary Watson, "Free Agency," *Journal of Philosophy* 72, p. 108.

among *forms* (*kinds*, *categories*, *species*, *sorts*, or *types*) of practical explainer but only made a distinction among the contents of some one form. Watson says, "Frankfurt's position resembles the platonic conception in its focus upon the structure of the 'soul'. But the two views draw their divisions differently. Whereas Frankfurt divides the soul into higher and lower orders of desire, the distinction for Plato…is among independent sources of motivation."[137] Frankfurt's conception of how to "draw the divisions," it seems, is really the deep source of the difficulty. In opposition, Watson insists that the distinction between "valuing and desiring" at the foundation of his own account of inner conflict and the structure of the soul "is not, it is crucial to see, a distinction among desires or wants according to their content."[138]

The object of my immediate interest is not inner conflict, but what is far more abstract—the distinction between form and content that Watson reaches for when criticizing Frankfurt's attempt to explain such conflict. Despite the importance of the distinction to contemporary work on practical reason, surprisingly little is said about what it is to classify elements of practical psychology according to their *form* as opposed to their *content*, and, I suppose, about what it is for someone's will to have structure or complexity. Indeed, too little is said about the mere fact that contemporary philosophers of practical reason must and do, at least implicitly, operate with such a way of carving things up—what we might regard as a way of carving practical reality at its joints, or as A. O. Lovejoy might have said registering a link in the chain of practical being.[139]

---

[137] Gary Watson, "Free Agency," p. 109.
[138] Gary Watson, "Free Agency," p. 102.
[139] The distinction between the form and content of an element of practical psychology has made countless appearances with varying degrees of explicitness in the tradition. It is, I think, the necessary background against which we can grasp, for example, Kant's distinction between autonomous and heteronomous principles of the will or Hume's insistence that passion and never reason moves the will. Indeed, the terrain of the contemporary literature on

*Individuating parts of the soul: Plato's problem*. So, we have not done anything to describe the relevant sense of *kind*, other than to say that when we use it, we purport to be carving practical reality at its joints. We have not addressed the problem of articulating a principle for individuating parts of the soul. I'd like to highlight this foundational problem and emphasize the difficulty of solving it by describing a familiar difficulty for Plato's tripartitioning.

Recall that in Book IV of the *Republic*, when Plato argues that there are, in fact, distinct parts of the human soul, he does so through the application of the *principle of opposites*: "The same thing will not be willing to do or undergo opposites in the same part of itself, in relation to the same thing, at the same time."[140] So, he says, "if we ever find this happening in the soul, we'll know that we aren't dealing with one thing but many."[141] The well-known problem for Plato is this: by applying the principle of opposites we would seem to be able to generate indefinitely many parts of the soul, and not simply three. When the fountain and trough are fifteen yards apart don't my thirst and hunger pull me in opposite directions? If so, why not recognize further parts, say, a thirst-soul and hunger-soul? Or suppose that I'm indignant because Shannon has ignored me and pushed to get revenge, but am also ashamed of my anger and pulled to turn away from revenge. Why not recognize further divisions of the soul here as well? In these cases we come upon different and, in some sense, conflicting practical pushes and pulls. Doesn't the principle of opposites force us to make further discriminations among parts of the soul or among kinds of practical capacity? It is safe to say that Plato would not incorporate this objection by treating it as an unnoticed consequence of his position: the main argument of the *Republic* for the rationality of justice turns on there being exactly three parts.

---

practical reason could not be mapped without implicitly deploying some such privileged scheme of classification. Of course, this is not to say that there is any agreement about how to individuate the relevant kinds or about what particular kinds there are, but only that some such concept figures in the neighborhood.

[140] *Republic* 436b.

[141] Ibid.

In any case, Plato must reject it because no adequate account of practical kinds permits their infinite and arbitrary proliferation. Any adequate account must address the *problem of proliferation*.

