
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU


HARVARD LIBRARY
Office for Scholarly Communication

Goodness and Desire

Citation

Boyle, Matthew and Douglas Lavin. 2010. Goodness and desire. In Desire, Practical Reason, and the Good, ed. Sergio Tenenbaum. New York: Oxford University Press.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:4879173

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

[To appear in S. Tenenbaum, ed., *Desire, Practical Reason, and the Good*, Oxford UP, forthcoming 2010]

Goodness and Desire[*]

Matthew Boyle and Douglas Lavin
Harvard University

> It is always the object of desire which produces movement,
> but this is either the good or the apparent good.
>
> Aristotle, *De Anima*, III.10 (433a28-29)

1. Introduction

Aristotle famously held that all desiring is directed toward some actual or apparent good, and a long train of celebrated philosophers have agreed with him, from Aquinas and Kant to Anscombe and Davidson. Philosophers writing in this tradition have tended to find Aristotle's proposition not merely correct but obvious, so obvious that the Scholastics codified it in a dictum: *quidquid appetitur, appetitur sub specie boni*, "whatever is desired is desired under the guise of the good." The last few decades, however, have seen a vigorous attack on this guise of the good thesis. The Aristotelian doctrine has been dismissed as not merely wrong but naïve, reflecting a picture of the human heart that is either hopelessly innocent or deliberately one-sided.[1] What accounts for this transformation? How could what seemed plain to so many come to seem so patently false?

Contemporary critics of the guise of the good thesis often support their rejection of it by citing putative counterexamples: cases in which people fail to desire what they regard as good (*accidie*), desire most what they do not think best (*akrasia*), or even desire what they do not regard as good at all, as when one person feels a purely malicious or spiteful urge to lash out at another. Milton's Satan, when he resolves "Evil be thou my good," expresses a

---

[*] For comments and suggestions, we are very grateful to Rachel Cohen, Wolfram Gobsch, Matthias Haase, Dick Moran, Jessica Moss, Arthur Ripstein, Sebastian Rödl, Tamar Schapiro, Kieran Setiya, Martin Stone, Gisela Striker, Sergio Tenenbaum, and to audiences at the universities of Basel, Colgate, Leipzig, and Toronto.

[1] A representative expression of this new standpoint occurs near the beginning of Michael Stocker's "Desiring the Bad" (1979). According to Stocker,

> …it is hardly unfair, if unfair at all, to suggest that the philosophical view is overwhelmingly that the good or only the good attracts. At least, this is how I am forced to interpret so many philosophers. This affords me no pleasure, since that view… is clearly and simply false. (p. 740)

For similar assessments, see J. David Velleman, "The Guise of the Good" (2000c), esp. pp. 118-119 and Kieran Setiya, "Explaining Action" (2003), esp. p. 353.

perversity that is extreme in its degree, but the kind of impulse he feels is hardly unknown to us: we are all familiar with the chasms that can open between what we want and what we take to be good or valuable. Gary Watson provides a vivid sketch of this parting of ways:

> The cases in which one in no way values what one desires are perhaps rare, but surely they exist. Consider the case of a woman who has a sudden urge to drown her bawling child in the bath; or the case of a squash player who, while suffering an ignominious defeat, desires to smash his opponent in the face with the racquet. It is just false that the mother values her child's being drowned or that the player values the injury and suffering of his opponent. But they desire these things none the less. They desire them in spite of themselves.[2]

Pointing to such cases, critics of the guise of the good thesis ask: How can it be true that desire is directed toward the good if sometimes it palpably isn't?

It seems doubtful, however, that attention to such cases by itself accounts for the contemporary turn against the guise of the good thesis. For whatever philosophers writing in the Aristotelian tradition meant by saying that we desire under the guise of the good, they were hardly unaware that people can want what they take to be wicked or worthless: they themselves emphasized such cases. What seemed to them obvious was not that perverse desires are impossible, but that they must be understood in a certain way: as involving a kind of mutiny of some of our motivational faculties, a mutiny that results in an object's being put forward as desirable by one faculty even as others deny its desirability. Contemporary critics of the guise of the good thesis, by contrast, do not feel pressure to interpret cases of perverse desire in this way: they do not see why *desiring* something need involve any tendency, however partial, to regard it as *desirable*. To understand the basis of the recent turn against the guise of the good thesis, we must understand the reasons for this change.

Why might it seem that desiring something must involve some tendency to see that thing as desirable? The intuitive reason seems to be this: a person who wants something can in general be asked *why* he wants it, and then he is expected to answer, not by giving some sort of psycho-history of his own desire, but rather by offering another sort of account – one that describes what we call his "reasons" for wanting the thing in question, the considerations he takes to "speak in favor" of pursuing that object.[3] The guise of the good thesis is, in effect,

---

[2] Gary Watson, "Free Agency" (1982), p. 101.

[3] For a *prima facie* justification of the guise of the good thesis by appeal to these sorts of observations, see for instance Joseph Raz, "The Guise of the Good" (this volume).

a claim about what sort of thing a reason is: namely, that it is a consideration that bears on what is good about what is wanted. Contemporary opposition to the thesis is bound up with a rejection of this claim. Thus many contemporary philosophers of action insist on a deep distinction between *justifying reasons*, which speak to the goodness of an object or course of action, and *explanatory reasons*, which speak to the question why, in fact, an agent was motivated to do a certain thing.

If an agent can answer the question why he wants to do something, then, on this view, he will be giving an explanatory reason, and while this *may* bear on what is good about the thing he wants, there is no obvious reason why it *must*.

This suggests that the real source of the turn against the guise of the good thesis lies in a shift in the understanding of reasons-explanation. The Aristotelian tradition sees such explanation as a particularly sophisticated form of *teleological* explanation, one that represents us as drawn, not simply toward ends that are in fact good, but toward goods that act on us in virtue of our representing them as such. Contemporary philosophers, by contrast, are nearly unanimous in their unwillingness to countenance an unreduced appeal to teleology. The fundamental explanation of action, they assume, must advert to some precedent psychological state of the subject, a state which is the efficient cause of certain bodily movements, and whose causal role in producing these movements can be specified without appeal to teleological notions.[4] It may be that our desires tend to move us toward things that are good for us; indeed, this tendency may figure in an evolutionary explanation of why *Homo sapiens* typically desire what they do. But even if it is a fact that our desires *tend* to

---

[4] This is not to say that contemporary authors are unwilling to speak of a person acting *in order to achieve a certain goal* or *because he had a certain aim*. Most authors admit that these are genuine explanations, and that they may be called "teleological" inasmuch as they relate the agent's action to an end. What they do not admit is that the goodness, or represented goodness, of the end plays an essential role in the explanation: on their view, to say that an action is explained by a certain end is really to say that it is explained by a certain precedent psychological state of the agent (his wanting to achieve that end) whose efficacy does not depend on its tendency to move the agent toward things that are good for him. This view of action-explanation accords well with the widespread view that, in general, superficially teleological forms of explanation should in principle be reducible to explanations of other kinds. As John Hawthorne and Daniel Nolan observe in a recent article on teleological explanation:

> [W]hen contemporary philosophers and biologists tell stories about… natural teleology they tend to proceed as if there is a different underlying explanation: superficial teleology gives way to an underlying reality that is not fundamentally teleological at all. This is so even in the case of mental activity. Teleology gives way to mental representations that play efficient causal roles (which in turn may enjoy deeper explanations that proceed via categories that are not mentalistic at all). ("What Would Teleological Causation Be?" [2006], p. 266)

move us toward what is beneficial to us, it is hard to see, from this standpoint, why they *must* do so, and still harder to see why they must *represent* their objects as good. For why shouldn't desires for things that the subject does not evaluate as good be capable of playing the same sort of causal role in producing behavior as desires which rest on such an evaluation? The characteristic explanatory role of desire seems to be one topic; the relation, if any, of desire to a sense of justification, another. It may be true that an agent can normally give his reason for acting in the sense of: the desire which explains his acting; but to infer that an agent must normally take himself to have a reason for acting in the sense of: a justification for doing what he does – this, it seems, is to equivocate between two different senses of "reason for acting." As Kieran Setiya remarks in an important recent discussion:

> [W]e need a causal-psychological account of taking-as-one's-reason, not an account that appeals to what is seen as good. When an agent acts on a reason, he takes it *as* a reason, but that means he takes it as *his* reason, not that he takes it to be a *good* reason on which to act.[5]

Our aim in this essay is to challenge the assumptions that make such a distinction seem necessary, and thus make the Aristotelian doctrine about the relation between goodness and desire seem puzzling. We shall argue that the primary source of contemporary opposition to the guise of the good thesis lies in a certain set of views about action and its explanation, views which form the core of the standard "causal theory of action." But, we shall suggest, these views imply substantive and questionable commitments about the shape an account of intentional action must take; and once these commitments are questioned, the guise of the good thesis reemerges, not as an isolated and dubious claim about what motivates people to act, but as a proposition belonging to an attractive and coherent account of *what action is*.

On the Aristotelian view, what underlies the fact that rational agents must desire under the guise of the good is a general point about the explanation of any *self*-movement or self-change, a point that applies, in different forms, to the explanation of behavior in nonrational animals, and even to the explanation of the nutrition, growth and reproduction of nonsentient

---

[5] "Explaining Action" (2003), p. 380. Compare also J. David Velleman's contention, in an influential paper on the guise of the good thesis, that

> even if desiring something consists in regarding the thing as good in a sense synonymous with "to be brought about," it isn't an attempt at getting right whether the thing really is to be brought about, and so it doesn't amount to a judgment on the thing's goodness. ("The Guise of the Good" [2000c], p. 117)

living things. What these various kinds of explanation have in common is that they all have a teleological structure; and it is characteristic of the Aristotelian tradition to claim that, quite generally, this sort of structure applies only to a subject that is the bearer of a certain sort of *form*, a form that constitutes a standard of goodness for the subject in question, but that is equally implicated in any explanation of what the subject itself does. The special feature of the application of this explanatory structure to rational creatures is that such creatures belong to a kind in which this connection between action and goodness becomes self-conscious: they are creatures whose action is expressive of and explained by a self-conception which implicitly involves a *conception* of their own form. This, it will emerge, is why rational self-movers must desire under the guise of the good.

Our project in what follows will be to fill out this Aristotelian view of action, and, we hope, to present it in such a way that it appears to be, not merely a set of strange and antiquated ideas about how the world works, but rather a serious and defensible attempt to understand what action is. Our main aim in doing this is simply to bring out what is at stake in the acceptance or rejection of the guise of the good thesis. Many discussions of the thesis treat it as meriting attention simply because it has a certain historical authority and also, perhaps, a certain intuitive appeal. If we are right, its real interest lies in its connection with an approach to action that stands opposed in fundamental ways to the approach that dominates contemporary action theory. We can hardly hope to make a conclusive case for this alternative approach here, but we hope at least to make clear what it is and why it might have a powerful appeal. We will make this case by first presenting a challenge to the causal-psychologistic approach typically presupposed by critics of the guise of the good thesis (§3), and then sketching the outlines of an alternative, one that takes its departure from a general thought about the connection between self-movement and the form or nature of a thing (§4), and that represents the guise of the good thesis as the upshot of this general thought when it is applied to rational creatures (§5). To clear the way for this enterprise, however, it will help first to make a few points about how to understand the thesis and how to assess it (§2).

2. The meaning of the thesis and how to assess it

Before deciding whether the guise of the good thesis is true, we must first consider what it means, and already here we face difficulties. Aristotle himself formulates the thesis in

various ways: sometimes as a claim about the object of desire, sometimes as a claim about the aim of every "action or pursuit."[6] On which formula should we focus? Moreover, what should count as the "object" of a desire, or the "aim" of an action? If I want to eat an apple, for instance, is the object of my desire *an apple*, or *eating an apple*, or perhaps *my eating an apple*? Does the guise of the good thesis apply to all of these sorts of "objects," or only to certain ones? Finally, what does it mean to say that the object of desire or the aim of action must present itself to the agent "under the guise of the good"? Does it mean that the agent must *believe* that the object is good, or that there is something good about it, or that it would be good if …? Or is the relevant attitude something altogether other than belief?

