

Self-Constitution

Agency, Identity, and Integrity

Christine M. Korsgaard

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© Christine M. Korsgaard 2009

The moral rights of the author have been asserted
Database right Oxford University Press (maker)

First published 2009

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
Antony Rowe Ltd, Chippenham

ISBN 978-0-19-955279-5 (Hbk)
978-0-19-955280-1 (Pbk)

10 9 8 7 6 5 4 3 2 1

The Metaphysics of Normativity

2.1 Constitutive Standards

2.1.1

In Chapter 1, I proposed that the principles of practical reason serve to unify and constitute us as agents, and that is why they are normative. Behind this thesis lies a more general account of normativity that I believe to be common to the philosophies of the three thinkers who are the heroes of this book: Plato, Aristotle, and Kant. According to this account, normative principles are in general principles of the unification of manifolds, multiplicities, or, in Aristotle's wonderful phrase, *mere heaps*, into objects of particular kinds (M 8.6 1045a10).

The view finds its clearest expression in the central books of Aristotle's *Metaphysics*, so that is the place to start.¹ According to Aristotle, what makes an object the kind of object that it is—what gives it its identity—is what it does, its *ergon*: its purpose, function, or characteristic activity. This is clearest in the case of artifacts, which are obviously functionally defined. An artifact has both a form and a matter. The matter is the material, the stuff or the parts, from which the artifact is made. The form of the artifact is its functional arrangement or teleological organization. That is, it is the arrangement of the matter or of the parts which enables the object to serve its function, to do whatever it does that makes it the kind of thing that it is. Say for instance that the function of a house is to serve as a habitable shelter, and that its parts are walls, roof, chimney, insulation, and so on. Then the form of the house is that arrangement of those parts that enables it to serve as a habitable shelter—or rather, to be more precise—it is the *way* the arrangement of those parts enables it to serve as a habitable shelter. The walls are joined at the corners, the insulation goes into the walls, the roof is placed on the top, and so on, so that the weather is kept out, and a comfortable environment is created within. That is the form of a house.

¹ The views that follow are primarily from *Metaphysics* 7–9.

On this view, to be an object, to be unified, and to be teleologically organized, are one and the same thing. Teleological organization is what unifies what would otherwise be a *mere heap* of matter into a particular object of a particular kind. Teleological organization, according to Aristotle, is also the object of knowledge. To know an object, that is, to *understand* it, is to see not only what it does and what it is made of, but also *how* the arrangement of the parts enables it to do whatever it does. After all, anybody knows that a house is a shelter, and anybody knows that its parts are walls and roofs and chimneys and things, and even roughly where they go. What distinguishes the architect is his knowledge of *how* the arrangement of those parts enables the house to serve the purpose of sheltering. And this means that according to Aristotle the form of a thing governs both theory and practice. To understand houses is to have their form in your mind, and to build one is to be guided by that form.

At the same time, it is the teleological organization or form of the object that supports normative judgments about it. A house with cracks in the walls is less good at keeping the weather out, less good at sheltering, and therefore is a less good house. The ancient metaphysical thesis of the identification of the real with the good follows immediately from this conception, for this kind of badness eventually shades off into *literal* disintegration. A house with enough cracks in the walls will crumble, and cease to be a house altogether: it will disintegrate back into a *mere heap* of boards and plaster and bricks.

2.1.2

It is essential here to observe the distinction between being a good or bad *house* in the strict sense and being a house that happens to be a good or bad *thing* for some external reason. The large mansion which blocks the whole neighborhood's view of the lake may be a *bad thing* for the neighborhood, but it is not therefore a *bad house*. The normative standards to which a thing's teleological organization gives rise are what I will call "constitutive standards," standards that apply to a thing simply in virtue of its being the kind of thing that it is.

An especially important instance of the constitutive standard is what I will call the constitutive *principle*, a constitutive standard applying to an activity. In these cases what we say is that if you are not guided by the principle, you are not performing the activity at all. In the case of essentially goal-directed activities, constitutive principles arise from the constitutive standards of the goals to which they are directed. A house-builder is, as such, trying to build an edifice that will keep the rain and weather out. But all activities—as opposed

to mere sequences of events or processes—are, by their nature, directed, self-guided, by those who engage in them, even if they are not directed or guided with reference to external goals. And the principles that describe the way in which an agent engaged in an activity directs or guides himself are the constitutive principles for that activity. So it is a constitutive principle of walking that you put one foot in front of the other, and a constitutive principle of skipping that you do this with a hop or a bounce. Or, to use a controversial example, it is a constitutive principle of thinking that you *swerve* when you see a contradiction looming ahead in your path. And in all these cases, we can say that unless you are guided by the principle in question, you are not performing that activity at all.