Adequately meeting it requires that Plato distinguish between mere opposition and deep opposition, between opposition which does not and opposition which does force us to recognize distinct parts of the soul. Specifying the latter is not something I will attempt on his behalf. Instead, I simply want to focus on a constraint which Plato employs in the specification of this concept: desires of the same kind cannot be in opposition in the right way. If this is correct as a matter of interpretation, then the argument for the partitioning depends on already having at hand something like a distinction between *kinds* of desire. And this does seem to be how Plato proceeds. For example, in the course of distinguishing the appetitive and rational parts he takes it as mostly obvious that there is something like the *class* of desires which are appetites and that this is to be distinguished from rational desires, wishings and willings;[142] in the course of marking off the spirited part from each of these, Plato appeals to a further, spirited kind of desire. We can just assume that Plato's appeal to kinds of desire allows him to defuse the problem I've raised about the unbounded fertility of the principle of opposition. The move itself deserves our attention.

The principle of opposites does not purport to be an account of what a part of the soul is, but merely a specification of a sign of the presence of distinct parts. The real work of showing that there are parts of the soul is being done by the distinction among kinds of desire: we individuate parts of the soul according to kinds of desire, one to one. This is a happy result, I think. It suggests that Plato need not have argued for the partitioning on the grounds that it is

[142] *Republic* 437c-d.

required to explain moments of conflict or discord. Other things equal, it is better not to rest one's account of the structure of the soul on the grounds that only it can explain the peculiar features of the defective instances.[143]

We were asking about how to individuate kinds of practical capacity, and by following Plato we've settled on individuating them according to the kind of practical explainer involved, or again, according to kinds of desire. Of course, to be told that we individuate parts of the soul by appeal to kinds of desire is not to make any progress towards the acquisition of a sense of the type of classification Plato and those in his wake operate with. Really, we are left with a very similar problem to the one with which we began. What is the relevant sense of *kind* here? What is a *kind* of desire? What is a *kind* of explainer of action? How do we individuate *these*?[144]

I will not try to give Plato's answer but will instead turn to Michael Bratman's very recent attempt to address this sort of question, and more generally questions about philosophical method. Bratman has said that his own work in practical psychology is informed by the *method of creature construction* developed in the mid-seventies by H. P. Grice. Bratman, quoting Grice, presents the basic idea as follows:

> 'To construct (in imagination, of course) according to certain principles of construction a type of creature, or rather a sequence of types of creature, to serve as a model (or models) for actual creatures.' Grice calls his creatures 'pirots' and writes: 'the general idea is to develop sequentially the psychological theory for different brands of pirot, and to compare what one

[143] See *Republic* 443-5 for the suggestion that the psychic harmony of the just person does not leave room for the sort of conflict which serves as the starting point of Plato's own argument for the partitioning: "He puts himself in order, is his own friend, and harmonizes the three parts of himself like three limiting notes in a musical scale—high, low, and middle. He binds together those parts and any others there may be in between, and from having been many things he becomes entirely one, moderate and harmonious" (*Republic* 443d).

[144] In "Plato's Theory of Human Motivation" (*History of Philosophy Quarterly* 1 (1984): pp. 3-21), John Cooper says that a part of the soul is a distinct source of desire and that we are to distinguish parts of the soul by appeal to the relevant *kinds* of desire. Obviously the real problem is just getting pushed back one square.

> thus generates with the psychological concepts we apply to suitably related actual creatures.'

In his own voice, Bratman continues, "my aim is to see a number of different models of agency as reasonable stages in a sequence of creature constructions."[145] Here Bratman is employing the method in an attempt to better understand "valuing and its relation to the will," but I am interested in his conception of the method—in particular, in its aptitude for shedding light on sort of classification we are engaging in when locating types, brands or kinds of practical item. In the passage above, Bratman is discussing kinds of agent, though there is a difference in kind of agent just when there is a difference in the kinds of practical capacity possessed.