We will say something about each of these questions in due course. For the moment, our purpose is simply to point them out, in order to forestall a too-hasty assessment of the thesis. Many discussions of the thesis, both by its defenders and by its critics, proceed by marshalling intuitions about cases. But if the meaning of the thesis itself is not obvious, neither can we treat it as obvious at the outset what sorts of cases would speak for or against it. Indeed, the idea that the thesis admits of a direct assessment by reference to examples rests on a dubious assumption about the kind of claim that is at issue.

To see this, it will help to consider a case in point. As we mentioned earlier, much of the debate surrounding the guise of the good thesis focuses on certain sorts of apparent counterexamples: for instance, cases in which an agent does something, or desires to do it, out of sheer spite or malice. Discussing such cases, Michael Stocker observes that

> When we feel furious, hurt, envious, jealous, threatened, frustrated, abandoned, endangered, rejected, and so on, what we often seek is precisely the harm or destruction of someone, and not always the "offending party": "If I can't have her, no one will." "So, you are leaving me after all I have done for you. Well then, take that." "You stole her from me, now it's my turn to get even." "The whole day has gone so badly, I might as well complete it by ruining the little I did accomplish." … Given such moods and circumstances, harming another can be the proper and direct object of attraction. (1979, p. 748)

Each of these lines of imagined dialogue vividly evokes a sort of situation in which a person wants to cause harm, and Stocker persuasively argues that the harm caused is not necessarily regarded as itself something good, or as tending to produce some further consequence that is

---

[6] For the former sort of formulation, see the passage from *De Anima* quoted in the epigraph above. For the latter formulation, see *Nicomachean Ethics*, I.1.

good. But does this disprove the claim that the object of desire is always wanted under the guise of the good? The answer depends crucially on what should count as the object of desire. Stocker assumes that "the object of desire," in the sense relevant to the guise of the good thesis, must be some end *achieved* by acting – someone's being harmed, or some further consequence resulting from someone's being harmed – rather than the desired action itself – my "getting even," in one way or another, for some perceived wrong done to me. If the latter is the object of desire, then the idea that the relevant desires represent their objects as good is no less plausible than the claim that to regard something as an act of "evening the score" is to regard it as good in at least one respect.[7]

Our purpose for the moment is not to give a detailed defense of this approach to Stocker's cases, but simply to bring out how supposed counterexamples to the guise of the good thesis can raise questions of interpretation whose resolution is no simpler than, and indeed closely intertwined with, the assessment of the thesis itself. As the literature on such cases amply illustrates, there is typically room for controversy about how to describe any given class of examples – about what to call the object of an agent's desire, how to characterize his attraction to it, and consequently, whether to regard the cases as genuine counterexamples. The fact that such controversies break out is not just a reflection of philosophers' limitless appetite for dispute; it reflects the fact that the concepts employed in describing the examples are the very concepts whose proper employment the guise of the good thesis seeks to characterize. To suppose that the thesis admits of direct refutation by counterexample must presumably involve supposing that it asserts a necessary connection between two concepts – desiring and regarding-as-good – each of whose principles of application are clear enough. For if this were not assumed, how could we be confident in any

---

[7] Admittedly, in some of Stocker's examples, the party with whom the agent seeks to "get even" is not a party who could reasonably be regarded as having given offense. Indeed, in some cases, it is not a person at all, but a thing (the world, the day) on which the agent acts out a fantasy of revenge. These sorts of knowingly unreasonable or fantastical actions raise interesting problems for action theory, but they do not interfere with the point that, inasmuch as the action is wanted as a way (however unreasonable or fanciful) of getting even, it is wanted under an aspect of the good. Such desires, even the irrational ones, seem to be primitive manifestations of the desire for justice, and surely this sort of desire cannot be assumed *not* to be a desire for a certain form of good.

This is not to deny that there can be desires to harm which are not retributive, perhaps desires which are purely sadistic. For *these* sorts of desires, however, it would be much less plausible to claim that the agent can simply want to do harm without wanting thereby to achieve some further end (getting pleasure, exercising power over others, etc.).

given instance that the case we were considering was a clear-cut case of desiring which was clearly not a case of regarding-as-good? But the guise of the good thesis is surely intended as a contribution to the clarification of each of these concepts: it aims to say something about what each is. This does not imply that the thesis is correct in what it says, but it does suggest that the thesis must be assessed, not simply by appeal to cases, but by a systematic investigation of the importance of the concepts it links – of the role, if any, that each should play in a sound theory of action, and of the role of our power of reason in giving rise to action. In the absence of such an investigation, what we will be tallying will be, at best, people's intuitions about when to use the English words "desire" and "good," and we will lack any principled basis for deciding which intuitions to trust.

We have been making these points with reference to authors who attempt to refute the guise of the good thesis by counterexample, but related criticisms apply to authors who attempt to defend the thesis by appealing to intuitions about what desiring is. According to T. M. Scanlon, for instance, the idea of a mere disposition to act, where such a disposition lacks any evaluative component, "does not in fact fit very well with what we ordinarily mean by desire" since "desiring something involves having a tendency to see something good or desirable about it."[8] To support this contention, Scanlon cites Warren Quinn's well-known example of a man who finds himself with an unaccountable impulse to turn on every radio he sees: such dispositions may be conceivable, but, Scanlon observes, their presence seems not to rationalize the actions they induce, and we are not inclined to call them "desires," or anyway not desires of the ordinary kind. There must be – Scanlon concludes – some connection between desire and positive evaluation, or at least between desire and the tendency to think that there is a "justifying reason" to pursue the desired object.

But even if this conclusion is sound – as we believe it is – we should want to be able to support it, not just by provoking intuitions about what desire *is*, but by arguing that this is what desire *must be*. After all, when we say that what we ordinarily mean by "desire" is a state that involves a tendency to regard the desired object as good, this is presumably not supposed to be merely a claim about what, as it happens, English-speakers are willing to call "desire": it is supposed to characterize a real, integral phenomenon that we pick out using this

[8] T. M. Scanlon, *What We Owe to Each Other* (1998), p. 38.

term. Moreover, the claim is presumably not that, as a matter of observed fact, there is a lawlike relation between two distinct kinds of psychological state, desiring and regarding-as-good. If the claim that desiring something involves having a tendency to see it as good were intended as this sort of empirical claim, then introspection and armchair reflection ought to give way to experiments and controlled studies. But this is clearly not the sort of validation that most advocates of the guise of the good thesis take it to require. The thesis is meant to state some sort of *necessary* truth about desire, something knowable on the basis of philosophical reflection rather than by observation and experiment. If this is the character of the thesis, however, it seems that it should be supported by principled argument, not just by intuitions about cases.

Defenders of the guise of the good thesis should also aim to give a principled defense of the thesis for another reason. For, however persuaded we are of the truth of the thesis, we will not really know, in the absence of a systematic investigation of the concepts it employs, what the thesis means. Our intuitive notion of "regarding as good" collects together diverse modes of consideration: a thing can be regarded as good, seemingly, in being regarded as enjoyable, conducive to health or well-being, profitable to oneself, just or fair, required by duty, expressive of friendship or kindness, etc. What is the principle that governs the inclusion of items on this list? Lacking such a principle, we will not be in a position to make well-grounded judgments about whether any given class of desires should or should not count as involving a presentation of something as good. But only an investigation of the systematic importance of the concepts of a good and of regarding something as good could supply us with such a principle.

Our project in what follows will be to mount a defense of the guise of the good thesis that brings out why the thesis *matters* by showing its place in a systematic account of the role of reason in action. We shall be arguing, in effect, that to understand what it is to desire something, in the sense that is relevant to the explanation of intentional action, we must understand desiring as bound together with representing-as-good; and we will be defending this claim, not as an observation about how we use the word "desire" or as an empirical law of human psychology, but as a necessary truth about a concept of central theoretical importance. But this need not commit us to denying that people sometimes desire things which in their considered opinion have nothing good about them. What we will deny is that it is possible to

say what sort of thing such perverse desires are perverse cases *of* without appealing to the idea of a capacity to represent as good. This connection, we shall argue, is what anchors our understanding of the very concept of desire as it applies to rational creatures.

3. The causal theory of action and the object of desire

To make a case for these connections, it is first necessary to contest a view about action and its explanation that is presupposed in much contemporary work in action theory. We believe that it is fundamentally this view, rather than the appeal of particular counterexamples, that underlies most contemporary opposition to the guise of the good thesis. In this section, we first sketch the view in question and then raise a difficulty for it. The view derives much of its appeal from its apparent inevitability. In pointing out a challenge it faces, we hope to show that it embodies a quite specific and questionable view about the shape a philosophical account of action must take.

The problem of action and the causal theory. To introduce the standard approach, and to bring out what can make it seem inevitable, it is useful to recall a common way of framing the question that a philosophical theory of action must answer. As David Velleman observes, philosophers of action tend to introduce their topic by quoting "a bit of Wittgensteinian arithmetic": "What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?"[9] Whatever Wittgenstein himself thought of the matter, contemporary philosophers who quote him typically assume that there should be a solution to this equation: a bodily intentional action should turn out to consist of a not-intrinsically-intentional bodily movement occurring in a context where certain further facts obtain. This much seems inevitable: for if I intentionally raise my arm, then certainly my arm rises; but not every arm-rising is an intentional arm-raising; so it seems that an intentional arm-raising must be an arm-rising about which certain further facts are true. And if this is right, then it seems that the task of the philosophy of action must be to specify which further facts are relevant.

Once this framework is in place, however, two points about the content of this specification will seem evident: first, that the facts in question must include facts about the

---

[9] Velleman 2000b, p. 1. Compare *Philosophical Investigations* (1972), §621.

causes of the relevant bodily movement; and secondly, that these causes must involve *mental states* of or *mental events* occurring in the acting subject. These assumptions constitute the framework within which the standard "causal theory of action" is elaborated. A bodily intentional action, it is held, consists in (1) a bodily movement (2) caused in some "right way" by (3) mental states or events of certain specific sorts. Different authors give different accounts of the sorts of mental states or events that are relevant, and there is debate over how to characterize the causal relation that must connect these with the resulting movement, but this general structure is common ground for most mainstream action theory.

Now, it is no surprise that the guise of the good thesis should look unmotivated to philosophers who conceive of action in this way. The whole point of the causal theory is to make the notion of an intentional action intelligible by showing that it just amounts to a bodily movement of an unproblematic sort with certain specific causal antecedents. If that is what an intentional action is, however, then the requirement that a representation of something as good should figure among these antecedents looks like a superfluous constraint on the mechanism. No doubt the agent who acts intentionally must act with the intent of doing something, and so a representation of *what he intends to do* must presumably figure among the causes of the movements he makes; but there is no obvious reason why the guidance of movement by such a representation need involve the presence of a *further* representation of that doing, or something achieved by that doing, as good. Any such constraint would seem to be a merely stipulative restriction on what we will count as "an intentional action," not a requirement that flows from the sheer idea of movement guided by thought – at least not if the causal theorist is right about what the guidance of movement by thought amounts to.

There might seem to be the possibility of motivating the guise of the good thesis by arguing that an agent's representation of what he intends to do must itself involve a representation of that action as good. For it is tempting to suppose that a conative attitude like wanting or intending must represent a certain action as *to be done*, and then to infer that this must involve representing it as *good* to do (or as something that ought to be done). But the causal theorist's analysis of the situation will not support this reading of the phrase "to be done." Insofar as it is right to say that an agent who wants or intends to do something must represent a certain action as "to be done," all this need mean, according to the causal theorist,

is that he represents it as what will transpire if his want or intention is fulfilled. To say that such attitudes represent their objects as "to be done," that is, is simply to mark something about the "direction of fit" of these attitudes: they represent a certain state of affairs, not as already obtaining, but as what shall come to obtain if the relevant attitude is satisfied. But it is not obvious why this must entail any positive evaluation of the state of affairs in question: to suppose that it must is to confuse a "to be done" whose modal counterpart is "shall" with a "to be done" whose modal counterpart is "should."[10] In the context of the causal theorist's general understanding of what action is, then, the idea that conative attitudes must represent their objects as good will seem to be merely an additional restriction on the content of these attitudes, and one without evident motivation.