2.1.3

The idea of a constitutive standard is an important one, for constitutive standards meet skeptical challenges to their authority with ease. Why shouldn't you build a house that blocks the whole neighborhood's view of the lake? Perhaps because it will displease the neighbors. Now *there* is a consideration that you may simply set aside, if you are selfish or tough enough to brave your neighbors' displeasure. But because it does not make sense to ask why a house should serve as a shelter, it also does not make sense to ask why the corners should be sealed and the roof should be waterproof and tight. I mean, of course you can ask these questions in a technical voice, you can ask how sealed corners and waterproofed roofs serve the function of sheltering. But once you've answered the technical questions, there is no further room for doubting that the constitutive standard has normative force. For if you fall too far short of the constitutive standard, what you produce will simply not be a house. In effect this means that even the most venal and shoddy builder must try to build a good house, for the simple reason that there is no other way to try to build a house. Building a good house and building a house are not different activities: for both are activities guided by the teleological norms implicit in the idea of a house. Obviously, it doesn't follow that every house is a good house, although there is a puzzle about why not. It does, however, follow that building bad houses is not a different activity from building good ones. *It is the same activity, badly done.*

2.1.4

Let's consider that puzzle. If building bad houses is the same activity as building good ones, why are there any bad houses? In the case at hand, we

have an object, a house, characterized by certain constitutive standards. It is in terms of those standards that we understand the activity of producing a house. The producer of the house looks to the normative standards that are constitutive of houses—in Aristotle’s terms, to its form—and tries to realize that form in appropriate matter—in building materials. Since building is a goal-directed activity, that is what the activity of building essentially is. The description of the form of a house could be read as a sort of recipe, or a set of instructions, for building a house: join the walls at the corners, put the insulation in the walls, put the roof on the top . . . So trying to produce a house is not a different activity from trying to produce a good house. One is trying to build a good house if one is building a house at all. How then is the shoddy builder even possible?

The problem is a general one, not limited to productive activities. Here are a couple more examples. In the *Groundwork*, Kant argues that hypothetical imperatives, the principles of instrumental reason, are analytic, because “whoever wills an end wills the means” is analytic (G 4:417). This seems to suggest that if you don’t will the means, then it logically follows that you don’t really will the end. But if that were true in the plainest sense, no one would ever be guilty of violating a hypothetical imperative. For if someone didn’t will the means, then it would follow logically that he didn’t will the end, and in that case, of course, he wouldn’t have violated the hypothetical imperative, which only tells him what to do if he *does* will the end. This, however, leaves us unable to give sense to the claim that instrumental principles are imperatives—for how can they be imperatives, if they are impossible to violate?² Later I will argue that the hypothetical imperative is constitutive of action (4.3; 5.1.3), but it cannot follow that it is not normative for it as well. Here’s another example that you might find more readily convincing. The presence of both a noun and a verb in an English sentence is *constitutive* of its being a sentence, that is, of its expressing a complete thought. Yet those of us whose work includes grading papers have all encountered the verbless string of words that wants to be a sentence and fails, and yet is not mere gibberish. There is such a thing as speaking English badly, and it is not quite the same as not speaking it at all, although—importantly—it tends in that direction. For if you ignore the rules of English altogether, what you speak will simply not be English.

² For a more complete version of this argument, see “The Normativity of Instrumental Reason” (CA essay 1).

2.1.5

So we are looking at a quite general problem about finding the conceptual space between performing an activity perfectly and not performing it at all, space into which we can fit the person who does it badly. Among the ancient Greek philosophers this seems to have been one of the standard puzzles about art or craft. At least it comes up in the first book of the *Republic* with respect to the art of ruling. Thrasymachus says that justice is the advantage of the stronger, for the rules of justice are imposed on the weak by the strong, and the strong rule for their own advantage. Socrates pretends to be puzzled by the question where justice lies when the strong make a law that is not *in fact* to their advantage (R 339c–e). Thrasymachus replies that the problem is the result of a loose way of talking. In the *precise sense*, he says, no craftsman, expert, or ruler, is a craftsman, expert, or ruler, at the very moment when he makes an error (R 340d–341a). In other words, Thrasymachus concludes you are not practicing an art at all if you practice it badly. Socrates proceeds to make mincemeat of Thrasymachus with this “precise sense” by showing that a ruler, in the *precise sense*, rules for the benefit of whatever he rules, and not for his own benefit (R 341c–343a).

