



The Origin of the Good and Our Animal Nature

Citation

Korsgaard, Christine. 2020. "The Origin of the Good and Our Animal Nature." *The Journal of Ethical Reflections* 1 (2): 7-28.

Permanent link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37366069>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

The Origin of the Good and Our Animal Nature

Christine M. Korsgaard

Harvard University

The best illustration is a doctor doctoring himself: nature is like that.
- Aristotle¹

ABSTRACT: We use the term “good” in two contexts: as the most general term of evaluation, and to refer to the final ends of life and action. I start from the question what evaluative and final goodness have to do with each other. Do we use the same term because when we talk about final goods, we are evaluating ends and lives? If so, how do we go about doing that? Most things are evaluated with respect to their fitness to perform their function, but ends and lives do not have functions.

I contrast three theories of the final good: the intrinsic value theory, the hedonist theory, and Aristotle’s account, which identifies a being’s final good with its well-functioning, a form of evaluative goodness. Aristotle’s theory suggests an illuminating relationship between evaluative and final goodness: a conscious being has a final good when she functions by having conscious states that track, and so enable her to pursue, her functional or evaluative goodness. It

¹ Aristotle, *Physics* II.8 199b 30-32. References to Aristotle’s works will be given by the standard Bekker page, column, and line numbers.

is therefore the nature of an animal to have a final good, and there are such things as final goods because there are animals. This theory explains the existence of final goods without any metaphysical appeal to intrinsic values.

KEYWORDS: Animals, Aristotle, Consciousness, Function, Good, Hedonism, Intrinsic Value, Kant

I. Introduction

This paper is part of a larger project in which I investigate the origins of value. Those who believe that there are intrinsic values—that some objects, activities, or entities simply have the *property* of being valuable—do not feel a need to answer questions about the origins of value. For them, value is just there. But I believe that all value is dependent on the existence of valuing beings. In this paper, I am going to defend an account of the good, derived from Aristotle, that grounds it in our animal nature, and explain why I think this view is superior to some of its rivals. One reason I think it is superior is that it enables us to explain why there is such a thing as the good.

II. A Puzzle about the Good

I want to begin my discussion by noticing something that I think we should find puzzling about our use of the concept “good.” The term “good” is used in two broadly different ways. First, “good” is our most general term of evaluation, a term we apply to nearly every kind of thing, or at least every kind of thing for which we have any use, or interact with. Think of the wide variety of things we evaluate as good or bad: cars, houses, machines and instruments, food, weather, days, prose, pictures, movies, people considered as occupying roles such as mother, teacher, son, doctor, and people considered just as people, among many other things. All of these things may be evaluated as good or bad. Evaluation is usually related to the purpose, role, or function of the entity that is judged good or bad: an entity is good in the evaluative sense when it has the properties that enable it to serve its function—either its usual or natural function or one we have assigned to it for some specific purpose. I will call that the *evaluative* good.

I call “good” in the second sense in which we use the term the *final* good, borrowing one familiar translation of the Greek word “*telos*.” Final goods are the ends of action. We suppose that something we call “The Good” is the end or aim of all our strivings, the *summum bonum*, a state of affairs that is desirable or valuable for its own sake. We are usually talking about a person’s final good when we speak of what is good *for* that person. That is, the things that are good or bad for a person are things that have an impact on his final good. We

sometimes call our own final good “the human good,” suggesting that things other than human beings have a final good of their own, and perhaps also that the good for a thing is relative to its nature.

The puzzle is simply this: what is the relation between the evaluative and the final sense of good? Why do we use the same word as a general term of positive evaluation, and to designate the final ends of our lives and actions? I think that most people do *not* find this puzzling because they think that the answer is obvious: when we talk about someone’s final good, we are still using the term evaluatively: we are evaluating the person’s life. I do not mean we are evaluating it morally. That would be an evaluation of the person himself. Rather, we are evaluating something about how the life goes and the total circumstances in which it is lived.

It is tempting to say that we are evaluating the *quality* of a person’s life. But the phrase “evaluating the quality” just says the same thing twice over, namely, that his life is a proper subject of evaluation, it is the sort of thing that can be of a high or low quality. But that is exactly the problem. Ordinarily, as I mentioned earlier, we evaluate things by asking whether they have the properties that enable them to perform their function, but a person’s life and circumstances, considered just as such, do not seem to have a function. So how do we go about evaluating them? If we ask whether, say, a car is good, we are asking whether it has the properties that enable cars to perform their function well: whether it handles well, gets good gas mileage, is safe, and things like that.