A defense of the guise of the good thesis, then, must begin by questioning the acceptability of the causal theorist's story about what action is. That is our aim in the remainder of this section: we shall argue that the causal theorist's project of explaining what it is for a bodily movement to be an intentional action by appeal to its psychological causes faces a basic difficulty, since the basic sort of psychological cause that such a theory must posit, a desire *to do something*, can itself be explained only by appeal to the notion of intentional action. It will emerge that the object of a desire to do is characterized by a distinctive kind of teleological organization, an organization which the causal theory cannot explain but must rather presuppose. It will be through seeking to understand this organization that we will, in the end, arrive at a rationale for the guise of the good thesis.

The immediate object of desire. To bring out the difficulty facing the causal theory, it is necessary to consider more carefully what sorts of representations must figure among the causes of a bodily movement if that movement is to constitute an intentional action. There is considerable controversy among causal theorists about exactly what sorts of mental items must play a role here: whether all that is needed is a desire to achieve some end and a belief about how to attain it, or whether there must be further causal factors in play: intentions, plans, self-referential beliefs about the causes of my movements, etc. The difficulty we wish to raise, however, concerns a minimal commitment that any plausible version of the causal

[10] This, in effect, is the diagnosis of the appeal of the guise of the good thesis suggested in Velleman 2000c – although stating it this way glosses over certain complexities in Velleman's account.

theory must make, namely the commitment that the causal antecedents of my intentionally doing *A* must include a desire *to do A*. It is hard to see how this could be denied: no doubt a person will often be moved to do *A* by further aims which are served *by* his doing *A*, but if he does *A*, and does so intentionally, then it seems that *one* thing that must move him is some motivating representation of doing *A* itself. Otherwise, whatever movements he makes will not themselves by guided by thought in the way that the adverb "intentionally" demands: they may be movements that are caused by a representation of some further end, but they will not themselves be realizations of any aim of the agent.

We have called this motivating representation a desire (and we will follow the usual practice of speaking indifferently of what the agent "desires" and what he "wants"), but for our purposes here, little hangs on this designation. All that is necessary is that the attitude in question should be a representation of what is to be done that plays a causal role in bringing the agent to do that very thing. We refer to this representation as a "desire" in deference to a long tradition of using this term to name the condition which produces animate movement,[11] but if some theorist wishes to insist that the attitude that moves us to act must be something else, we need not quarrel with this: the points we will make turn wholly on the *object* of the relevant attitude, the thing that we normally express with the infinitive phrase "to do *A*." The question we wish to press is whether the causal theorist is entitled to take attitudes toward this sort of object for granted in stating his theory; and if he is not, how his theory might do without them.

It is a striking fact about standard expositions of the causal theory that they tend to insist on rewriting desire-ascriptions which we would colloquially express by saying

(1) *S* wants to do *A*.

by transforming what follows "wants" into a proposition, as in

(2) *S* wants that *S* does *A*.[12]

---

[11] See, e.g., the quotation from Aristotle that appears as our epigraph.

[12] The assumption is often not made explicit, though it comes out in the widespread use of a schematic "*p*" to represent the object of desire. One author who gives explicit attention to the point is Alvin Goldman:

> My analysis of intentional action will make use of a certain species of wanting – viz., wanting to do certain acts. Such wants are not essentially different from other wants, like wanting to possess certain objects. Wanting an automobile consists (roughly) in feeling favorably toward the prospect of owning an automobile. Wanting to take a walk consists (roughly) in feeling favorably toward the prospect of one's taking a walk. (1970, pp. 49-50)

This insistence on treating desire as taking a propositional object is of course not new to the causal theory: the idea that desire is a "propositional attitude" goes back at least to Russell.[13] But this view of desire has a special attraction in the context of the causal theorist's project: it helps to hold in place the idea that the satisfaction-conditions of desire do not raise any special problems over and above those raised by propositional attitudes in general. What is it for a desire to be satisfied? It is – causal theorists characteristically answer – simply for the condition it sets on the world to have come true, for a certain fact to have come to obtain. If desire can be treated in this way, as having for its object what Michael Smith calls a "way the world could be,"[14] then the idea that desire represents its object as *to be done*, whereas belief represents its object as *the case*, is to be explained, not by appeal to a distinction between two fundamentally different sorts of representational content, but rather by reference to the different causal relations in which representational states with the same sort of content normally or properly stand. And this is how causal theorists characteristically explain the distinction. Smith provides a helpfully explicit statement of this idea:

> the difference between beliefs and desires in terms of direction of fit comes down to a difference between the counterfactual dependence of a belief and a desire that *p* on a perception that *not p*: roughly, a belief that *p* is a state that tends to go out of existence in the presence of a perception that *not p*, whereas a desire that *p* is a state that tends to endure, disposing the subject to bring it about that *p*.[15]

Other causal theorists would explain the "direction of fit" of desire somewhat differently, but the general approach pursued here seems to be crucial to the theory: for it is crucial to the theory that what it is for an agent to perform an intentional action should be explicable in terms of concepts independent of the concept of intentional action itself. In particular, a causal theorist must suppose, on pain of circularity, that the concept of desire (or whatever representational state is supposed to guide action) need not itself be explained by

---

Notice how, in the last sentence of this remark, wanting *to take a walk* becomes wanting *one's taking a walk*, which includes a subject term. Since Goldman elsewhere uses "*p*" as his schema for the object of desire, he presumably intends that the complex noun phrase "one's taking a walk" be transformable, in turn, into a proposition.

[13] See Bertrand Russell, *An Inquiry into Meaning and Truth*: "We pass next to the analysis of 'propositional attitudes', i.e., believing, desiring, doubting, etc., that so-and-so is the case" (1992, p. 21).

[14] Cp. Smith 2004, p. 165.

[15] Smith 1987, pp. 53-54.

appeal to the concept of intentional action. But then the *object* of desire must presumably be some outcome of an unproblematic kind, some recognizable state of affairs, and the fact that this outcome is achieved *through intentional action* must be explained in terms of a certain special sort of causal process having contributed to its coming about. The assumption that "to be doneness" can be factored out of the object of desire, so to speak, and represented as part of way in which the attitude in question relates to its object, as part of its "direction of fit," is thus crucial to the project. For if this were not possible – if the kind of desire that gives rise to action were itself explicable only as a desire to act intentionally in a certain way – then the explanatory program of the causal theory would be compromised. We would not be able to say what an intentional action is by appeal to the idea of a not-intrinsically-intentional bodily movement plus certain psychological causes operating in the right way, for the nature of the relevant psychological causes could not be explained without appeal to the notion of intentional action itself.

The idea that the object of desire is a certain state of affairs or proposition, the nature of whose obtaining can be treated as unproblematic, is thus a crucial prop to the causal theory. But what is the state of affairs toward which a desire of form (1) is directed? On the face of it, such desires do not appear to take a propositional object: what follows the attitude verb is not a whole proposition but merely a certain sort of verbal predicate, and one that appears in a curiously infinitival form. And although grammar may permit us to transform this object (awkwardly) into apparently propositional structures like the one in (2), problems begin to arise as soon as we ask when the condition that *S does A* counts as fulfilled.[16] For what condition on the world is set by the freestanding English sentence "*S* does *A*"? When does this "state of affairs" obtain? At any given time, a particular action will be either still underway, in which case we describe it in the progressive, "*S* is *A*-ing," or already complete, in which case we describe it in the perfect, "*S* has *A*-ed."[17] By using a verb in the infinitive to

---

[16] In offering (2) as the proposition-directed counterpart of (1), we are following a widespread practice, but actually we feel unsure exactly what English grammar requires in the propositional complement here. Perhaps it should be in the subjunctive, as in

(2') *S* wants that *S* do *A*.

But whatever form of the verb "to do" appears linking "*S*" and "*A*", the question will remain: when does this supposed "state of affairs" obtain?

[17] Or in the simple past: "*S A*-ed." In either the perfect or the simple past, the verb phrase expresses *perfective aspect*: it represents the event in question as completed rather than underway. For convenience, we will focus on

express what is wanted, desire-ascriptions of form (1) escape the problem of specifying which of these alternatives is in question; but we cannot construct a complete proposition setting a definite truth-condition without making a choice.[18] Should we, then, say that the propositional correlate of (3) is

(3) *S* wants that *S* will have *A*-ed ?

This seems wrong inasmuch as it does not capture the desire *to bring the relevant state of affairs about* that is a part of (1). Would my desire to go to the store count as fulfilled if I were blown there by a powerful wind or miraculously transported there by God? The relevant want might lapse, of course, if my only reason for wanting to go to the store was that I wanted to have got there. Still, it seems that my original want would not have been *fulfilled* unless my having arrived was my own doing. But on the other hand, (1) certainly does not merely amount to

(4) *S* wants that *S* is *A*-ing

for this would be compatible with *S*'s being indifferent about whether he actually completes the relevant action.

Nor, again, does it help to conjoin the two propositions to form

(5) *S* wants that *S* is *A*-ing and *S* will have *A*-ed

for the want involved in (1) is not just to have been in two unrelated states, but to have arrived in the end state in virtue of having done all of the *A*-ing that was required to *A*. But if we opt for something like

(6) *S* wants that *S* is *A*-ing as much as necessary in order that *S* will have *A*-ed

then it seems that such understanding as we have of the contained proposition here depends on our knowing that (6) is supposed to be equivalent to (1). For how much *A*-ing is necessary? Not much if I am blown by a wind, but that is plainly not what is intended: the point is that I should do enough of it to satisfy the want *to do A* – where this means, presumably to effect the condition of my having done *A* through a process of intentional

---

the contrast between the progressive and the perfect, but it is really this contrast between forms that express imperfective aspect and forms that express perfective aspect that is crucial to our argument. For further discussion of this distinction and its relation to English verb forms, see Comrie, *Aspect* (1976).

[18] It would not help to suggest that the agent wants that there be a certain event, as in

(2'') S wants that (∃*e*)(*e* is a doing of *A* & *e* is by *S*).

For when "is" there such an event? Does this require the truth of "*S* is doing *A*", or "*S* did *A*", or what? Each

action.

This is a long-winded way of bringing out something about the object of a desire to do *A* which is perhaps obvious, but which presentations of the causal theory tend to ignore: namely, that the object wanted is not some final state of affairs which might be brought about either intentionally or non-intentionally, but rather an object which "exists" only insofar as a certain intentional action has been carried through to its completion. This makes difficulties for any attempt to explain such wanting in the causal theorist's way, by appeal to the generic notion of representation of a "way the world could be" and the generic idea of a causal tendency. For to want *to do A* is not merely to want to be in some terminal state. It is, as we have seen, not to want to be in any mere state, but rather to want what is essentially a goal-directed course of action: "enough *A*-ing to have *A*-ed," as we were led to put it. And equally, to want to do *A* is not merely to want that the relevant course of action should occur just anyhow. It is, as we have seen, to want to be oneself the source of the relevant action: to want it to be *one's own doing*. When I want to do *A*, in short, the content of my want is of a form such that the world can only come to conform to that content insofar as it not only comes *to be* a certain way, but does so as the outcome of a goal-directed process guided by the agent. Indeed, even this way of putting the matter leaves the outcome and the process too external to one another: to represent my doing *A* is to represent, as it were, a kind of state of affairs whose obtaining *is* my having intentionally caused it to be.[19]

These observations do not constitute an objection to just any philosopher who asserts that an intentional action is a bodily movement with certain psychological causes, but they do present difficulties for philosophers who put forward this proposition in a reductive spirit, as a step towards an analysis of the concept of intentional action in terms of concepts better understood – philosophers who seek, by means of this connection, to give a substantive

---

answers would set off a chain of difficulties similar to the one we describe below.