In fact the “precise sense” or perfect version of an activity stands in a complex relation to the activity, because it is at once both normative and constitutive. Although it is not true that you are not performing an activity at all unless you do it precisely, it is true that you have to be *guided by* the precise version of the activity in order to be performing the activity at all. And at the same time the precise sense sets normative standards for the activity. It is tempting to say that the actual activity must *participate* in the perfect or precise one. In other words, Plato’s Theory of Forms is true for activities.

The shoddy builder doesn’t follow a different set of standards or norms. He may be doing one of two things. He may be guided by the norms, but carelessly, inattentively, choosing second-rate materials in a random way, sealing the corners imperfectly, adding insufficient insulation, and so on. But he may also, if he is dishonest, be doing this sort of thing quite consciously, say in order to save money. In that case, surely we can’t say that he is trying to build a good house? No, but now I think we should follow Socrates’s lead, and say that he is not trying to build a house at all, but rather a sort of plausible imitation of a house, one he can pass off as the real thing. What guides him is not the aim of producing a house, but the aim of producing something that will fetch the price of a house, sufficiently like a real house that he can’t

be sued for it afterwards. Socrates, in the passages from the *Republic* that I have already mentioned, makes rather a fuss about this point, insisting that a craftsman in the precise sense is not a money-maker, but simply a practitioner of his craft (R 341c–342a).

2.1.6

So on this conception, every object and activity is defined by certain standards that are both constitutive of it and normative for it. These standards are ones that the object or activity must at least try to meet, insofar as it is to be that object or activity at all. An object that fails to meet these standards is bad in a particular way. It will be useful to give this kind of badness, badness as judged by a constitutive standard, a special name, and in English we have a word that serves the purpose well: *defect*. So in the somewhat special sense that I will be using the term, a house that is so constructed as to be ill-adapted for sheltering is *defective*; while a house that blocks the neighborhood's view, though it may for that reason be a bad thing, is not *a defective house*. Since the function of action is self-constitution, I am eventually going to argue (Chapter 8) that bad actions, *defective* actions, are ones that fail to constitute their agents as the unified authors of their actions.

2.1.7

Constitutive standards are important, I claimed above, because they meet skeptical challenges with ease. But the importance of the idea is deeper than that, for I believe—and I know this is more controversial—that the *only* way to establish the authority of any purported normative principle is to establish that it is constitutive of something to which the person whom it governs is committed—something that she either is doing or has to do. And I think that Kant thought this too. The laws of logic govern our thoughts because if we don't follow them we just aren't thinking. Illogical thinking is not merely bad, it is *defective*, it is bad *as* thinking. The laws of the understanding govern our beliefs because if we don't follow them, we just aren't constructing a representation of an objective world (9.7.5). And as I will argue, the laws of practical reason govern our actions because if we don't follow them we just aren't acting, and acting is something that we must do. A constitutive principle for an inescapable activity is unconditionally binding.

How could it be otherwise? Constitutive standards have unquestionable authority, while external standards give rise to further questions, and leave space for skeptical doubt. How then can we ever give authority to an external standard, except by tracing its authority back to a constitutive one? Consider

again that house that blocks the neighbors' view of the lake. Why shouldn't the house-builder build it? For I'm supposing that we all do agree that really, after all, he shouldn't do it, in spite of the fact that it wouldn't therefore be a *defective* house. Well, perhaps he identifies himself as a good neighbor, a citizenly type, and doesn't need to ask why he shouldn't build a house that is a blight on the neighborhood. Or perhaps he loves his neighbors, and wouldn't want to harm them. Or perhaps—to anticipate the success of the views we are working on here—it would be morally wrong to build a house that blocks the view of the neighbors, and so although it might be all very well as a bit of *house-building*, it would be *defective* as an *action*.

2.1.8

There is another reason why the idea of a constitutive standard is important—or rather, this is the same reason, described a different way, coming from a different direction. It is that we *need* the concept of the *defective*, in the sense described above. Say we have two objects, call them A and B, and they are in some respect different from each other. They have some different non-accidental properties. Now we need to distinguish two ways that A and B can be different from each other in this way: A can be a different *kind* of thing from B, or A can be a *defective instance* of the same kind of thing as B. Suppose A is a defective instance of the same kind of thing as B. Then say we have two objects Y and Z, which differ in regard to the same property, but which *are* of different kinds. Should we treat these two cases, the case of A and B and the case of Y and Z, any differently? Does it matter what *kinds of things* things are?³ Why shouldn't all that matters be the properties themselves? If properties are all that matter, then we need not—and cannot—distinguish the different from the defective: different collections of properties will just be different, and that is all.