But when we ask whether a person's life is good, we do not seem to be asking anything *except* whether the person whose life it is achieves that thing we call "The Good." This becomes particularly obvious if we want to leave open the possibility that the human good is something that is in a certain way external to life itself—the way it is on some conceptions of, say, nirvana, or salvation. People who believe in such final goods do believe that human life has a purpose, but that purpose seems to be to enable us to achieve that thing we call "The Good." But then *what* are we evaluating when we talk about "The Good"?

One might be tempted to answer that we are evaluating possible ends: things we might pursue for their own sake. But then what is the function of an end—or, perhaps to put it more intelligibly, what makes something fit to be an end? Apparently, that it is something good for its own sake, or finally good. This is the kind of consideration that drove G. E. Moore to the view that there simply *nothing* we can say about the good except that it is good, in just the same way that there is nothing we can say about red except that it is red: it is just a property.² It follows from Moore's theory that there is no way we can know what is good in the final sense except by a power of rational intuition that functions like a sense. If Moore is right, there is no point in trying to identify the human good through philosophical argument: we just have to *focus* our powers of

² In G. E. Moore, *Principia Ethica*.

intuition. But I believe that the puzzle has an answer, and that the situation is not as hopeless as Moore made it seem.

III. Three Theories of the Good

I will come back to the puzzle, but first I want to describe three theories of the final good I am going to canvass in my search for an answer, and note where they stand on a certain question: namely, whether and how the final good for a being is relative to its nature. According to what I will call the intrinsic value theory – this is essentially G. E. Moore’s theory – certain objects, states of affairs, activities, or forms of experience have the property of being intrinsically valuable. On this theory, it is natural to suppose that the good for a sentient being consists in exercising his ability to experience, appreciate, or participate in these intrinsically valuable things. According to this view, the good for such a being is relative to his nature, but only in the sense that his nature determines which kinds of intrinsic value he is able to experience, appreciate, or participate in. The human good is a richer thing than the good for a non-human animal, because our nature enables us to enjoy the objective values of art, literature, science, philosophy, and humor, say, while another animal’s nature might enable her to enjoy only such simple intrinsic values as pleasurable experience, family affection, and so on. We can intelligibly say that it is better to *be* a human being than to be another sort of animal, since human beings get to participate in a wider range of intrinsically valuable activities. Plants and inanimate objects, on

this view, do not have a final good at all since they are unable to participate in valuable activities or have valuable experiences.

According to the second view I will consider, hedonism, the good just *is* pleasurable experience or consciousness and the absence of painful experience or consciousness. What makes a being capable of having a final good is simply that the being is conscious and can experience pleasure and pain. Otherwise, her good is not relative to her nature. On this theory it is a real question whether some of the other animals might not have a better life, or at least be capable of having a better life, than human beings, given their apparent enthusiasm for simple and readily available joys. Although I will treat it as a separate theory, hedonism, I believe, has an inherent tendency to collapse either into a version of the intrinsic value theory, or into a version of the third view I am about to describe. Obviously, it is possible to regard hedonism simply as a particular instance of the intrinsic value theory, one that singles out conscious experience as the only possible bearer of intrinsic value. But I think this way of looking at hedonism does not do justice to the intuition that has made hedonism seem plausible to so many thinkers, which is precisely the idea that the final good must have an irreducibly *subjective* or relational element. That is, what makes hedonism seem plausible is precisely the idea that the final good for a sentient being must be something that can be felt or experienced *as a good* by that being. It is something that can be perceived or experienced as welcome or positive from the being's own point of view, and that is therefore relative to the being's own

point of view. The intrinsic value version of hedonism tries to capture the essentially subjective element of the final good by attaching objective intrinsic value to a subjective experience, but when this move is made the essentially *relational* or *relative* character of subjectivity tends to drop out. The goodness of the experience is detached from its goodness *for* the being who is having the experience, and instead is located in the character of the experience itself. This defect shows up most clearly in utilitarian versions of hedonism, which allow us to add the goodness of pleasant experiences across the boundaries between persons or between animals. There is no subject *for whom* the total of these aggregated experiences is a good, so the aggregate good has completely lost its relational character: the goods are detached from the beings for whom they are good. This relational element of value, I believe, is better captured by the third theory I am about to describe.