[19] Certain sophisticated versions of the causal theory attempt to capture something like this point by making the content of the motivating want or intention self-referential: by stipulating, e.g., that I do *A* intentionally only if I am moved to do *A* by a desire that *this very desire* should be the cause of my doing *A* (compare Harman 1976, Searle 1983, Ch. 3, Velleman 1989, Setiya 2003). We believe that this sort of maneuver would fall prey to a version of the difficulty we have raised, since the phrase "doing *A*" remains in the content clause, and this sort of content is one to which the causal theorist is not entitled to appeal. Our aim here, however, is simply to show how the presence of this phrase makes a difficulty for the most straightforward and natural version of the causal theory. More complicated versions of the theory make the difficulty harder to detect, but we believe that they do not ultimately eliminate it.

answer to Wittgenstein's question.[20] For they show that the relevant psychological causes must include representations whose nature can be explained only by appeal to the notion of intentional action itself: representations which represent their object as "to be done" in a sense which must mean precisely "to be done intentionally."[21] But if we cannot hope to give an account of wanting to do *A* that is independent of the idea of intentional action, then neither can we hope to give a reductive account of intentional action as a matter of movement caused in the right way by such wantings. An account of what it is to want to do *A* must rather presuppose an account of action, an account of the kind of event (*S*'s doing *A*) whose coming to be is the subject's intentionally causing it to be.[22]

4. Action, Form, and Goodness

We can hardly claim to have shown that causal theorists have no room for maneuver in responding to this difficulty, but we hope at least to have brought out a significant commitment that such theorists must undertake: they are committed to treating the sort of object at which an agent primarily aims, a *doing of A*, not as an unproblematic given, but as something itself requiring analysis. If the foregoing considerations are sound, constructing an account of action that eliminates reference to this sort of object is more difficult than is commonly supposed. The ease with which we pass from "my raising my arm" to "my arm's rising" can encourage us to think that we can hang onto the sort of thing whose coming to exist would satisfy an agent's aim while subtracting its intentionality; but in fact, as we have

---

[20] Some philosophers who would regard themselves as advocates of a causal theory of action do not have this ambition: Donald Davidson (1980b, 1980c) and Jennifer Hornsby (1980, 1995), for instance, argue that the intentional actions must have certain psychological causes, but do not aim for a reductive account of intentional action in these terms. What we have been saying does not constitute an objection to such views, though we believe that the kind of investigation of action we go on to propose in the following sections does, if it is sound, raise questions about the sort of theorizing about action that these authors undertake. For the present, however, our interest is in the kind of causal theory that aims at reduction. For it is the assumption that *this* sort of causal theory is possible which supports the widespread conviction that *doing something intentionally* is not a special and irreducible kind of event or process, but an event or process of some more generic kind with certain special causes. This is the conviction we ultimately aim to challenge.

[21] Difficulties would also arise, we believe, if we investigated the notion of *cause* on which such an account must rely. It is widely recognized that not just any sort of causal relation will do: the cause must operate "in the right way." Davidson famously doubted whether there could be a noncircular account of what "the right way" must be (1980b, p. 79). Contemporary causal theorists tend to suppose that Davidson was wrong about this. The foregoing considerations are an attempt to show, from a different angle, why he was right to doubt the possibility of reduction.

[22] For a more systematic argument for this conclusion, to which the present discussion is much indebted, see

seen, the sort of object that an agent primarily aims to realize is precisely: an action consisting of phases ordered intentionally toward a certain end (his having *A*-ed). No doubt it is true that, where such an object has come to exist, it will be possible to describe the movements that have occurred in terms which do not imply that an intention has been realized therein. But it does not follow that it must be possible to give an account of what it consists in for an intentional action to have occurred employing only such terms; and indeed, we have seen at least *prima facie* reason to suppose that this is *not* possible.

The importance of this result, for our purposes, is that it forces us to face anew the question what intentional action is. The upshot of the last section is that we cannot explain the nature of this sort of event or process – an agent's intentionally doing *A* – by appeal to the idea of some unproblematic kind of happening with certain specific psychological causes. Rather, we must explain the nature of the relevant psychological causes by appeal to their directedness toward precisely this sort of process or event. But this simply returns us to the question how to characterize the sort of process or event in question. The schema "*S* is (intentionally) doing *A*" marks the object of our interest, but invoking it does not by itself clarify how expressions of this form function, or what sort of conceptual surrounding they need to get a grip.

Our aim in the remainder of the paper is to argue for a broadly Aristotelian view of the required surrounding: a view on which the general notion of a goal-directed action is explained by linking it with the idea of the *form* of the subject that acts, and the more specific notion of an intentional action must be explained by linking it with the idea of a form that essentially involves the capacity for self-consciousness. This amounts to a fundamentally different sort of approach to understanding action from the sort pursued by causal theorists: the aim is not to specify conditions under which some not-intrinsically-intentional process amounts to an intentional action, but rather to explain intentional action as an irreducibly distinctive *type* of process, one that is to be characterized by bringing out the specific implications involved in positing a process of this kind, and by clarifying what *other* sorts of propositions must also be true of something that can be the subject of such processes. The guise of the good thesis will turn out to be a commitment that follows from this Aristotelian approach to saying what action is.

Michael Thompson, *Life and Action* (2008), Part II.

To bring out the attractions of this approach, it will be useful first to reflect on how to characterize the wider genus of which intentional action is a species. Contemporary discussions of the guise of the good thesis often treat it as an isolated claim about rational, intentional action. If that were right, the thesis would look implausible from the start. For clearly not all action is rational action. Animals and infants lack the powers distinctive of a rational agent: they cannot deliberate about reasons, form prior intentions, or reflect self-consciously on what they are doing. But although they are not *rational* agents, it would be perverse to deny that they are agents in some sense: not just the passive subjects of various events and processes, but the active source of certain happenings in which they are involved. When the dog *runs excitedly to the door* or the infant child *turns toward the sound of its mother's voice*, these are clearly self-originated, goal-directed actions of a sort. If they are actions, however, and if "action" here does not mean something utterly different from what it means in the rational case, then it seems that an account of rational action should conform to a shape that applies, at least in its general outlines, to the nonrational case as well. But if nonrational action can be accounted for without any reference to the notion of goodness, why must reference to this notion suddenly appear in the rational case?

This line of thought suggests that a principled defense of the guise of the good thesis should begin with a defense of a more basic principle, one that posits a connection between goodness and action-in-general. And this, in fact, is how we will argue for the thesis: by showing that it is the product of (i) a general point about the connection between the idea of action and the idea of the good of a thing and (ii) a special point about the shape this connection takes in the rational case. Briefly, the thought is this. The most general idea of an *action* is the idea of a movement or change which in some sense comes "from the subject," rather than merely being the result of forces acting on the subject "from without." But this distinction between *self*-movement or *self*-change, on the one hand, and movement or change whose cause is external, on the other, must be drawn against the background of an idea of form which brings with it a standard of goodness. This connection holds for action in general, but in the rational case it takes a particular shape. For to be a rational creature means just this: to live by *thought*, which is to say by the employment of concepts. Hence the shape which the general connection between action and goodness takes in a rational creature will be one that involves thought: rational action is a kind of movement that has its source in a subject's

power to bring things under the concept *good*. The remaining pages attempt to fill in the picture just sketched. The rest of this section discusses the first part of the argument: the general connection between self-movement and goodness. The next section discusses the special form this connection takes in the rational case.

Characterizing self-movement. There is broad agreement among contemporary philosophers of action that autonomous, rational action is not just an isolated topic, but one species of a wider genus that also includes more primitive forms of goal-directed activity. But how can we define the wider category? What *is* "goal-directed activity," and how is it distinguished from other kinds of worldly happening?

It will not do simply to say that a goal-directed activity is anything that can fill the "*A*" position in the schema

(7) *S* is doing *A*.

It is true that such propositions in general describe the here and now by relating it to a possible future situation: to assert a proposition of form (7) is to posit an outcome toward which *S* is tending: namely, its having done *A*. And by the same token, to assert (7) is to leave logical space for the possibility of failure: it can be true that *S* was doing *A* but never did *A*. But not every instance of this schema is goal-directed in the interesting sense. The bare schema "*S* is doing *A*" admits of such substitutions as "The tree is falling over," "The smoke is rising to the ceiling," "The tidal wave is rushing toward the shore"; and these are clearly not instances of genuinely goal-directed activity. To count as genuinely goal-directed, one wants to say, a movement or change must in some sense *come from* the subject, and must be *for the sake of* the end. This, indeed, is a traditional way of defining an agent: agents are self-movers, things that themselves pursue ends. But when does a movement count as a self-movement, or as done for the sake of an end?

Having rejected the sort of approach that seeks to specify when a movement counts as an intentional action by appeal to its psychological causes, and having in any case widened our focus from intentional action in particular to goal-directed activity generally, we must seek a characterization of self-movement that does not appeal to specific causes, but rather identifies self-movement as a distinctive *type* of event or process. We must, that is, identify a kind of event-description that is as such a description of a movement by a subject for the sake

of an end, a type of proposition of form (7) from which we can correctly infer *S is doing A goal-directedly*. But what role does the proposition *S is doing A* have to be playing in order for this adverbial attachment to be appropriate?

We can make progress on this question by observing that the truth of some but not all progressive propositions can be explained by adverting to further, more embracing progressive propositions with the same subject. Thus we can offer such explanations as

Why is the plant budding?

—Because it is growing a new leaf.

Why is the cat crouching there waiting?

—Because it is stalking a bird.

Why is he mixing eggs and flour?

—Because he's baking a cake.

But of course we do not suppose that we can explain the rain's falling to the earth by the fact that it is watering the plants, or indeed by reference to any larger purpose belonging to the rain. To suppose that we *could* explain the rain's falling in this way would be to take, as they say, an animistic view of nature. And nor will we suppose that the rain's falling to the earth itself explains the lesser phases in which this process consists: to suppose that the rain is falling past the treetops because it is falling to the earth would be equally animistic. These observations are clues to the solution of our problem: they suggest that a subject *S* is capable of the kind of goal-directed activity of which inanimate nature is incapable just if it is a potential subject of explanations of the form

(8) *S* is doing *A** because *S* is doing *A*

where, intuitively, doing *A** is a way or means or part of doing *A*. We can call such propositions *judgments of individual teleology*, for they explain a lesser activity in which an individual subject is engaged by linking it to a more-encompassing activity in which that same subject is engaged, and thereby represent the accomplishment of the more-encompassing activity as a sort of *end* whose pursuit can explain things done in the service of it.[23]

---

[23] There may be explanations which fit the grammatical pattern of (8) but which do not imply a teleological relationship between the lesser process and the greater one: for instance, "The dryer is shaking because it is running its spin-cycle." If this is a genuine explanation, however, it seems intuitively clear that it is an

This observation suggests a way of stating a condition under which a progressive proposition of form (7) ascribes goal-directed activity to its subject: it does so just if it can figure on the right-hand side of a proposition of form (8) – i.e., if it can figure as the explainer of a less embracing progressive with the same subject.[24] A proposition that can play this role represents its subject as engaged in a process which can explain its own realization, a process that can be the cause of its own coming to be, in whatever sense of cause is implied by the "because" in such explanations. And it seems that for *any* temporally-extended process of *A*-ing, if the process is goal-directed, there will be some true explanation of this form in which the process figures. For if the process of *A*-ing is temporally extended, there will be moments at which *S* is *A*-ing but it is not yet true that *S* has *A*-ed – moments at which the process is underway but not yet complete. But if there is no part or phase of this process, *A**, which is occurring at some such moment and which can be explained in a judgment of form (8), then it seems that there can be no basis for saying that the accumulation of parts or phases toward the completion of the larger process, *S*'s having *A*-ed, is goal-directed: the aim of the whole played no role in the realization of the parts. So it seems that, at least for temporally-extended processes of doing *A*, it is a necessary and sufficient condition for their being goal-directed activities that they should be capable of figuring in true explanations of form (8).[25]

An indication that this characterization of goal-directedness is on the right track is that it gives a clear interpretation to the intuitive ideas that a goal-directed movement must (i) come from the subject and (ii) be made for the sake of the end. For on the one hand, we have

---

explanation of a different kind from the kind offered when a proposition of form (8) is used to mark a teleological connection: the dryer's shaking is not, intuitively speaking, a way or means or part of its completing its spin cycle. One test of whether an explanation of form (8) is genuinely teleological is whether it can be transposed into an explicitly teleological form, as in:

(8') *S* is doing *A** in order to do *A*.