Well, consider again the case of instrumental reasoning. Kaspar says he resolves to begin a course of exercise tomorrow, in order to get in shape, but he does not do it. If he has changed his mind about the value of

³ One place the question of difference and defect comes up in the philosophical literature is in discussions of the moral standing of animals. The so-called “marginal cases” argument holds that if we accord defective human beings a certain moral standing, then there is no reason not to accord that standing to animals who lack the property with respect to which the human being in question is defective. I believe that this argument is mistaken. I think that a better argument can be made for according moral standing to the other animals. I sketch such an argument in my “Fellow Creatures: Kantian Ethics and Our Duties to Animals” and in “Interacting with Animals: A Kantian Account,” forthcoming in *The Oxford Handbook on Ethics and Animals*, ed. Thomas Beauchamp and R. G. Frey (Oxford, 2010).

getting in shape, or if he was lying when he announced his resolution, his volition is merely different from what we expected it to be. But if he does not exercise because he is suffering from weakness of the will, his volition is defective: he has performed an abortive act of will. It must be possible to distinguish these two cases. If his shifting volitions can only be different, and not defective, then he has not violated, and cannot violate, any principle or norm.⁴

Or to take once again a similar but more troublesome case, think of language. Someone violates the rules of English as you understand them. Is he speaking a dialect, or making a mistake? Perhaps he is speaking a dialect—we must certainly admit the possibility, to avoid intolerance—and then what he is doing is simply something different. These cases can be vexed—the adults may regard as merely erroneous what the children take to be a legitimate form of slang, for instance. But if everyone who speaks differently is allowed to counter criticism with the claim that he is simply using a different dialect, then there are no rules of English.

Another example: some physical differences, say hair color, are just that—differences. We regard others as defects, and those who suffer from them as unfortunate. Sometimes people try to deny this, often from laudable motives of respect and consideration. Being deaf, they claim, is not a defective condition, but is just a difference—the source of a different way of learning from and communicating with others. But we offer those who suffer from defective conditions special aid and compensation from society. If they were only different, why should we do that?

Distinguishing cases of difference from cases of defect can be difficult. As some of the examples I've given show, it can even be politically charged or delicate. It can also be largely pragmatic. Being short makes it harder to do certain things, just as being deaf does, but we do not regard this as a handicap. Perhaps this is because in a species like ours, not all of a single height, some people will necessarily be short. Some differences become defects only when they reach certain extremes. We should grant all these points about how hard it is to distinguish the different from the defective. Nevertheless, we need the concept of the defective for all sorts of purposes. And if we try to banish the concept of the defective from the world altogether, we will banish normativity along with it: nothing will violate any standard that necessarily applies to it; everything will just be different. And that is why we need constitutive standards.

⁴ As I argue in “The Normativity of Instrumental Reason” (CA essay 1), especially pp. 48–50. But for some complications about that argument, see also the discussion below at 4.3.4.

2.2 The Constitution of Life

2.2.1

In 1.4.3, I mentioned what I called “the paradox of self-constitution.” How can you constitute yourself, create yourself, unless you are already there? And how can you need to constitute yourself if you *are* already there? With Aristotle’s view before us, we are now ready to start working our way towards the solution of this problem.

Aristotle extended his account of artifactual identity to living things with the aid of the view that a living thing is a thing with a special kind of form.⁵ A living thing is a thing so designed as to maintain and reproduce *itself*: that is, to maintain and reproduce its own form. It has what we might call a self-maintaining form. So it is its own end; its *ergon* or function is just to be—and to continue being—what it is. And its organs, instincts, and natural activities are all arranged to that end. The function of a giraffe, for instance, is to be a giraffe, and to continue being a giraffe, and to produce other giraffes. We might therefore say that a giraffe is simply an entity organized to keep a particular instance, a spatio-temporally continuous stream, of giraffeness going—primarily through nutrition—and also to generate other instances of giraffeness, through reproduction. A healthy giraffe is one that is well-organized for keeping her giraffeness going, while an unhealthy giraffe suffers from conditions that tend to her disintegration. So health is not, strictly speaking, a *goal* for giraffes, but rather is our name for the inner condition which enables the giraffe to successfully perform her function—which is to go on being a giraffe. This parallels the way in which, as I said in 1.1.5, goodness is not a goal for people, but rather is our name for the inner condition which enables a person to successfully perform her function—which is to maintain her integrity as a unified person, to be who she is. This is why Plato and Aristotle always compared health to virtue.