Let's get acquainted with the creatures and ascend a portion of Bratman's *scala naturae*. Creature 1, the ground level, has only the capacity to act on the basis of *mere desires*. Working our way up, we encounter Creature 2 which is in addition able to subject its wants to reflective scrutiny and thereby to alter what it wants; that is, it is a subject of what Bratman calls *considered desires*. Still, Bratman tells us that each of these Creatures "is moved to act by its strongest desire."[146] In contrast, Creature 3 has the ability to deliberate about what ultimately to do in the face of conflicting wants. However, Bratman tells us that the deliberative weights assigned to each desire "correspond to the motivational strength of the associated, considered desire. The outcome of such deliberation will match the outcome of the causal, motivational processes envisaged in our description of Creature 2."[147]

---

[145] Michael Bratman, "Valuing and the Will," *Philosophical Perspectives: Action and Freedom*, vol. 14 (2000), p. 250.
[146] Michael Bratman, "Valuing and the Will," p. 251.
[147] Michael Bratman, "Valuing and the Will," p. 252. I have a difficult time grasping this step: it seems to offend a Gricean constrain on creature construction, namely that "no psychological concept can be instantiated by a priot without the supposition of behaviour which manifests it. An explanatory concept has no hold if there is nothing for it to explain. This is why 'inner states must have outward manifestations'" (H. P. Grice, "Method in Philosophical Psychology (From the Banal to the Bizarre)" [Presidential Address] in *Proceedings and Addresses of the American*

Another rung up is Creature 4, which in addition to acting on wants, considered or otherwise, acts on *plans* and *policies*. But what are those? If we look to other work of Bratman's we see that a plan is always a plan *to do* A, where A is a relatively specific course of action extended over time. A policy is also oriented towards action, it is a policy *to do*, but to do *generally*.[148] So, we seem to move from Creature 1 (merely a subject of wants) to Creature 4 (also a subject of plans and policies) when what the creature is after is no longer, say scratching its arm now (want), but rather, swimming across the lake (plan), or swimming across the lake every morning (policy).

One has to wonder whether there is any real progress here, whether Bratman is carving at the joints. I mean for all that Bratman has said, we can insist that there are other Creatures between 1 and 4. How has he overlooked the possibility of tasking-agents? The tasking-agent is more than Creature 1 but less than a planning agent. What it aims at doing does not take much time at all, say, swimming across the kiddie-pool, or copying out the first line of *The Decline and Fall of the Roman Empire*, as opposed to the objects of plans, say, swimming across the lake, or copying out all of Gibbon's work. These creatures perform mere tasks but do not execute plans in Bratman's sense. Moreover, how has Bratman overlooked the possibility of trendy-agents? The trendy-agent is more than Creature 1 but less than a policy agent. A creature that acts on policies, acts on something with a general content, doing A in C, whereas the more limited trendy-agent only acts on fads, it does A* in C* until next season.

I am aware that these suggestions are absurd. But why are they? Without a principled answer there will be no more to Bratman's method of creature construction than the

---

*Philosophical Association* [1974-1975]: 23-53, p. 39). In any case, I do not want to get hung up on any of the details, so let's proceed.

[148] Michael Bratman, "Reflection, Planning, and Temporally Extended Agency," *The Philosophical Review* 109, p. 7.

employment of an entertaining rhetorical device in which putative elements of our conceptual scheme, as well as their relations of dependence, are rendered vivid by imagining Creatures, as Plato does when asking us to animate the parts of the soul.[149] But that is not a method for determining when such distinctions are to be made, much less an account of the kind of classification one is engaging in. Have we missed something in Bratman's account of philosophical method? Something that will help us to understand what we are doing when we "draw the divisions," to use Watson's expression?

Maybe we have. Bratman says, "at each stage in the sequence I will try to identify an issue or problem that suggests some sort of modest addition to or extension of the earlier design."[150] He seems to suggest that we are only justified in introducing a new creature when we have described "an issue or problem" afflicting the less advanced creature and which the subsequently introduced practical capacity is suited to solve. This seems to be a specification of, at least, a necessary condition on taking a step in the construction of creatures and a partial specification of the relevant difference in kind. Even if we were willing to grant this, I'm not at all sure that the method of creature construction would be able to address the *problem of proliferation*. Just as we were able to overpopulate the taxonomy of creatures by using different concepts, e.g., task and fad, or by taking shorter steps along the dimensions that Bratman regards as significant, e.g., time and generality, we can introduce issues and problems in smaller increments as well. We would thereby generate a problem for Bratman similar to the one that arises for Plato.