The third view, and the one I wish to defend, is a version of the account suggested by Aristotle's famous function argument in Section Seven of Book One of the *Nicomachean Ethics*. Aristotle says:

Presumably, however, to say that happiness is the chief good seems a platitude, and a clearer account of what it is is still desired. This might perhaps be given, if we could first ascertain the function of man. For just as for a flute-player, a sculptor, or any artist, and in general, for all things that have a function or activity, the good and

the 'well' is thought to reside in the function, so it would seem to be
for man, if he has a function.³

Aristotle's point is not that human life has a purpose and our good rests in serving that purpose.⁴ Rather, characterizing the view in an abstract and somewhat cumbersome way, Aristotle's idea is that the good for a being consists in the well-functioning of that being as the kind of being that it is, in circumstances that are conducive or favorable to its overall well-functioning. I will make this idea less abstract later on.

Now Aristotle's view, like the hedonist view, could conceivably be reduced to a version of the intrinsic value theory. We could say either that the well-functioning of a sentient being has intrinsic value, or that a sentient being is well-functioning when he participates in whatever intrinsically valuable activities his nature makes possible for him. But I think Aristotle's view is more interesting if we combine it with a Kantian approach to value, which makes all values relative to what we might broadly call our valuing capacities: the capacity to find something pleasant, interesting, enchanting, satisfying, or stimulating – and of course to experience the opposite responses as well. Sticking to the human case for now, according to this view, what makes a state of affairs or an object or an experience valuable is precisely its capacity for eliciting joy, interest, or

³ Aristotle, *The Nicomachean Ethics*, NE 1.7 1098a 22-29.

⁴ See my "Aristotle's Function Argument" for further defense.

appreciation from human beings, together with the value that, according to Kant, we necessarily set upon our humanity itself. Science and philosophy are valuable for human beings *because* they engage and arouse the intellectual faculties of human beings. Art and music are valuable for human beings *because* of their capacity to elicit complex and satisfying perceptual experiences refined by thought and emotion. Fine food and wines are valuable for human beings *because* of our capacity to aestheticize the appetites we share with the other animals. Love and friendship are valuable because of the human social needs they satisfy and the human powers they arouse in us. And so on. If we ask why we treat the things that are good *for* human beings as *finally* good, as valuable for their own sakes, Kant's answer, given by his argument for the Formula of Humanity, is that we take these things to be valuable because of the value that we necessarily place on ourselves. To take what is *important to you* to be *important, period*, is just what it *means* to place a value on yourself.⁵ And to render these values normative is, accordingly, to express the value you place on yourself and on humanity generally, by willing these values as laws. So what makes our well-functioning good is not that it has intrinsic value, or that it consists in the pursuit of intrinsic value, but rather that we ourselves value it and confer normativity upon it by

⁵ See my "Kant's Formula of Humanity" especially pp. 122-123; and *The Sources of Normativity*, §§3.4.5-3.4.9, pp. 117-123.

willing it as a law that this value should be realized. In what follows I will be arguing that we have to do that, because of our animal nature.

According to this Aristotelian theory, values are relative to a being's nature. It is good for a human being to philosophize or explore nature or fall in love in just the same way it is good for a horse to run, or a whale to breach, or for a tiger to hunt, or for an insect to pollinate its characteristic plant—because that is the fulfillment or realization of an animal's nature—because that is how he functions. On this view, interestingly, we cannot say that it is better to be a human being than to be another kind of animal, or if we can, it can only be with reference to kinds of goods that we share with the other animals—to the goods that pertain to our animal nature as such, rather than to specifically human goods.

IV. The Nature and Content of the Good

Before I continue, I want to mention two other theories of the final good commonly discussed in the literature, namely perfectionism and eudaimonism, and explain why I am not discussing them here. This will give me an opportunity to clarify something about the nature of the question I am trying to raise.