It is important to notice the complex role that teleological organization plays with respect to the giraffe’s activities and actions. The giraffe’s actions are both dictated by, and preservative of, her giraffeness. A good giraffe action, such as nibbling the tender green leaves at the tops of the trees, keeps the giraffe going, for it provides the specific nutrients needed to constantly restore and refurbish her giraffeness through the nutritive processes. Yet the giraffe’s action is one to which she is prompted by instincts resulting *from*

⁵ To the aforementioned central books of Aristotle’s *Metaphysics*, now add *On the Soul*, especially Book 2. *Physics* 2 is also helpful.

her giraffe nature. This is related to an apparent difference between living things and artifacts, which is that living things are made of parts that strictly speaking cannot exist independently of the living things themselves. You can't build a giraffe out of tender green leaves, but a giraffe's nutritive processes turn tender green leaves into the kinds of matter out of which a giraffe *is* built—giraffe tissues and giraffe organs and so on. Furthermore, the living tissues that make up organisms are comparatively fragile, and in need of constant renewal. It follows from all this that if a giraffe ceases her activities—if she stops nibbling the tender green leaves, or stops digesting them when she does—she will fall apart. And this means that, strictly speaking, being a giraffe is not a state, but rather an activity. Being a giraffe is *doing* something: a giraffe is, quite essentially, an entity that is always *making* herself into a giraffe. In fact, the *entity* that I just mentioned is derivative, arrived at only by an artificial freezing of the observer's mental frame, for nothing that stops working at being a giraffe, that stops making herself into a giraffe, will remain a giraffe for long. So to be a giraffe is simply to engage in the activity of constantly making yourself into a giraffe: this is what a giraffe's life consists in.

2.2.2

I said that living things are apparently different from artifacts because, strictly speaking, the parts of living things do not exist separately from the living things themselves. But actually, speaking *very* strictly, this is true of artifacts too—their parts cannot exist independently of the artifacts themselves. For example, large slabs of sheetrock or plaster can exist apart from houses, but *walls* cannot, for walls are functionally defined, and a slab of sheetrock or plaster that isn't part of a house cannot divide one room from another, or help to hold up a roof. But perhaps the only reason to bother making this point is to support the parallel with organisms.

But perhaps not. If we don't draw the parallel, and regard artifacts as having separately existing parts, then it seems as if artifacts are, or can be, static entities, not essentially activities, the way living things are. And I don't think they can. An artifact is defined in terms of its *essential* activity: it is the thing that can perform that activity. But in fact, most artifacts cannot perform their activities all by themselves. They need either a power source, or to be wielded by a human being, or both, before they can perform their functions. It isn't quite right, then, to say of the vacuum cleaner in your closet that it "can clean floors," since, actually, until you plug it in and wield it, it cannot. So strictly speaking, artifacts, when they are just sitting around doing nothing on our

shelves or in our closets, are *incomplete objects* that will only start to perform their function when some last part is plugged in or inserted. In fact, the truth about this matter looks as if it may be depressing: there is no such artifact as a vacuum cleaner at all. Instead, what you call your vacuum cleaner is actually an entity that, when properly incorporated by you, makes *you* into a vacuum cleaner.

All sorts of strange conclusions follow from this line of thought: reality is essentially activity, for all static entities are in general only the result of freezing the observer's mental frame; all those objects in your attic and garage are not entities after all, but only half-constructed heaps waiting to be finished; and indeed there is no such thing as an artifact, although human beings and the other tool-using animals throw themselves into an enormous variety of artifactual modes . . . Okay, I'll stop.

2.2.3

To be a giraffe is simply to engage in the activity of constantly making yourself into a giraffe: this is what a giraffe's life consists in. And for the same reasons that we considered earlier there is no real difference between the activity of living a giraffe's life, and the activity of living a healthy giraffe's life, for in order to live a giraffe's life, you must follow the teleological principles implicit in the form of giraffeness, the constitutive principles of being a giraffe. And so leading the life of an unhealthy giraffe is not a different activity from leading the life of a healthy giraffe. *It is the same activity, badly done.*

2.3 In Defense of Teleology

2.3.1

We are almost ready to solve the paradox of self-constitution, but first I want to address another issue. The account of the normativity of practical reason that I am working on here grounds normative standards in a frankly teleological, Aristotelian, conception of objects and activities. Many philosophers are worried by teleological ways of conceiving the world. Hasn't Aristotle's idea that there are natural purposes, or that the world and the things in it were made for a purpose, long since been discredited by the Modern Scientific World View? My response to these worries will come in three parts: first, I will give an account of the target and scope of the teleological conception I propose to use; second, I will give an account of what justifies its use; and finally I say a few words about the resulting status of teleological claims.