One reason for raising the question of the form-content distinction as it operates in work on practical reason is that it needs to be resolved in order to proceed further in analytical

[149] *Republic* 588.
[150] Michael Bratman, "Valuing and the Will," p. 250.

philosophy of practical reason. Moreover, even if one thinks that the only results available in analytical philosophy of practical reason are the ones I have already gotten, and nevertheless thinks that we human beings have practical capacities that outrun anything captured by instrumentalism, as I suggested in the last chapter, then one needs something to say about what that something more is. In what way is there a *structural* difference in the character of our abilities to employ thought in the service of action? Nevertheless, I do not know how to say what the form-content distinction comes to, especially when it is unmoored from the traditional apparatus of transcendental argument, as it would have to be when characterizing practical capacities possessed by us, but not necessarily possessed by any agent at all.[151] Here I only express the hope that there is some value in dwelling on the problem.

When I was despairing out loud about these difficulties, Michael Thompson grabbed the last volume of Marx's *Capital*, turned to the end of the last chapter—the chapter in which Marx confronts the question of the sort of division we are making when dividing society into classes—and read something simultaneously demoralizing and uplifting, and certainly amusing. I'll end with that:

> The owners merely of labour-power, owners of capital, and land-owners, whose respective sources of income are wages, profit and ground-rent, in other words, wage-labourers, capitalists and land-owners, constitute then the three big classes of modern society based upon the capitalist mode of production…
>
> The first question to be answered is this: What constitutes a class?—and the reply to this follows naturally from the reply to another question, namely: What makes wage-labourers, capitalists and landlords constitute the three great social classes?
>
> At first glance—the identity of revenues and sources of revenue. There are three great social groups whose members, the individuals forming them, live on

[151] Sebastian Rödl argues in a fascinating series of lectures on related topics that it is impossible to have an account of what a kind of rational capacity is in the absence of a more encompassing story about why possession of such a capacity is necessary in anything that can think. I have learned very much from these lectures, even if I haven't taken this point to heart.

wages, profit and ground-rent respectively, on the realization of their labour-power, their capital, and their landed property.

However, from this standpoint, physicians and officials, e.g., would also constitute two classes, for they belong to two distinct social groups, the members of each of these groups receiving their revenue from one and the same source. The same would also be true of the infinite fragmentation of interest and rank into which the division of social labour splits labourers as well as capitalists and landlords—the latter, e.g., into owners of vineyards, farm owners, owners of forests, mine owners and owners of fisheries.

[Here the manuscript breaks off].[152]

[152] Karl Marx, *Capital*, *volume three* (New York: Random House, 1977), chapter 52, Classes.

## BIBLIOGRAPHY


Allison, Henry. 1990. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Anscombe, G. E. M.. "Before and After", *Philosophical Review* 73 (1964): 3-24.
--------- 1989. "Von Wright on Practical Inference" in *The Philosophy of Georg Henrik von Wright*, ed. Schilpp. La Salle: Open Court.
--------- 2000. *Intention*. Harvard University Press.

Anselm, St. 1967. *Truth, Freedom, and Evil: Three Philosophical Dialogues*, ed. and trans. J. Hopkins and H. Richardson. New York: Harper and Row.

Aquinas, St Thomas. 1920. *Summa Theologiae*, trans. Fathers of the English Dominican Province. London: Burns Oates.

Augustine. 1998. *The Confessions*, trans. H. Chadwick. Oxford: Oxford University Press.

Baier, Annette. 1971. "The Search for Basic Actions", *American Philosophical Quarterly* 8:161-170.

Bennett, Jonathan. 1989. *Rationality: An Essay Toward Analysis*. Indianapolis: Hackett Publishing.

Blackburn, Simon. 1995. "Practical Tortoise Raising", *Mind* 104: 695-711.
--------- 1998. *Ruling Passions*. New York: Oxford University Press.

Braithwaite, Richard. 1966. "Causal and Teleological Explanation", in *Purpose in Nature*, ed. John Canfield. Englewood Cliffs: Prentice-Hall.

Bratman, Michael. 2000. "Valuing and the Will", *Philosophical Perspectives: Action and Freedom*, vol. 14: 249-265.
--------- 2000. "Reflection, Planning, and Temporally Extended Agency", *The Philosophical Review* 109: 35-61.