This will be possible at least in the case of animate teleology: there may be reservations about making the transposition in the case of plants. But even there, it seems that such an explicitly teleological description should apply, if not at the level of the individual plant, then at least at the level of the kind to which it belongs:

(8'') *K*s do *A** in order to do *A*. (They grow leaves in order to absorb sunlight, roots to take up water and nutrients from the soil, etc.)

A fuller treatment of these topics would need to investigate the special features of the explanatory structure that exists when a proposition of form (8) is used in the way that we intuitively recognize as teleological. We do not attempt this here, however: our aim is just to sketch a general approach and bring out some of its attractions.

[24] Compare the characterization suggested at Thompson 2008, p. 112.

[25] The consequences of this point for the understanding of intentional action are explored in greater depth in Lavin, "Must There Be Basic Action?" (MS).

identified a kind of explanation that says why a subject is up to one thing by adverting to something else that same subject is up to, rather than by tracing the subject's activity to the influence of some other thing. And on the other hand, this kind of explanation says why the subject is up to something by saying what further end the subject's doing that thing would serve. Moreover, our characterization seems to be extensionally correct. For, in the first place, judgments that ascribe intentional actions meet our criterion: my intentionally doing one thing can explain my intentionally doing another thing. But our criterion is also satisfied by other kinds of progressive judgments: as the previous examples indicate, this general form of explanation can apply to nonrational animals and indeed to plants. Its application marks the feature of living things we are tracking when we say that what goes on with them is subject to teleological explanation.

Our criterion thus captures what is right in the idea that the sphere of the goal-directed is wider than the sphere of the intentional. At the same time, it leaves us with an agenda of problems: to explain, in an equally abstract way, what distinguishes the category of goal-directed progressives that ascribe, not merely the kind of purposiveness that we find in plant life, but the more determinate kind that we find among *animate* creatures, ones that can act in response to the promptings of sensation and appetite; and again, to explain what distinguishes within the latter category the further subcategory of animate goal-directed progressives that ascribe rational, intentional action.

Action and form. For the moment, however, our aim is not to take up these special problems but to make a general point about the conditions under which any judgment implying goal-directedness gets a grip. The general point is this: judgments ascribing goal-directed self-movement or self-change only have application to things to which a certain kind of *form* is attributable, a form that in turn makes room for a notion of what is *good* for things of that kind.

We will try to give a principled account of the connection between goal-directedness and form in a moment, but first let us simply observe that, for many such judgments, the point is obviously true. If we begin by considering, not rational action, but the more rudimentary kinds of goal-directed self-change or self-movement characteristic of plants and animals, it is plain enough that recognizing them involves bringing to bear a conception of the nature of the

kind in question. In recognizing that a certain plant is, e.g., budding – as opposed to, say, developing a cancer – we are relating what is going on with it to a more general conception of how things go in the life of that kind of plant. In recognizing a cat as pursuing a mouse or as fleeing in response to a loud noise, we regard processes in which it is presently engaged as organized by general aims that belong to it as a cat. In these sorts of instances, at least, the general idea of processes in which a certain individual figures as an agent pursuing a goal seems to get a grip only against a certain sort of background: only inasmuch as the individual in question is regarded as an instance of a certain kind of thing, a kind with a certain characteristic form or nature, a kind to which certain ends and activities belong as such.

It is one of Aristotle's characteristic thoughts that ascriptions of self-movement or self-change to individuals presuppose such a background: for he holds that, in general, the topics of movement and change are to be understood against the background of the natures of things, where a nature is a "principle of motion and change," so that movements and changes that do not come about by chance are understood as the realizations of potentialities belonging to the natures of things that move and change; and he holds that the possibility of *self*-movement and *self*-change must be understood against the background of the special sort of form or nature that he calls a "soul."[26] This Aristotelian thought is the inspiration for our account, but our aim here is not to do Aristotle exegesis, but to show why the thought might be attractive in its own right. If ascriptions of self-movement to individuals presuppose a conception of the forms that those individuals bear, why is this so? In the remainder of this section, we will sketch an account of this connection. Our argument will be less than decisive at several points, but it should at least clarify why, given certain general views about the role of kinds in explaining the activities of individuals, the idea of a connection between self-movement, form, and goodness will seem compelling.

The first step is to note the striking and remarkable way in which ascriptions of goal-directed activity link a subject's present to a certain future condition toward which it is tending. It is not merely that, if *S* is doing *A* goal-directedly, and if nothing interferes, then it

---

[26] For the idea of nature as a principle of movement and change, see *Physics*, II-III, and for the connection between nature and form, see especially *Physics* II.1 and II.7. For the idea of the soul as the form of a living thing, see *De Anima* II, in particular II.4 for the idea of the soul as the cause of specifically vital movement and change. See also the discussions of the general connection between nature and teleological (or final) explanation in *Physics* II.8, and of the connection between teleological understanding and the soul in *Parts of Animals* I.1.

will eventually be true that *S* did *A*: that much is also true of "The smoke is rising to the ceiling." What is distinctive of goal-directed progressives is their implication that aspects of *S*'s present activity are happening *because* they tend toward this future situation (*S*'s having done *A*). Now, that should sound strange: how can a situation that does not yet exist explain a situation that does exist? A standard sort of objection to teleological explanation is that it would require some sort of "backward causation" or "pull from the future," and that such notions are unintelligible. But if we are convinced that progressive propositions which imply goal-directedness are sometimes true – and it is hard to see how we could even begin to think about living things and their activities without taking this for granted – then our task must be to understand what it can be about a subject that allows such propositions to be true of it. What can it be about an individual here and now in virtue of which its present activity is explained by an as-yet-unrealized end toward which it is tending?

We can begin to see how this might be made intelligible by noting that goal-directed progressives characterize a subject here and now by relating it, not necessarily to *the actual* future, but rather to *its own* future – to a possible outcome that would count as something the subject itself effected, rather than something that merely happened to it. Part of the point here is not special to goal-directed progressives in particular: in general, a progressive proposition of the form

(7) *S* is doing *A*

is not necessarily falsified because the relevant future state of affairs (*S*'s having done *A*) does not come to obtain; it is falsified only if this was not the state toward which *S* was tending, the state which would have come to obtain had nothing interfered with its activity. In this sense, any progressive proposition of form (7) relates its subject, not to the actual future whatever it may be, but rather to a possible future that would count as the subject's own. Now, the crucial Aristotelian thought is that the distinction between a future that counts as the subject's own and one that does not must be drawn against the background of a conception of *what the subject is* and of what belongs to being that kind of thing – that is, of the form it bears and the nature of things that bear this form. This claim may initially sound dark and metaphysical, but we can bring it down to earth by restating it as a point about the relation between truths of the form we have been considering and truths of certain other characteristic shapes. The thought, in effect, is that where there are truths of the form (7), there must also be true

judgments of form-attribution of the form

(9) *S* is an *F*

and true *form-characterizing judgments* of the form

(10) *F*s do α (in conditions *C*)

where the description of the activity characteristic of the kind, α, need not in general be identical to the description that characterizes what the individual is doing (*A*), although in the simplest sort of case it might be. In the more general case, doing *A* will be some specific form or manifestation of α-ing, as rolling down this hill is a specific manifestation of rolling (*S is rolling down this hill; S is a bronze sphere; Bronze spheres roll (when on uneven ground)*). The relation that must obtain between *A* and some corresponding α would not be easy to specify, but in any case the Aristotelian thought is: there must be one.

Propositions of form (10) are general judgments about *F*s, but they are stated without a quantifier, and are meant to state what are intuitively truths about the nature of *F*s, truths that hold "if nothing interferes" or "other things equal" in the sense that they hold, not necessarily without exception, but rather in such a way that an *F*'s doing α in conditions *C* is to be expected, while an *F*'s *not* doing α in conditions *C* calls for special explanation.[27] The idea is that if propositions of forms (7), (9), and (10) hold, and if *A*-ing is a form or manifestation of α-ing, then it is intelligible how it could already be true to say of *S* that it is headed toward the condition of having *A*-ed, and will reach that condition if nothing interferes. A subject can be *tending* toward a certain result, in the manner of (7), even though that result may not actually come about, precisely because a subject can have certain general *tendencies*, where describing a tendency is describing not how things *will* come out in any given instance but how things *do* come out if nothing interferes. But – the Aristotelian thought goes – general tendencies belong to individuals only inasmuch as those individuals bear some general form: for the very idea of a tendency is the idea of what happens "if nothing interferes," and the distinction between interference and noninterference would have no application to an individual thing if

[27] Such judgments are characteristically expressed using sentences of the type linguists call "generics." For an overview of the characteristic significance of such sentences and the problems they raise, see the Introduction to Carlson and Pelletier 1995. For illuminating discussion of the relevance of generic propositions to the understanding of teleology in living things, see Moravscik 1994 and Thompson 2008, Part I. The more general Aristotelian thought that, to make sense of the notion of an individual substance undergoing motion or change, we must see that substance as belonging to a certain substantial kind characterized by various general ways of

there were not some basis for distinguishing between those episodes in its existence which manifest the sort of thing it is and those which do not. The case where nothing interferes is the case where a thing manifests its own nature, the traits that characterize the sort of thing it is. So in committing ourselves to the distinction between interference and noninterference, we commit ourselves to regarding individuals as bearers of forms, and the forms themselves as characterized by general principles of movement and change, principles which are explanatory, other things equal, of the activities of the individual bearers of the forms in question.

The connections described so far hold not just for individuals that can be the subject of goal-directed progressives, but for any individual that can be the subject of progressive judgments of the form *S is doing A*: this attribution of a process-in-progress is possible only against the background of various general tendencies, and such tendencies can belong to an individual only in virtue of its being a certain kind of thing, the bearer of some general form. The special problem about goal-directed progressives was their implication that the outcome toward which *S* tends in doing *A* is in some sense what explains the parts or phases in which the realization of this tendency consists. The problem of understanding how goal-directed progressives can be true is thus the problem of understanding, not just how a certain future can be a subject's own, but how it can be so in a way that explains the subject's present activity.