First the target and scope of teleological thinking. The Aristotelian conception that I have just laid out identifies objects as having an internal

teleological organization. This is clearest in the case of living things, where the claim is simply about how the living thing's organs and activities are conceived and explained as contributing to its life. A living thing is not assigned a purpose outside of itself—its “purpose,” or more properly function, is to be what it is, to live its particular form of life.⁶ Thus there are no such claims here as that horses are meant for riding into battle or that cows are meant for human beings to eat or that women are meant for housework or that oil is meant for lamps and automobiles. The teleological claims are made at the level of the individual object: they are claims about its internal organization. It is of course true that we can identify something as having an internal teleological organization only to the extent that we can identify it as *doing something*. Serving a human purpose is one recognizable way of *doing something*; but doing what we ourselves do—namely, living—is another. (Even in the case of an artifact its purpose need not be thought of as external to the object, since in the case of an artifact it is the whole nature of that object to serve the purpose in question.) In fact what I want to claim—although I will have to be a little vague here—is that this is how we pick out the object, how it emerges from what Kant called the sensible manifold as a unified thing. That is to say, we pick out an object as a region of the manifold that appears to be doing something, and we understand it as a single and unified object by understanding it as internally organized for doing whatever it does.⁷

This brings me to the second point—the justification of teleological thinking. That justification falls into two related parts. The first is the claim I have just made. Teleological thinking need not be grounded in a claim about the world. It may be grounded in a claim about how human beings conceptualize the world. The idea, of course a Kantian one, is that human beings are faced with the task of carving the sensible manifold into objects. The claim is that we pick out objects by identifying functional unities. *Very* roughly speaking, the idea is this: in dividing the world into objects, we need some reason for carving out more particular unities from the sensible manifold. And the kind of unity that grounds the identification of a particular object is a functional unity. To put it a bit fancifully, when a cluster of forces are all contributing to something that we, by our admittedly human standards, would call a *result*, then we bunch these forces together, and call them an object. When a cluster of natural forces works together to produce something I can sit down on, say

⁶ For a more complete account of what I think Aristotle means by function, and a defense of his “function argument” in NE 1.7, see my “Aristotle’s Function Argument” (CA essay 4).

⁷ I think this is roughly what Kant means by “reflective judgment.”

a flat rock, then I call it a *seat*. When I try to reproduce that cluster of forces, I call the result a *chair*. When a cluster of forces works together to maintain and continually reproduce that same cluster of forces, or a cluster of forces spatio-temporally continuous with itself, thus constantly making itself and copies of itself, then I call it a *living thing*.

And that has implications for the status of the resulting teleological claims. If we are to pick out self-maintaining regions of the manifold as living things, of course, there must be such things, so I do not mean to imply here that living things are merely human constructs, or anything of that sort. Not that that would necessarily be so bad. “Seats” are human constructs, since the concept of a seat is relative to the purposes of an erect-standing creature so constructed as to be able to sit down. “Chairs” like other artifacts are human constructs, but then, no one doubts *that*—the concept and its object are born together in the original craftsman’s mind. For all that, however, there are chairs. Why do we pick self-maintainers out of the manifold as a kind of thing? As anyone who watches animals knows, animals or at least middle-sized multicellular animals in fact recognize one another as fellow animals without any fancy powers of conceptualization, so perhaps this question needs no answer. But our later recognition of living things as self-maintainers could have been inspired by the analogy with ourselves. Nothing I’m saying here is incompatible with a Darwinian account of how the world became populated with items fit to be thus conceptualized. And nothing I’m saying here is incompatible with all the ways in which the Darwinian account implies that teleological thinking can be wrong. We can wrongly assign a purpose to a useless vestigial organ, for example. We can conceive of something as relative to our purposes, when it has interests of its own that make a different understanding of its organization available. So there is no claim here that everything has one and only one purpose that is in fact its natural purpose. The claim is simpler—it is that the way we conceptualize the world, the way we organize it into a world of various objects, guarantees that it will appear to be teleologically organized at the level of those objects.