Brandom, Robert. 1994. *Making It Explicit*. Cambridge: Harvard University Press.

Broome, John. 1997. "Reason and Motivation", *Proceedings of the Aristotelian Society*, suppl. vol. 71: 131-146.
--------- 1999. "Normative Requirements", *Ratio* [new series] 12: 398-419.

--------- 2004. "Practical Reasoning", in J. Bermudez, ed. *Reason and Nature: Essays in the Theory of Rationality*. Oxford: Oxford University Press.

Butler, Joseph. 1983. *Fifteen Sermons Preached at the Rolls Chapel*, reprinted in *Five Sermons Preached at the Rolls Chapel and A Dissertation upon the Nature of Virtue*, ed. Stephen Darwall. Indianapolis: Hackett Publishing.

Churchland, Paul. 1970. "The Logical Character of Action-Explanations", *The Philosophical Review* 70: 214-236.

Clark, Philip. 2001. "Velleman's Autonomism", *Ethics* 111: 580-593.

Cooper, John. 1984. "Plato's Theory of Human Motivation", *History of Philosophy Quarterly* 1: 3-21.

Danto, Arthur. 1968. "Basic Actions", in *Philosophy of Action*, ed. Alan White. Oxford: Oxford University Press.

Darwall, Stephen. 1983. *Impartial Reason*. Ithaca: Cornell University Press.
--------- 1985. "Kantian Practical Reason Defended", *Ethics* 96: 89-99.
--------- 1992. "Internalism and Agency", *Philosophical Perspectives* 6: 155-174.
--------- 1995. *The British Moralists and the Internal 'Ought': 1640-1740*. Cambridge: Cambridge University Press.
--------- unpublished. "Two Dogmas of Empiricism in Ethics".

Davidson, Donald. 1980. *Essays on Actions and Events*. New York: Oxford University Press.

Dreier, James. 1997. "Humean Doubts about the Practical Justification of Morality", in G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*. New York: Oxford University Press.

Engstrom, Stephen. Unpublished. *Contradictions in the Will: An Essay on Kant's Formula of Universal Law*.

Foot, Philippa. 2001. *Natural Goodness*. Oxford: Oxford University Press.

Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person", *Journal of Philosophy* vol. 68, no. 1: 5-20.

Galton, Antony. 1984. *The Logic of Aspect*. Oxford: Clarendon Press.
--------- 1987. "The Logic of Occurrence" in *Temporal Logics and Their Appications*. London: HBJ Publishers.

Gauthier, David. 1985. "The Unity of Reason: A Subversive Reinterpretation of Kant", *Ethics* 96: 74-88.

--------- 1986. *Morals By Agreement*. New York: Oxford University Press.

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

Grice, H. P.. 1975. "Method in Philosophical Psychology (From the Banal to the Bizarre)" [Presidential Address] in *Proceedings and Addresses of the American Philosophical Association*: 23-53.

Hegel, G. W. F.. 1991. *Elements of the Philosophy of Right*, ed. Allen Wood, trans. H. B. Nisbett. Cambridge: Cambridge University Press.

Hill, Thomas. 1992. "The Hypothetical Imperative", in *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca: Cornell University Press.

Hubin, Donald. 2001. "The Groundless Normativity of Instrumental Rationality", *Journal of Philosophy* 98: 445-468.

Hornsby, Jennifer. 1980. *Actions*. London: RPK.

Hume, David. 1975. *An Enquiry Concerning the Principles of Morals*. ed. L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press, 1975.
--------- 1978. *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and rev. P. H. Nidditch. Oxford: Clarendon Press.

Kant, Immanuel. 1996. *Groundwork of the Metaphysics of Morals*, in *Practical Philosophy*. Ed. and trans. Mary Gregor. Cambridge: Cambridge University Press.
--------- 1996. *Critique of Practical Reason*, in *Practical Philosophy*.
--------- 1996. *The Metaphysics of Morals*, in *Practical Philosophy*.