As the translation of the passage from Aristotle I just quoted suggests, people often identify the good with happiness. The theory that happiness is the good is sometimes called eudaimonism, from the Greek word that gets translated

“happiness” here—*eudaimonia*. One reason I will not be discussing the view that the good is happiness here is that the notion of happiness is almost as obscure as the notion of the final good. In fact, we can interpret “happiness” so broadly that it means pretty much the same thing as the final good—although with a nod towards the subjective element in that idea that I have already mentioned in connection with hedonism. On the other hand, if we interpret “happiness” so that it means something more specific than that, then the theory that happiness is the good is not a theory of the good in the sense that I am talking about here, as I will explain in a moment. Aristotle’s view is also often classified as a version of what is called “perfectionism,” the view that the good for a human being rests in the development or realization of human capacities.⁶ On the assumption that a thing’s well-functioning is expressed in the realization of its natural capacities, it makes sense to attribute this to Aristotle. But some perfectionists seem to have a further idea in mind—one of using our capacities and powers to the fullest, or maximizing their use or something like that. It is not really clear why the bare idea of well-functioning should involve the idea of using one’s powers to the *fullest*. However that may be, there is a reason I have not included these two theories among my candidates, although this reason is a little difficult to explain.

These theories are most naturally understood as theories about the *content* of the

⁶ The term “perfectionism” is also used to characterize certain moral theories, generally theories which take the goal of moral action to be the maximizing or promoting of the human good in the perfectionist sense. I am only talking about perfectionism as a theory of the good here.

finally good – about what in particular is finally good. Whereas I am looking for an account of what the good is that would enable us to pick out its content, one that would enable us to say which things are good and explain why they are, or at least to say how we would go about picking which things are good. I might say that what I am asking about in this paper is not the *content* of the good, but the *nature* of the good, on the assumption that if we knew what sort of thing the good is, we would be able to apply that knowledge to discover which things are good. The distinction is clearer in some theories than others, and Moore's theory enables me to show you what I have in mind. Moore's theory about the *nature* of the good is that it is an intrinsic property, discerned by intuition. His theory about its *content* is that consists in aesthetic experience, friendship, knowledge, and so on. Now, in spelling out my three theories, I mentioned that either Aristotle's theory or hedonism may be regarded as applications of the intrinsic value theory, that is, claims about which things have intrinsic value, and so as claims about the content of the good. But I am treating these theories as theories about the nature of the good.

That this sort of slippage is possible, I think, shows that philosophers do not usually make a very firm distinction between views about the nature of the good and views about the content of the good. But this is no accident, because the difficulty of making this distinction is reflected in the puzzle itself. If we do not know what exactly we are evaluating or how exactly we are evaluating it when we claim something is good in the final sense of good, it is not surprising

that we cannot firmly distinguish between talking about the nature of the good and merely identifying its content. Most of the time, we know what we are doing when we identify something as evaluatively good. I want to know what we are doing when we identify something as finally good.

The fact that I am looking for a theory of the nature of the good in the sense I have just tried to explain will make the fact that I have included hedonism on my list seem peculiar, for surely it is most naturally interpreted as a theory of the content of the good. I have included it because I think, for reasons already mentioned, that it captures something important about the *nature* of the good—namely, the subjective and relational aspect of the good. As for Aristotle, explaining his account as an account of the nature of the good will be the work of the rest of this essay.

In fact, I hope you will not be too disappointed to learn that for purposes of this essay I am not going to say much at all about the content of the human good. I will take it for granted that any of these accounts might plausibly pick out the sorts of activities, experiences, and achievements that most of us are tempted to think must constitute the human good—love, knowledge, participation in social and civic life, intellectual and aesthetic experiences and activities, standing in a proper relationship to humanity and nature, moral virtue, significant achievement—all the usual things. Such things might be taken to be intrinsically valuable, or the sources of our deepest and steadiest pleasures, or the manifestations of the well-functioning of our nature, depending on which

theory is correct. My interest is rather in the metaphysical question what sort of thing we are talking about when we talk about the final good for a certain kind of being.

V. The Metaphysical Background to Aristotle's Theory

I am interested in Aristotle's view because it represents an interesting way of relating the evaluative and the final senses of good. Earlier I mentioned the view that we use the same term for the evaluative and final good because talk of the final good involves the evaluation of a life. According to Aristotle's view, we also use the same term because both the evaluative and the final good are matters of well-functioning. To say that something is evaluatively good is to say that it has the properties that make for well-functioning, and to say that something achieves its final good is to say that it in fact functions well. Importantly, though, we should say it achieves the final good not merely when it functions as well as it can given the circumstances, whatever they are, but when it functions well in circumstances that allow or and perhaps even facilitate its functioning well.