2.3.2

The idea that teleological thinking is inherent in our powers of conceptualization is a development of a point that is implicit in what I have already said. A teleological conception of the world is essential to our functioning as agents. We need the world to be organized into various objects in order to act. To recognize an object as *doing something* or as producing a *result* of some kind is to identify it with reference to our own purposes and powers of

action. Since we must act, the world is for us, in the first instance, a world of tools and obstacles, and of the natural objects of desire and fear. An object is identified as a locus, a sort of force field, of particular causal powers, and the causal powers in question are identified as those we might either use or have to work against. And if we did not identify objects in this way, we could not act at all.

Let me put this point more specifically. As I will be arguing later on, Kant's hypothetical imperative is a normative principle essential to, constitutive of, action itself (4.3). To act is essentially to take the means to your end, in the most general sense of the word "means." And to take the means to your end is, as Kant himself pointed out, to determine yourself to cause the end—that is, to deploy the objects that will bring the end about. Thus action requires a world of objects conceived as the loci of causal powers. You intend to cut, for instance, so you look for a knife, conceiving the knife as the cause of cutting.

Now perhaps some people suppose that as long as you conceive the knife merely as *the cause of the cutting*, rather than as *for the purpose of cutting*, you are not conceiving of the world teleologically. The view that the knife is *the cause of the cutting* is mechanistic. But is it? In the purest form of the mechanistic view, the knife is not the cause of the cutting. It is rather—say—the knife wielded by the hand directed by the brain operating through the nervous system stimulated by certain forces determined in turn by certain events caused in certain ways. Assuming something like determinism is true at the level of middle-sized objects, the cause of the cutting is the state of the world a nanosecond ago determining the state of the world now. Why then do we say that *the knife*, rather than the state of the world a nanosecond ago, is the cause of the cutting? That is easy—because we can *use* the knife for cutting. From the purely mechanistic point of view, the identification of a particular object or even a particular event as *the cause* of another is artificial, a piece of shorthand, a sort of conception of thumb, if I may put it that way. The teleological view—the view of the world as a realm of tools and obstacles—stands behind the slightly artificial idea that particular objects are "causes." But the teleological conception of the world is essential to creatures who are inside of the world and must act in it.

2.3.3

The teleological view of the world as a realm of tools and obstacles, of objects of desire and fear, the conception of the world from which as agents we must start, is modified by rationality in two ways. One modification occurs

within the teleological conception itself, and as I will argue is an inevitable development of it. It is the moral conception of the world. To act, I have already suggested, is to determine yourself to be the cause of a certain end. So to act *self-consciously* is to conceive of *yourself* teleologically—as the cause—that is, the *first* cause—of a certain end. It is to conceive yourself as an agent, as efficacious to achieve certain subjectively held ends. Thus in addition to tools and obstacles and objects of fear and desire, a rational, self-conscious agent comes to conceive the world as containing agents, with ends of their own. She comes to conceive the world, in Kant's language, as a Kingdom of Ends: a whole of all ends in themselves or first causes, with the ends that each sets before him or herself (G 4:433).

The other modification, which eventually emerges as an alternative conception, is the scientific or mechanistic conception of the world (6.1.6). It is a conception that results among other things from pressing the notion of cause, as I did above, until the idea of *a* cause within the world begins to look spurious. Or, to put the same point another way, it is the result of pressing our understanding of the world until the idea of an object, as a unified and independent being within the world, begins to look spurious. You think you're an object, indeed even an agent, but to a flea or a nit you are merely a rather nutritious and specific region of the environment, like a Pacific island. If the flea or nit could think, it would think itself an object, perhaps even an agent, but to the cells in its body it is merely a rather nutritious and specific environment . . . and so on. Even we self-identifying self-conscious and supposedly self-maintaining substances fail to see how thoroughly embedded we are in an environment that supports us from outside, how thoroughly our perceived internal unity and cohesion depends on what goes on around us. A chemical change, a rise in the temperature, a stray bullet, and the transient whirling vortex of forces that thought itself an immortal *thing* puffs away . . .

Are the teleological and moral conceptions of the world then related to the Scientific World View as illusions to fact? If that were so, whose illusions would they be?