Now, if the general problem about how a certain future can be a subject's own is solved by noting that subjects of change belong to kinds characterized by general tendencies to change, we should expect this special problem to be solved by noting that subjects capable of goal-directed activity belong to kinds of a special sort, kinds characterized by tendencies with a distinctive structure. Indeed, the general shape of the required structure follows from the very statement of the problem. How can the outcome toward which *S* is tending explain the occurrence of the process that proceeds towards it, and explain it in such a way that the whole process is intelligible as a case of self-movement or self-change, a process "coming from the subject" in the intuitive sense that we are seeking to explicate? Well, in the first place, the tendencies that characterize the kind to which *S* belongs must be tendencies toward some definite outcome or final condition, not merely tendencies that have no intrinsic limit,

---

moving and changing has been defended in a contemporary context in Wiggins 2001, Chapter 3.

like the tendency of fire to spread or of iron to rust. They must, as we can put it, be tendencies toward a definite *end*. Furthermore, they must be tendencies whose being set in motion is in some sense due to the subject itself. They must, in other words, be tendencies of a special sort, which we can call *powers*, to mark the fact that the normal cause of their coming to operate (or, as we can say, to *act*) is not just some alien force acting on the subject, but a cause whose operation itself expresses the subject's nature.[28]

What this might mean, more concretely, is that the cause which summons a power to act does not depend on conditions whose holding is simply an accident given the nature of the kind of thing in question. That its powers are brought to act must be something for which the kind by its nature provides: if a given power's acting depends on its being in condition *C*, then the kind must have further powers which tend to secure that *C* holds – as always, if nothing interferes. And conversely: if *P* is a power of a kind *K*, then the acts of *P* must themselves be explanatory (other things equal) of *K*s being in a condition in which various of their other powers operate. This is what secures that the end toward which an act tends is explanatory of the processes that contribute to its own realization: for by contributing to the existence of the kind of thing whose nature it is to realize that sort of end, the achievement of that end is, in a sense, *the explanation of itself* – not through some occult form of backwards causation, but by being an effect whose coming to be contributes to the very conditions that make its own coming to be no accident. In short, on this broadly Aristotelian view, a subject that is capable of goal-directed activity must belong to a kind characterized by powers that form a sort of

---

[28] Once again, it will plainly be a complicated matter to say how an instance of goal-directedly *A*-ing must be related to the general description of what a given power is a power to do (α) if the *A*-ing is to instance or belong to α-ing. Indeed, the potential complexity involved here will presumably increase with each step up the "ladder of nature": from plants, to nonrational animals, to rational animals. In the vegetative case, what any individual plant can be said to do goal-directedly will probably cleave pretty closely to what it belongs to the general powers of things of that kind to do. Already in the case of nonrational animals, however, the presence of the power of perception (to say nothing of capacities for learning, adaptation, etc.) will introduce possibilities of goal-directed activity which have no direct counterpart in the kind: for, e.g., in virtue of being able to perceive a certain mouse, an individual cat can get into the act of *hunting that mouse*. Whatever powers of cat-kind this activity instances will certainly not be powers to do *that*: they will presumably involve only a power to *hunt mice* or possibly just to *hunt things* that meet certain general parameters. And the distance between any general description of the power exercised and the appropriate descriptions of the action undertaken will receive another and much more radical multiplication in a rational creature, for an individual rational creature will be in a position to determine the *how* and the *wherefore* of its action in all sorts of ways that are not anticipated in the powers belonging to its kind. Nevertheless, if the Aristotelian view is right, representing a rational creature as engaged in intentional action will involve representing what it is doing as in *some* way connected with the operation of powers characteristic of its kind. We say more to develop and defend this view about rational creatures in the next section. For further discussion, see Boyle, "Rational Animals and Rational Powers" (MS).

self-sustaining system. For it is against this background that the idea of a subject capable of moving or changing *itself* becomes intelligible: acts of self-movement or self-change are understood as movements or changes that are explicable by reference to such powers, and the goal-directedness of such acts consists in the fact that their tending toward a certain end is no accident given the general nature of the kind of subject that undergoes them.

This description of the structure of powers that characterizes a kind capable of goal-directed activity is highly abstract, but it should become more tangible if we reflect on how aptly it describes the sort of order we find in living things. Any given kind of living thing is characterized by a manifold of powers directed toward various ends, powers which constitute a sort of self-maintaining system: one such that the realization of any one of its ends supplies the condition for the realization of various others, and these in turn of others, in such a way that the kind "makes itself exist," so to speak. This self-maintenance occurs at two levels: first, individuals of the kind in question maintain their own existence through such processes as nourishing themselves, defending themselves from threats, healing from injuries, etc.; and secondly, the kind itself maintains its existence through reproductive processes in which members of the kind make others like themselves. For any given kind of living thing, we can describe powers that it has which subserve each of these two sorts of self-maintenance, powers whose various acts contribute to fulfilling the conditions in which life of that kind can continue. And it is characteristic of these powers that they not only contribute variously to the maintenance of the kind of living thing in question, but thereby contribute to the maintenance of themselves and one another in sound order: by seeking out and consuming nourishing food, a creature makes it possible for its injuries to heal; by healing its injuries, it makes it possible to seek out and consume nourishing food, etc. Indeed, this reciprocal interdependence extends to all of the essential powers of a living thing, for precisely insofar as they are essential, they are each needed to contribute to the maintenance of the system of which they are powers, but equally they each depend on all the other powers to operate in a way that maintains that system, and thus makes each power possible.[29]

The foregoing has been a highly programmatic attempt to characterize the background that must be in place if there is to be room for the ascription of goal-directed activity to

[29] The resulting system thus constitutes the kind of unity that Kant called a "natural purpose," which is "the cause and effect of itself." Compare Kant, *Critique of Judgment* (1987), Part II, Division I, esp. §§3-4.

individuals. Our proposals are obviously disputable at many points. Our purpose has merely been to sketch an approach, one that we hope looks recognizably Aristotelian, and that contrasts in important respects with the sort of approach pursued by contemporary causal theorists. Whereas the causal theorist's approach to characterizing action is reductive, this approach might be called holistic: it aims, not to replace talk of intentional processes (or teleological processes more generally) with something less mysterious, but to dispel the air of mystery in a different way, by showing how ascriptions of the relevant processes have a place within an intelligible system of types of proposition – a kind of system that we cannot easily do without if we hope to make sense of living things at all. If we are right that progressive propositions about individuals presuppose general truths of form (10), and that goal-directed progressives, in particular, can apply where such truths characterize a certain sort of self-maintaining system of powers, then some such system must be the background against which goal-directed progressives can be true of individuals at all.

Form and goodness. Our aim was not only to bring out a connection between goal-directed activity and the presence of a certain sort of form, but also to bring out a connection between this sort of form and the application of the concept *good*. We can now finally turn to this latter connection.

It is not a new suggestion that what it is for *S* to be good depends on what kind of thing *S* is. Peter Geach famously suggested that "good" is a "logically attributive adjective" which has sense only in relation to some substantial kind, and which can vary in sense depending on which kind is in question; and a number of other authors have argued that at least certain uses of evaluative language in application to the states and activities of individuals presuppose judgments about the kinds to which they belong and the natures of those kinds.[30] Thus a rosebush is said to be doing poorly if it does not bud and flower in the season proper to rosebushes, and a human being's tooth-development is said to be defective if he does not come to have the normal thirty-two adult teeth. The evident aptness of such evaluations, as far as they go, suggests that at least *some* sorts of evaluations are kind-relative. But it certainly seems a long and dubious way from here to the topic of intentional action; and

[30] See Geach 1956 and Anscombe 1958, and for more recent representatives of this "Aristotelian naturalist" tradition, Foot 2002 and Thompson 2008, Part I.

lacking a principled account of *why* the kind to which individuals belong should constitute a standard for the deeds of those individuals, it is easy to doubt whether a kind-relative notion of goodness has any bearing on the deliberation of a rational agent.

We will consider how these issues bear on the deliberation of a rational agent in the next section. Before turning to that topic, however, we need to say something general about why, if being a goal-directed agent presupposes being the bearer of a form in the sense described above, such forms should equally constitute evaluative standards for the acts of the agents who bear them. The answer will be, in a way, disappointingly quick. It is that this is something we have already conceded in all but name in assigning the notion of form the place we have given it in our account. To represent an individual as the bearer of a form, in the sense we have been specifying, is to represent that individual as a sort of thing that as such pursues certain ends, ends that stand, when things are going well, in a sort of balance or equilibrium, a balance on which the existence of such things depends. To the extent that such a thing achieves those ends, it succeeds in pursuits that belong to it as such. And by the same token, to the extent that it fails, it fails in pursuits that belong to it as such. Inasmuch as the form in question is essential to individuals that bear it, these pursuits belong inalienably to those individuals: they cannot cease to be pursuers of these ends without ceasing to be. And inasmuch as their particular doings are to be understood as acts of powers directed toward certain general ends, these ends will be the measures of those acts, in the way that any act is a success or failure in virtue of its fulfilling or not fulfilling its end. That attributing a form to a thing, in this sense, involves attributing to it something that is a standard or measure of its activity, a standard relative to which it may be acting well or poorly, is thus a truism, not a controversial addition to what has already been said.

A certain standard of goodness for a thing follows inevitably from its belonging to a kind characterized by a functionally-organized system of powers: this, we suppose, is the crux of Aristotle's famous "function argument."[31] If the objection to this is that it illegitimately infers an "ought" from an "is," we are not sure that we understand the charge. The sort of "saying what a thing is" that is at issue here is: ascribing to it a certain form, where a form is something that as such involves directedness toward certain ends. If the question is supposed

[31] See *Nicomachean Ethics* I.7.

to be why the thing at issue ought to pursue those ends, we ask: from what standpoint is this question posed? If the thing in question genuinely is a bearer of such-and-such a form, then it *is* a pursuer of such-and-such ends, and essentially so. *It* can no more renounce these ends than it can cease to be itself. But if the objection is that there can be no such thing as a "form" in the sense that would validate these claims, then we would want to dispute this, though to confront the various challenges to this notion would be too large a task to take on here. We hope the foregoing discussion suggests, at any rate, that the costs of giving up this notion would be significant. For it suggests that the notion belongs, not simply to some strange pre-modern metaphysical outlook, but to a characterization of the underlying structure of forms of thought and speech that we all constantly employ, and whose soundness few philosophers seriously question. If the Aristotelian standpoint on goal-directed activity is right, then to regard something as a goal-directed agent is necessarily to regard it as the bearer of a certain form, and thus as directed toward a certain system of goods, goods the pursuit of which orients, more or less remotely, its various particular doings.

5. Goodness and reason

But what does all this have to do with the claim that we must act under the *guise* of the good? How do the points we have made about the connection between self-movement and goodness underwrite a connection between intentional action and the *representation* of something as good? The connection, we shall suggest, turns on a point about the special form that this connection takes in a creature capable of *reasoning* about how to act.

Human beings are rational creatures; our capacity for action is a rational capacity. What does this mean? Thomas Aquinas explains the point as follows:

> The will is a rational appetite. Now every appetite is only of something good. The reason of this is that the appetite is nothing else than an inclination of a person desirous of a thing towards that thing. Now every inclination is to something like and suitable to the thing inclined… But it must be noted that, since every inclination results from a form, the natural appetite results from a form existing in the nature of things, while the sensitive appetite, as also the intellective or rational appetite, which we call the will, follows from an apprehended form. Therefore, just as the natural appetite tends to good existing in a thing; so the animal or voluntary appetite tends to a good which is apprehended. Consequently, in order that the will tend to anything, it is

> requisite, not that this be good in very truth, but that it be apprehended as good.[32]

This passage is useful for our purposes because it argues for the guise of the good thesis in just the way that we have been seeking to argue for it: by deriving it from a more general thought about the connection between goodness and the inclination to do something, together with a thought about the special character of this connection in the rational case. The focus of this last section will be on this connection between the thought that we possess a faculty of *rational* inclination and the claim that we pursue the *apprehended* or *apparent* good.

Kinds of inclination. Aquinas says that whereas the "natural appetite" of nonsentient agents like plants results from "a form existing in the nature of things," and tends toward a good that exists in them *simpliciter*, both the "sensitive appetite" of nonrational animals and the "rational appetite" of human beings follow, in different ways, from an "apprehended form" and tend toward "a good which is apprehended." His point can be expressed as follows. A plant just belongs to a certain living kind, a kind that has a particular way of inhabiting its environment, taking in nourishment, making possible its own growth and reproduction, and so on. In representing a certain plant as engaged in some teleologically-structured process (growing a new leaf, taking in nourishment from the soil, etc.), we see what is happening with it as oriented toward obtaining something suitable to that kind of plant. We thus represent the prospect of obtaining the relevant good as in a certain way already active in what the plant is doing here and now, even though the good in question has yet to be obtained: the plant's present activity is already informed by its natural tendency, as a plant of such-and-such a kind, to pursue certain goods. Or speaking in Aquinas's way: the form of the relevant good already exists in the plant's nature, as something toward which it is naturally inclined.