However, this is all very abstract, and in order to make it less so, it will be necessary for me to say a little about the metaphysical conception behind it. According to Aristotle, any substance or entity has a function. This is because according to Aristotle, a substance or an entity is matter so organized as to serve some purpose or function, to do something. Specifically, every entity can be

analyzed as a form in a matter. The matter is the material or parts of which the entity is composed, while the form is the arrangement of the matter or the parts that enables the entity to serve its purpose, or to do whatever it characteristically does. Of course, the idea is clearest in the case of an artifact or a machine. A car is, say, engine, gas tank, chassis, wheels, etc. organized in such a way as to form a guidable means of human transport, or something like that. The engine, gas tank, chassis, wheels, and so on are the matter or the parts; the form is that arrangement of those parts that enables the car to serve as a guidable means of human transport. In the case of an artifact, we identify the function or purpose of the entity in question by reference to our own purposes or that of its inventor.

Aristotle extended this basic idea—that a substance is a functionally organized unity—to living things by means of a thesis about what a living thing essentially is. A living thing is a substance so arranged as to secure the continuing existence of its own form. It does this in two ways: through nutrition, which enables it to preserve a continuing spacio-temporal stream of matter in its own arrangement or form, and through reproduction, which enables it to impose its form on other bits of matter. In other words, a living thing has a form that maintains matter in that very form. That is its function. A living thing functions well, essentially, when it manages to stay alive and reproduce. This metaphysical thesis does not imply that living things, like artifacts, were created by a designer for the purpose of preserving themselves and their forms. Instead, it simply asserts that that is what a living thing *is*. We identify a certain bit of matter as a

living thing or organism when it is so organized as to preserve its own form in these ways, when it has a self-maintaining form.

Each kind of organism has its own specific ways of carrying out its nutritive and reproductive activities, or its own form of life. We can identify it simply as the substance or entity that leads that form of life, or whose matter is organized in such a way that it maintains its form by living that form of life. Thus a dandelion is an entity that maintains its form through dandelion activities, such as spreading dandelion seeds on the wind, and a porcupine is an entity that maintains its form through porcupine activities such as defending itself with quills. In each case the function of the entity is simply to be what it is, to lead the kind of life it characteristically lives.

But we can also draw broad distinctions among types of life forms. Plants are the basic form of living organism, characterized simply by the powers of nutrition and reproduction. Animals, as Aristotle understands them, are characterized by an additional set of powers that determine the way they carry out the nutritive and reproductive functions—namely, the powers of perception and action, where action is understood basically as locomotion guided by perception. The idea of an animal, as Aristotle understands it, is the idea of an entity that preserves her form in part through her consciousness of her environment, and her resulting ability to respond to her environment in ways that serve to maintain her form. The idea is not, of course, that the animal *aims at* the preservation of her form, if that is understood to mean that the animal

consciously entertains such an end. Rather, the idea is that the *way* an animal functions is by having instinctive evaluative attitudes—desire and aversion, pleasure and pain, fear and interest—towards things that affect her functioning. Although I am obviously using the notion of an evaluative attitude very broadly, it may seem extravagant to ascribe evaluative attitudes of any kind to simpler animals. But I think that an animal's experiences must, at some level, however primitively, be aversive or welcome, in order to play a role in the animal's self-maintenance at all. If her perceptions are to guide her towards what she needs and away from what threatens her, they must render some things attractive and some repulsive. That is still an evaluative state, and that is all that the argument will require.

I should note that the Aristotelian categories of plant and animal are not precisely coextensive with the contemporary scientific use of the terms: an animal now is understood to be, very roughly speaking, a complex, multicellular organism that feeds on other organisms, as opposed to plants that can convert sunlight or perhaps methane into energy. Nowadays scientists believe that some organisms, such as fungi, do not fit into either of these categories, and some animals, such as sponges, do not fit Aristotle's definition very well. But of course there is broad overlap between Aristotle's categories and the modern one, and it is no accident. While it is not necessary that a creature that feeds on other life forms be percipient and mobile, that is how most kinds of animals function. In any case, it is animals in Aristotle's sense that I am talking about when I talk

about animals in this paper: beings who are guided by their evaluative attitudes to respond with appropriate actions to events in their environment, and who in that way preserve and maintain their forms.