2.4 The Paradox of Self-Constitution

2.4.1

Now we are ready to talk about the paradox of self-constitution. According to the Aristotelian picture of the nature of living things, a living thing is engaged in an endless activity of self-constitution. In fact to be a living thing is just to be self-constitutive in this way: a living thing is a thing that is constantly making

itself into itself. But notice that the apparent paradox involved in the idea of self-constitution does not seem to arise here. No one is tempted to say: “how can the giraffe make itself into itself unless it is already there?” The picture here is not of a craftsman who is, mysteriously, his own product. The picture here is of the self-constitutive process that is the essence of life. The paradox of self-constitution, in this context, is no paradox at all.

And the same applies to personhood. Aristotle believed that there are three forms of life, corresponding to what he called three parts of the soul.⁸ Each supervenes on the one below it. At the bottom is a vegetative life of nutrition and reproduction, common to all plants and animals. According to Aristotle, animals are distinguished from plants in being alive in a further sense, given by a functionally related set of powers that plants lack. Aristotle emphasizes perception and sensation, but notes that these are necessarily, or at least usually, accompanied by imagination, pleasure and pain, desire, and local movement (OS 2.2 413b22–24). What is distinctive of animals is that they carry out part of their self-constitutive activities through action.

The third form of life, distinctive of human beings, or as I will say, of persons, is the life of rational activity. Rational activity, as I have already suggested, is essentially a form of self-conscious activity, and it is this that leads to the construction of personal identity. Thus personhood is quite literally a form of life, and being a person, like being a living thing, is being engaged in an activity of self-constitution.

In other words, what it is to be a person, or a rational agent, is just to be engaged in the activity of constantly making yourself into a person—just as what it is to be a giraffe is to be engaged in the activity constantly making yourself into a giraffe.

2.4.2

One way to bring out the force of this point is in terms of the idea of practical identity. In 1.4, I proposed that we constitute our own identities in the course of action. In choosing in accordance with the principles of a form of practical identity, I claim, we make that identity our own.

It is sometimes said, in opposition to this sort of point, that it involves an overly voluntaristic conception of identity. I did not choose to be an American citizen, or my parents’ daughter. Even many of my personal friendships, the older ones especially, are as much the outcome of circumstance as of choice. So I am these things—this country’s citizen, these people’s daughter, this

⁸ These views are found especially in *On the Soul* 2–3. See also NE 1.7 1097b32–1098a5.

person's old friend—*perforce*, and not because I chose to be them. And yet these identities give rise to reasons and obligations, as much as the ones that I do more plainly choose, like a profession or an office or a friendship quite deliberately sought out. But I want to argue that while that is true in one way, in another way it is not. For whenever I act in accordance with these roles and identities, whenever I allow them to govern my will, I endorse them, I embrace them, I affirm once again that I am them. In choosing in accordance with these forms of identity, I make them my own.

The idea that to be a person is to be constantly engaged in making yourself into that person helps to explain what is going on in this debate. To see how this works, consider one of the standard dilemmas of contemporary moral philosophy. Some people have complained that the Kantian self is “empty.”⁹ If you conceive yourself simply as a pure rational agent, and are not committed to any more specific conception of your identity, you are, as it were, too distant from yourself to make choices. There are two problems here. The more formal problem is that it looks as if your empty self can have no reason to do one thing rather than another. But even if you can find some particular reasons, there is also a problem about wholeheartedness, about commitment. How can you be a true friend, a true citizen, a true Christian, say, if the relevant commitments are always up for question and open to choice? The self, it is argued, must be not empty but rather determinate and full: it must take certain identities and relationships as unquestionable law.

And then of course the other side replies that there are also two problems with the determinate self. In the first place, the determinate self is not free, for its conduct is governed by a principle or a law which is not reason's own. In the second place the determinate self must in the end be unjust. For tolerance requires exactly that distance from our roles and relationships that the defenders of the determinate self deplore. “Christianity is my religion, but just in the same way, Islam is his,” says the tolerant person. Tolerance demands that you see your religion not as *you* but as *yours*, yourself not as essentially a Christian but as essentially a person who *has* a religion—and only one of many you might have had. So you cannot identify with your religion all out and still be a tolerant person. Or so says the defender of the empty self.

Now this is a false dilemma, arrived at by an artificial freezing of the observer's mental frame. It assumes that the endorsement of our identities, our self-constitution, is a state rather than an activity. If self-constitution were a state we would be stuck on the horns of this dilemma. Either we must

⁹ See for instance Michael Sandel in *Liberalism and the Limits of Justice*.

already have constituted ourselves—in which case the self would be full and determinate. Or we must not have done so yet—in which case the self would be empty.

But we don't have to choose between these two options, because self-constitution is not a state that we achieve and from which action then issues. As I will try to show in the course of the next four chapters, it is action itself.