A nonrational animal too acts in fulfillment of various teleologically-oriented inclinations, but it stands in a different kind of relationship to these inclinations in virtue of having the powers of sensation and appetite. It does not just inhabit an environment; it can perceive its environment and react to it. It does not just take in nourishment; it can seek it out. In general, animals do not just *have* inclinations to pursue certain goods as a part of their nature; their nature involves their *apprehending* particular things that are good for them

---

[32] *Summa Theologica*, IAIIae.q8.a1.

(which is not to say: apprehending *that particular things are good for them*) and pursuing those apprehended things. So those things toward which an animal is inclined – the forms that its inclinations seek to realize – are not just fixed by its nature but are present to it in virtue of particular perceptible things having made an impression on it. In this sense, its way of life essentially involves apprehension: it pursues apprehended forms.

A rational creature is different again. Such a creature does not merely have certain purposes that it naturally pursues, as a plant does, and it does not merely feel appetites for perceived goods, as a nonrational animal does. A rational creature is one that apprehends its environment and its good in a still more profound sense: it is a creature that brings its representations under general concepts, and this means that it can not only acquire particular representations of the world through perception but think in the abstract about what is true, and also that it can not merely have particular desires but reflect on how to attain particular goods, on how to combine various sorts of goods in a life well-lived, and on the notion of *a good* as such. With these powers, moreover, comes the capacity for a distinctive kind of self-movement, one that involves the ability to think about what to pursue and how to obtain it. The fact that rational creatures do not merely have certain things that are goods for them, or merely desire things that are in fact good – the fact that they can reflect on the concept of a good life – is an aspect of their practical self-consciousness.[33]

The life of a rational creature is thus profoundly different from the life of a nonrational creature, but, if Aquinas is right, this difference does not disrupt the general connection between self-movement and goodness. Rather, it transforms the kind of good that is in question, and makes an individual creature's relationship to its own good correspondingly more complex. For – taking the second point first – a rational creature's self-movement is

---

[33] If this seems a dark suggestion, compare it with Kant's well-known claim that the imperative "ought" expresses "the relation of an objective law of reason to a will that is not necessarily determined by this law because of its subjective constitution" (*Groundwork*, Ak. IV:413). Kant's claim suggests a way of looking at the significance of normative language more generally: namely, as expressing the idea of a distinction, and the possibility of an accord, between how something is actually and how it is in the nature of that thing to be (a distinction, as Kant would put it, between how the thing is actually and the "law" that governs it). To think of the objective laws of something's constitution in this way is to think of them as specifying its good. And if this is right, then a creature that can think about what it ought to do and what is good for it would be thinking in terms that presuppose a conception of *its own* constitution. Such thinking expresses not just awareness of myself as an individual, but an at-least-implicit idea of what *kind of thing* I am: for when I think of how I ought to be, I am thinking in a way that must measure how I actually am against some general standard, a law I am under, or nature I bear, in virtue of the kind of thing I am.

mediated by *thoughts* about, and in the favorable case by knowledge of, its own good: its capacity for action is a capacity to pursue what it *takes* to be good. And by the same token, the life of a rational creature will be one that essentially involves exercise of the power of reason, and will imply the existence of goods of a specifically rational kind (e.g., contemplation, friendship, justice). The concept *good* to which a rational creature attains will be the concept of what is good in the life of that kind of rational creature, and this will not be merely determined by its "nature," if this means something independent of its reason. It would be a mistake to suppose that a rational creature is merely a creature that has a certain naturally-given inclinations and is capable of thinking about how to put that form into practice. That is not the sort of nature a rational creature has: its nature is to live a specific sort of rational life, and its good will involve whatever features life by reason must involve.

The rational apprehension of goodness. To substantiate this standpoint on rational agency in detail would require another paper, but the basic points follow from the considerations of §4, together with the thought that a rational agent is one which determines how it acts by thinking about how to act. For to say that a rational agent is one whose thinking about how to act determines how it acts is to say that it is an agent whose doing *A* depends on its capacity to reflect on the question "What shall I do?", in such a way that its acceptance of the answer "*A*" can constitute its setting to do *A*.[34] At a minimum, this will involve the power to consider the relation of means to ends: a rational agent will be one who, if he is intentionally doing *A*, and if he recognizes that doing *A** is means of doing *A*, can thereby come to do *A** because he is doing *A*. In short, a rational agent will be one whose thought can make true propositions of the form

(8) *S* is doing *A** because *S* is doing *A*.

This is the kind of proposition that was the focus of our discussion in §4, the kind on which our characterization of goal-directed movement hinged. But observe that, when applied to

[34] This is not to suggest that, whenever an agent intentionally does one thing with a view to doing another, this must be the product of prior deliberation about this course of action: the claim that intentional action must always be the product of prior deliberation is both factually untrue and subject to difficulties of principle. But this is not our claim: we hold that the capacity for intentional action requires a *capacity* for reflection on the question "Why do *A*?" Only some intentional actions are the result of actual prior reflection on this question, but even in the case of actions that are not the result of prior reflection, the relevance of the capacity for such reflection is evident in the fact that an agent is *able* to say what he is doing and why if the question arises.

rational agents, propositions of form (8) take a special turn: they require for their truth that the subject should *know* the relevant explanation to be true. A subject who does not know that he is doing *A* is not doing *A intentionally*, and a subject who does not know that he is doing *A** because he is doing *A* is not doing *A* with the intention* of doing *A*.

Now, if a rational agent is one whose self-movement is subject to explanations which are essentially known to their subject, what follows about the way in which such an agent must think of his own actions and of what explains them? Well, if we are right that, in general, truths of form (8) presuppose general truths about the kind to which the subject belongs, truths which carry implications about what is good for things of that kind, then a rational agent, in knowing what it is doing and why, will *know* truths with such presuppositions.[35] Moreover, his knowledge of such truths will not be merely an observer's knowledge: his taking such facts to hold will, as we have seen, be essential to their holding. This implies that, on pain of regress, his grounds for believing that he is doing *A** because he is doing *A* cannot be grounds for believing that this fact *already* obtains – for until he takes the relevant fact to obtain, it does not. He must, if he is to be rational in making such a judgment, have another kind of ground: a ground for believing that, in accepting a certain justification for doing *A**, he will be making it the case that he is doing *A**. But what kind of grounds could these be? There is at least this constraint on them: they must be grounds that are potentially such as to warrant a judgment of the relevant kind; and we have argued that the relevant kind of judgment is one that presupposes facts about what is good for bearers of his sort of form. But then, if the considerations of the preceding section are correct, it follows that a rational agent's grounds for making a judgment of form (8) must be grounds that bear some relation to general sorts of aims belonging to creatures of the form that he bears.

This conclusion obviously needs further explanation, and we will elaborate on it in a moment; but first let us note how well it coheres with some familiar facts about how we know what we want and what we are doing. We noted in §1 that a person who wants something can in general be asked *why* he wants it, and then he is expected to answer, not by describing how, as a matter of past history, his want arose, but rather by explaining what, here and now,

---

[35] Note that we are not claiming that, in general, if I know that *p*, and *p* presupposes *q*, then I know that *q*. That an agent who knows a truth of form (8) about himself must have views about what good would be served by his so acting will follow only given further considerations which we explain below.

speaks in favor of pursuing the aim in question. Similar points apply, *mutatis mutandis*, to an agent's knowledge of why he is doing what he is doing: if *S* is doing *A* intentionally, then *S* is expected to be able to say why he is doing it by giving the reasons he takes to speak in favor of doing *A*. Moreover, as G. E. M. Anscombe observed in a famous discussion of practical reasoning, explanations of this sort have a characteristic structure: where there is a substantive answer to the question "Why are you doing *A*?" (or "Why do you want to do *A*?"), a full statement of this answer terminates in some "desirability characterization" which represents what is sought as desirable or *good* in some intelligible respect (pleasant, appropriate, conducive to health, beneficial to a friend, required by fairness, satisfying of some intelligible life-ambition, …).[36] The agent may be wrong in supposing that the desired object would really *be* good in the relevant respect. Indeed, an agent may do something although, in his considered judgment, it is not a good thing to do at all. But even in this sort of case, if the agent has acted intentionally, we expect him to be able to explain what he did by saying what it was that seemed *prima facie* attractive about it, what form of good it at least appeared to promise. And, as Anscombe remarks, "the good (perhaps falsely) conceived by the agent must *really* be one of the many forms of good" (1963, §40).

At least as a characterization of the presumptively normal case, these Anscombean descriptions seem evidently true: a rational agent is normally assumed to be in a position to explain his own doings and wantings-to-do, and the kind of explanation that is called for here is one that relates the doing in question to something that might intelligibly be taken to be some form of good. But why *should* the primary explanation of what an agent wants or is doing be one that he can himself supply, and why should it take this particular shape? Our account of the conditions under which a rational agent can know himself to be the subject of a fact of form

(8) *S* is doing *A** because *S* is doing *A*

suggests answers to these questions. For we have observed, first, that a rational agent's grounds for such a judgment must be grounds for believing that in so judging he is *constituting* the explanatory relation at issue; and we have argued, secondly, that his grounds for so judging must ultimately advert to the way in which the present goal-directed activity

[36] Anscombe 1963, §37. See also §§38-40 for further elaboration of the point.

serves some general sort of aim belonging to creatures of his form. The first point speaks to the first question: he must be able to give the explanation because this is a sort of explanatory relation that holds only insofar as the subject takes it to hold. The second point speaks to the second question: the explanation must terminate in a desirability characterization because what is articulated in such a characterization is precisely an account of the general aim that this case of goal-directed activity instances, and this is the sort of ground that any truth of form (8) must ultimately have.

An excessively idealized picture? The preceding paragraphs are an attempt to give some content to the Thomistic-Aristotelian idea of a hierarchy of different kinds of life, distinguished by different sorts of relation to form, but united by an abstract structure which connects the ascription of self-movement to individuals with the recognition of them as belonging to kinds that supply norms for their activities. This picture of a "ladder of nature," with nonrational animals occupying one rung and rational animals occupying another, higher rung, is often dismissed as reflecting a discredited worldview, one that is incompatible with Darwinian biology and with a more general commitment to naturalism. It would be impossible for us to address this general misgiving here. We can only ask the reader to consider whether our claims that the recognition of self-movement brings with it a certain kind of explanatory framework, and that the recognition of rational self-movement involves the application of a distinctive version of that framework, need stand in tension with a naturalistic understanding of the evolutionary background and material basis of vital processes. Our own view is that, if the standpoint articulated here is in tension with anything, it is only with certain philosophically-contentious interpretations of what naturalism must involve.

Having tabled this broad misgiving, however, there is a more specific concern that we can address. This is the concern that our account of rational agency is excessively *idealized*. This concern can take two forms. First, our account of a rational creature's relationship to its own good may seem to demand too much intellectual sophistication, more than actual human beings typically possess. Secondly, even if it is granted that rational creatures must be capable of conceiving of their own good, the claim that intentional action in general reflects a subject's taking what he is pursuing to be good may seem to imply a naïve or pious view of

what moves us to act. We address these objections together because, although they take issue with different aspects of our conception of rational agency, we believe they are both rooted in a failure to grasp the logical character of the guise of the good thesis. Answering the objections will give us an opportunity to clarify this character.