Because he has the powers that make agency possible, Aristotle believed, an animal *lives* or *has a life* in a sense that a plant does not.⁷ Animals have experiences and they act, they *do* things, in a sense that plants do not: it is natural for us to describe animals, even fairly primitive ones, as hunting, eating, mating, defending themselves, raising their young. But the capacities for feeling and action are not just powers added, so to speak, on top of the animal's nutritive and reproductive life. They are powers that exist in the first instance as a way the way the animal carries out the tasks of nutrition and reproduction that an animal shares with plants. The animal's capacity for action shapes the way she gets food and produces offspring. However, as a result of having these powers animals also do things that plants do not do at all—they enjoy and suffer from their lives, and as a result they may do other things that plants do not do, like, say, loving, or playing. These facts make the “life” of an animal a different sort of thing than the “life” of a plant.

Aristotle thought that human beings, as rational animals, formed a distinct, third kind of being, with a third kind of life. I will not attempt to say

⁷ Aristotle refers to various senses of “life” (“the life of nutrition and growth” “the life of perception” “an active life of the element that has a rational principle”) in the function argument, in the argument following the lines I quoted earlier. (NE I.7 1097b20-1098a5)

here what Aristotle understands by “rationality.” But his idea is that the capacity for rationality changes the way we carry out the functions we share with the other animals, just as the capacity for action changes the way animals carry out the functions they share with plants. And, as in that case, it also adds to our repertoire of activities, expanding those to include such purely human activities as, say, the disinterested pursuit of knowledge and aesthetic activities. But the main change is that with rationality comes the power of choice, in a distinctive sense not shared by the other animals. For a non-human animal’s way of life is mapped out for her, at least broadly, by her instincts; and any two members of a given animal species basically live the same sort of life (unless the differences are biologically fixed, as by age and gender, or by kinds as among bees). A human being, as a rational being, therefore has a life in a different sense from this, for a human being has, and is capable of choosing, what we sometimes call a “way of life.” Thus rational nature, or personhood, introduces a new a form of functioning, and so a new form of life.

VI. Health and Goodness: Some Objections

In fact, this last consideration points in the direction of an answer to a possible objection to Aristotle’s theory. By now it should be obvious that Aristotle’s theory suggests that the final good for an organism is essentially to be healthy—more properly speaking, to lead a healthy life of its kind in circumstances favorable to its leading such a life and continuing to lead such a

life. And while that may be a plausible thing to say about the good for a plant or an animal, it may seem to be too thin as an account of the human good. But as I have just explained, it is part of Aristotle's view that, in virtue of rationality, human beings have a life in a sense that animals do not. The well-functioning that constitutes the human good is well-functioning in the specific kind of life made possible by the capacity for rational choice. It would take me too far afield to defend this claim now, so for now I will just assert that what it is to be well-functioning in that form of life is not merely a matter of health, at least as we ordinarily understand the idea of health.

But there is one ramification of Aristotle's theory of the human good that is worth mentioning. According to the other two theories of the good I have described, whether there is any connection between being a good person in the evaluative sense and achieving the final good seems to be an open question. Of course, among the circumstances that are part of the final good for an entity might be some that require that she have certain evaluatively good attributes, but whether this is so and which attributes are required is an empirical matter, or at any rate one to be settled by further argument. Suppose that human beings are good in the evaluative sense when we are morally virtuous. Whether that promotes our ability to participate in intrinsically valuable activities or to have pleasurable experiences is an open question. But in Aristotle's theory, the connection between being good in the evaluative sense and achieving the good in the final sense is not merely an empirical one, because both kinds of good

essentially involve well-functioning. *If* Aristotle is correct in thinking *both* that moral virtue is essential to human well-functioning – of course that is a very big “if” – *and* if he is correct in thinking that the final good is to be a well-functioning member of your kind, it will fall out as a kind of necessary truth that virtue is essential to the achievement of the good.⁸

Before I go on, I want to mention some other objections to Aristotle’s view that are connected to the fact that the good for an organism, considered as such, is to lead a healthy life. These objections have to do with the way we talk. I have claimed that a thing is good in the evaluative sense when it has the properties that enable it to perform its function well. But we do not usually call a plant or an animal “good” in virtue of being healthy. We might say that a healthy animal is a good specimen, but when we speak that way we are not talking merely about the fact that he is healthy – rather, we are talking about the fact that his health makes him useful to us in some way, say as an object of study, or as breeding livestock. And that brings me to the other side of this objection. We *do* tend to describe living entities as evaluatively good when they have properties that enable them to serve *our own* purposes. As we tend to use the evaluative notion, a good horse is one good for riding, good corn is corn good for the eating, and, if you would believe the caretakers of suburban lawns, the only good dandelion is a dead one. As we normally use the terms, then, we would not call a well-functioning dandelion good, and what we would call, say, “good corn” does not necessarily

⁸ But notice that it does not fall out that virtue is sufficient to the achievement of the human good.

have the properties that make a corn plant flourish. It seems conceptually possible, for instance, that the sweetest corn, best for the eating, might fail to reproduce well, or something like that.