Korsgaard, Christine. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
--------- 1996. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
--------- 1997. "The Normativity of Instrumental Reason", *Ethics and Practical Reason*.
--------- 1998. "Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind," *Ethics* 109: 49-66.
--------- 1999. "Self-Constitution in the Ethics of Plato and Kant", *The Journal of Ethics* 3: 1-29.
--------- unpublished. Locke Lectures, Oxford University 2002.

Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.

Marx, Karl. 1977. *Capital*, *volume three*. New York: Random House.

McDowell, John. 1998. "Might There Be External Reasons?" in his *Mind, Value, and Reality*. Cambridge: Harvard University Press.

Moravcsik, Julius. 2000. "Essences, Powers, and Generic Propositions", in *Unity, Identity, and Explanation in Aristotle's Metaphysics*, ed. T. Scaltsas, D. Charles and M. L. Gill. Oxford: Clarendon Press.

Müller, Anselm. 1979. "How Theoretical is Practical Reason?" in *Intention and Intentionality: Essays in Hornor of G. E. M. Anscombe*, eds. C. Diamond and J. Teichman. Ithaca: Cornell University Press.

Nagel, Thomas. 1970. *The Possibility of Altruism*. Princeton: Princeton University Press.

Parfit, Derek. 1997. "Reasons and Motivation", *Proceedings of the Aristotelian Society* suppl. vol. 71: 99-130.

Parsons, Terence. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge: MIT Press.

Penner, Terry. 1970. "Verbs and the Identity of Actions: A Philosophical Exercise in the Interpretation of Aristotle", in *Ryle: A Collection of Critical Essays*, eds. O. Wood and G. Pitcher. New York: Anchor Books.

Plato. 1992. *Republic*, trans. G. M. A. Grube and C. D. C. Reeve. Indianapolis: Hackett Publishing.

--------- 1995. *Phaedrus*, trans. A. Nehamas and P. Woodruff. Indianapolis: Hackett Publishing.

Railton, Peter. 1997. "On the Hypothetical and Non-Hypothetical in Reasoning about Belief and Action", *Ethics and Practical Reason*.

--------- 1999. "Normative Force and Normative Freedom," *Ratio* (new series) XII: 320-353.

Rawls, John. 2000. *Lectures on the History of Ethics*. Cambridge: Harvard University Press.

Rousseau, Jean-Jacques. 1978. *On The Social Contract*, ed. Roger D. Masters, trans. Judith R. Masters. New York: St. Martin's Press.

Ryle, Gilbert. 1949. *The Concept of Mind*. New York: Barnes and Noble Press.

Scanlon, T. M.. unpublished. "Structural Irrationality".

Searle, John. 2001. *Rationality in Action*. Cambridge, MA.: MIT Press.

Smith, Michael. 1987. "The Humean Theory of Motivation," *Mind* 96: 36-61.

Sobel, David. 2001. "Subjective Accounts of Reasons for Action", *Ethics* 111: 461-492.

Stout, Rowland. 1996. *Things That Happen Because They Should*. Oxford: Oxford University Press.

Szabo, Zoltan. 2004. "On the Progressive and the Perfective", *Nous* 38: 29-59.

Tenenbaum, Sergio. 2003. "Speculative Mistakes and Ordinary Temptations: Kant on Instrumentalist Conceptions of Practical Reason", *History of Philosophy Quarterly* 20: 203-223.

Thompson, Michael. Forthcoming. *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, MA.: Harvard University Press.

Velleman, David. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Vogler, Candace. 2002. *Reasonably Vicious*. Cambridge, MA.: Harvard University Press.

Wallace, R. Jay. 2001. "Normativity, Commitment and Instrumental Reason", *Philosophers' Imprint*, vol. 1, no. 3: http://www.philosophersimprint.org/001003.

Watson, Gary. 1975. "Free Agency", *Journal of Philosophy* 72: 205-220.

Williams, Bernard. 2001. "Some Further Notes on Internal and External Reasons", in *Practical Reasoning*, ed. Elijah Millgram. Cambridge: MIT Press.

Wolf, Susan. 1990. *Freedom within Reason*. New York: Oxford University Press.

Wood, Allen. 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.
--------- 1984. "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy*, ed. A. Wood. Ithaca: Cornell University Press.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*, trans. G. E. M. Anscombe. New York: Macmillan Publishing.