Consider first the charge that the account demands too much intellectual sophistication on the part of rational creatures. Can't children and unreflective adults possess the capacity to act for reasons even though they never have reflected on the general concept *good*? Of course they can; but to focus on the question whether rational creatures in general must make explicit use of this concept is to miss our point. To see this, compare this charge of intellectualism with a similar charge that might be brought against the claim that rational creatures must believe under the guise of the true. Here too it is natural to object that only a quite sophisticated reasoner gives thought to the concept *true*. But the point of saying that rational believers must believe under the guise of the true is not to insist that all their beliefs take the form "*p* is true," but rather to point out something about the manner in which they must consider any belief of the form "*p*." A rational believer is one who can reflect on his grounds for holding any given belief, one who can put to himself the question whether *p* and why. "Truth" names, so to speak, the dimension in which answers to the question whether *p* add up: a consideration is relevant only if it speaks to whether *p* is true, and a creature that is not capable of distinguishing relevant from irrelevant answers to this question is not yet capable of reflecting on its grounds for belief. Thus any rational creature must at least implicitly possess the idea of truth: its capacity for reflection involves a capacity to bring the standard of truth to bear, even if it has never reflected on this standard as such or mastered a term that designates it. It may not have the predicate "is true" in its vocabulary, but the standard that this predicate designates provides the structure of its reflection. Moreover, even an unreflective rational subject who believes a proposition accepts it *as true* in a plain enough sense: such a subject has the power to reflect on what he believes, and if on reflection he does not take the relevant proposition to be true, he will in so judging have changed his belief about it.

Similarly, the point of saying that rational agents must act under the guise of the good is not to insist that their practical reasoning must arrive at conclusions of the form "Doing *A* would be good," but rather to point out something about the manner in which they must

consider the question whether to do *A*. Just as a rational believer is one who can reflect on his grounds for belief by putting to himself the question "Why *p*?", so a rational agent is one that can reflect on his grounds for action by putting to himself the question "What speaks in favor of doing *A*?" Just as "truth" names the standard we apply in answering the former question, so "goodness" names the standard for answers to the latter: it specifies the topic on which a consideration must bear in order even to be a candidate answer to the question. And although there may be rational creatures which have not reflected on this topic as such or learned a term that designates it, still any such creature must grasp this topic at least implicitly, for it must recognize what *kind* of answer the question "What speaks in favor of doing *A*?" calls for, on pain of not being a rational agent at all. If this is right, then the intentional actions of a rational agent express his regarding those actions *as good* in a plain enough sense: for such an agent has the power to reflect on how to act, and if on reflection he does not accept that a given way of acting has at least something good about it, he will in so doing have changed his mind about whether to do it.

But isn't this account of the motivation of intentional action too pious? Don't we sometimes intentionally do things that we judge to be wicked or worthless? No doubt we do, but care must be taken in interpreting this fact. It is sometimes suggested that the capacity for rational reflection effects a radical separation between an individual subject's thinking about what to do and any general facts about what kind of creature that subject is and what is good for such creatures.[37] On this view, the power to reflect on the question "What should I do?" is the power precisely to *transcend* any allegiance to the goods of one's kind – not, as we maintain, the power to relate to those goods in a way mediated by the capacity for self-consciousness. This conception of reason as the power to transcend any merely automatic allegiance to a certain list of goods does, of course, contain an important truth: namely, that a rational creature is one that can ask itself, with regard to any given putative good, "Why should I care about that?" In acknowledging this truth, however, we should not forget how one goes about answering this sort of question. Suppose the putative good in question is, say, justice to one's fellow human beings, and suppose that some agent seeks to call the authenticity of this good into question: what sorts of considerations might figure in his

[37] For this sort of view, see for instance John McDowell, "Two Sorts of Naturalism" (1995).

critique? Well, he might, for instance, condemn justice as Thrasymachus and Callicles do, by arguing that the life of a just person is cowardly and slavish. But notice that in doing so he would be condemning one putative good by appealing to others – in this instance, the goods of a courageous and free existence. And surely some such appeal would occur in any intelligible answer to this sort of question. A rational subject can call the value of any given good into question, but such questioning must, if it is to be intelligible as reasoning at all, appeal to other goods whose value is not in question.

Even an agent who does something he takes to be wicked or worthless can normally say why he did it, at least in the sense that he can say what feature of the action was *prima facie* appealing to him. Moreover, the action's at least *seeming* attractive to him in the relevant respect is precisely what explains his performing it, in spite of his reservations. If this is right, then such examples do not show the guise of the good thesis to be false. Indeed, they rather tend to confirm it. For, as Aquinas notes in the passage quoted at the beginning of this section, the presence of reason in us is exactly what makes room for willings that do not tend toward things that are good "in very truth": once we are equipped with the thought that rational action involves an exercise of the power to judge things good, we are in a position to recognize that there can be appetites which the subject does not judge or believe to be good at all. Such appetites would be akin to perceptual appearances: they would be *appearings-good*, just as perceptual appearances are appearings-true. They would present themselves as *prima facie* grounds for choice, just as perceptual appearances present themselves as *prima facie* grounds for judgment. But just as a subject can disbelieve the testimony of his senses even as the appearances persist (e.g., I can disbelieve that the stick in the water-glass is really bent although it continues to look bent), so a subject can disapprove of the urgings of his appetites even as these urgings persist (e.g., I can think that lust is encouraging me to do something bad, although the impulse to do it remains as strong as ever). The objects of such appetites are not believed good, but the appetites themselves continue to represent those objects as good in a perfectly intelligible sense: they present their objects as having features that make them *prima facie* desirable, even if we doubt that something with those features would really be desirable.[38]

---

[38] For development of the idea that appetites can be understood as appearings-good, and the use of this idea in replying to supposed counterexamples to the guise of the good thesis, see Sergio Tenenbaum, *Appearances of*

We thus arrive at the traditional view that when a person intentionally does something he judges to be wicked or worthless, this reflects a kind of mutiny of some of his motivational faculties, a mutiny that involves the action's being put forward as desirable by appetite or passion, even as reason denies its desirability. This sort of mutiny is certainly possible, but what is not possible is that it be the normal case. For a subject possess the power of practical reason only if his reflection on what to do is in general determinative of what he actually does do, and we have seen that a rational subject's reflection on what to do must be at least implicitly sensitive to his judgments about what would be good to do. Hence a rational agent must in general judge his actions to have something genuinely good about them; and even in the cases where he does not, this will be because he is subject to an appearance of goodness which engages his power to act in conscious pursuit of ends, even as it overpowers his judgment about what is really good.

Conclusion. The thesis we have been defending is abstract. This bears emphasizing because it brings out how many questions of substantive moral philosophy are left open by our conclusions. We have concluded that a rational agent must act under the guise of the good in the sense that he must in general pursue ends in virtue of taking there to be something good about those ends. We have granted that his recognition of this standard of goodness may be merely implicit, and we have said about the relevant standard only that it must reflect more general facts about the kind to which the agent belongs. We have not said anything about the content of this standard, not even whether the propositions that characterize it depend merely on abstract facts about what it is to be a rational being (this would be a kind of Kantian view), or on more determinate facts about what it is to be a *human* rational being (this, we believe, would be closer to Aristotle's view). In any event, to investigate the content of our concept *good* would require a different kind of reflection from the one in which we have been engaged: we have merely been investigating the *form* of this concept, and the kind of role it must play in our practical thought. But this investigation suffices to answer philosophers who would deny that this concept has any role to play in our wanting and willing.

Our project in this paper has been to make a case for the Aristotelian doctrine that rational agents must desire under the guise of the good; but we hope that even readers who are

---

the Good (2007).

not convinced by this case will now possess a clearer view of the context into which the guise of the good thesis fits. Most contemporary opponents of the thesis discuss it in isolation from broader issues about teleological explanation and the relation between an individual creature and its kind. Our aims have been, first, to show how the thesis fits into this larger framework, and secondly, to suggest that the framework itself is not just some antiquated worldview that we can brush aside, but a perceptive analysis of forms of thought that are essential to our everyday understanding of ourselves. If the thesis turns out to articulate a commitment involved in the application of these forms of thought, then it cannot be lightly dismissed, for we can no more brush aside these forms than we can brush aside our capacity for choice itself.

REFERENCES

Anscombe, G. E. M. 1958. "Modern Moral Philosophy." *Philosophy*, 33, 124, pp. 1-19.
_____. 1963. *Intention*. Second Edition. Oxford: Basil Blackwell.

Aquinas, Thomas. 1948. *Summa Theologica*. 3 Vols. Trans. Fathers of the English Dominican Province. New York: Benziger Brothers.

Aristotle. 1984a. *The Complete Works of Aristotle*. 2 Vols. Ed. Jonathan Barnes. Princeton: Princeton University Press.
_____. 1984b. *De Anima*. In Aristotle 1984a.
_____. 1984c. *Metaphysics*. In Aristotle 1984a.
_____. 1984d. *Nicomachean Ethics*. In Aristotle 1984a.
_____. 1984e. *Parts of Animals*. In Aristotle 1984a.
_____. 1984f. *Physics*. In Aristotle 1984a.

Boyle, Matthew. MS. "Rational Animals and Rational Powers."

Carlson, Gregory N. and Francis Jeffry Pelletier, eds. 1995. *The Generic Book*. Chicago: University of Chicago Press.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Davidson, Donald. 1980a. *Essays on Actions and Events*, Oxford: Oxford University Press.
_____. 1980b. "Freedom to Act." In Davidson 1980a.
_____. 1980c. "Mental Events." In Davidson 1980a.

Foot, Philippa. 2002. *Natural Goodness*. Oxford: Oxford University Press.

Geach, P. T. 1956. "Good and Evil." *Analysis*, 17, 2, pp. 33-42.

Goldman, Alvin I. 1970. *A Theory of Human Action*. Princeton: Princeton University Press.

Harman, Gilbert. 1976. "Practical Reasoning." *The Review of Metaphysics*, 29, 3, pp. 431-463.

Hawthorne, John and Daniel Nolan. 2006. "What Would Teleological Causation Be?" In John Hawthorne, *Metaphysical Essays*. Oxford: Oxford University Press.

Hornsby, Jennifer. 1980. *Actions*. London: Routledge.
_____. 1995. "Agency and Causal Explanation." In John Heil and Alfred Mele, eds., *Mental Causation*. Oxford: Oxford University Press.

Kant, Immanuel. 1987. *Critique of Judgment*. Trans. Werner J. Pluhar. Indianapolis: Hackett Publishing Company.

_____. 1993. *Grounding for the Metaphysics of Morals*. Trans. James W. Ellington. Indianapolis: Hackett Publishing Company.

Lavin, Douglas. MS. "Must There Be Basic Action?"

McDowell, John. 1995. "Two Sorts of Naturalism." In Hursthouse, Lawrence, and Quinn 1995.

Moravcsik, Julius. 1994. "Essences, Powers, and Generic Propositions." In *Unity, Identity, and Explanation in Aristotle's Metaphysics*, ed. T. Scaltsas, D. Charles, and M. L. Gill. Oxford: Oxford University Press.

Raz, Joseph. Forthcoming. "The Guise of the Good." This volume.

Russell, Bertrand. 1992. *An Inquiry into Meaning and Truth*. London: Routledge.

Scanlon, T. M. 1999. *What We Owe to Each Other*. Cambridge: Harvard University Press.

Searle, John R. 1983. *Intentionality*. Cambridge: Cambridge University Press.

Setiya, Kieran. 2003. "Explaining Action." *Philosophical Review*, 112, pp. 339-393.

Smith, Michael. 1987. "The Humean Theory of Motivation." *Mind*, 96, pp. 36-61.
_____. 2004. "The Structure of Orthonomy." In *Agency and Action*, ed. John Hyman and Helen Steward. Cambridge: Cambridge University Press.

Stocker, Michael. 1979. "Desiring the Bad." *Journal of Philosophy*, 76, pp. 738-753.

Tenenbaum, Sergio. 2007. *Appearances of the Good*. Cambridge: Cambridge University Press.

Thompson, Michael. 2008. *Life and Action.* Cambridge: Harvard University Press.

Velleman, J. David. 1989. *Practical Reflection*. Princeton: Princeton University Press.
_____. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.
_____. 2000b. "Introduction." In Velleman 2000.
_____. 2000c. "The Guise of the Good." In Velleman 2000.

Watson, Gary. 1982. "Free Agency." *Journal of Philosophy*, 72, pp. 205-220.

Wiggins, David. 2001. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

Wittgenstein, Ludwig. 1972. *Philosophical Investigations.* Trans. G. E. M. Anscombe. Cambridge: Cambridge University Press.