I do not think that either of these facts about the way we talk should worry us. Although we do not call healthy animals “good” ones in virtue of their health, there is an obvious continuity between a well-functioning artifact and a healthy animal: both have the properties that enable them to do what they do successfully. But I will sometimes call “good” in the sense of having properties that make a thing well-functioning of its kind the “extended-evaluative” sense, to remind you that it includes both the ordinary evaluative sense and the organic idea of being healthy. This is reflected in the fact that the ordinary evaluative sense of good and the idea of being healthy seem to support the notion of “good for” and “bad for” in similar ways. Fatty foods are bad for you, and impure gasoline is bad for your car. That is suggestive, because this use of “good for” provides an apparent link between goodness in this extended-evaluative sense and goodness in the final sense. In fact, the idea of health in general seems interestingly poised between the evaluative and the final sense of goodness. For to say that an organism is healthy is clearly to evaluate how well it is functioning, and yet most of us would agree that health is at least an important part of the final good – and perhaps nearly the whole of it for some of the other animals.

As for the other side of the linguistic awkwardness – that the organisms we do call “good” are not necessarily healthy and thriving – that is no problem at

all. That simply reflects our own tendency to regard plants and animals as instruments, and to evaluate them as if they were a kind of artifact created for our use. And even though we do that, we do not usually go so far as to talk about what is “*good for*” say, plants, with reference to the ways in which plants are good for us; rather, we use it to refer to their own well-functioning. So, for instance, we might note with regret that the fertilizer we are using is just as good for the weeds as it is for the grass. Aristotle’s view that a thing’s final good is its own well-functioning actually explains why we can say this sort of thing, and why we say it in the case of living entities but not artifacts. Since an artifact exists and has a function only with reference to us and our needs, there is no real room for opposition between its good and our own. But because a plant’s function is defined with reference to its maintenance of its own form, to its ability to lead its own distinctive kind of life, such an opposition is possible: what is good for it may not be good from our point of view.

VII. What Kinds of Things Have a Final Good?

But is goodness in the extended-evaluative sense connected to final goodness in the way Aristotle supposes? Some of you are probably already growing restive under the idea, vaguely implied by what I have just been saying, that there *is* such a thing as a “final good” for, say, a car. We certainly do say that it is *good for* the car to drive it once in a while, meaning that driving the car regularly keeps it functional. But we do not really think of a car as having a final

good: nothing that happens to it is really for the sake of the car: usually, it is for the sake of the car's owner. We think of only certain kinds of beings as having a final good, while the notion of well-functioning extends much more widely. Indeed, according to Aristotle's metaphysics, it extends pretty much to anything we can recognize as an entity at all.⁹ But I do not think this is as grave a problem for Aristotle's view as we might at first think, for reasons I have already touched on in the discussion above. The good for an artifact is wholly relative to the good of the being who will use it. In fact, sometimes when we talk about what is "good for" an artifact, it is fundamentally unclear whether we are really talking about something that enables it to perform its function, or something that would give it other properties we would like it to have. This is because it is fundamentally unclear whether we should count, as part of its function, its having all of the properties we would like it to have. A common example of what I have in mind is when we say that something is good for an artifact, meaning that it will enable the artifact to keep functioning and last for a long time. We prefer artifacts that last for a long time, and that makes us think of artifacts rather as if they were organisms, for it *is* part of the function of a living thing to last—that is, to keep itself alive. But self-maintenance is not, or at least not obviously, part of the

⁹ Stones, unless regarded as missiles or pieces of pavement, cannot be well-functioning, nor can topographical entities like mountains. For Aristotle, the stones are not a problem—in an important sense, they are not really entities at all, but "mere heaps" of matter. For a defense see my "Aristotle's Function Argument." I am not sure what to say about topographical entities.

function of an artifact. We even speak of an artifact in these contexts as having a “life.” Using good gasoline, we say, will extend the life of your car. But does that make it better at performing its function, which is serving as a means of transport? We do not need to answer this question, because the whole issue arises simply because the good of artifacts is just a projection of their goodness for us. But the things that are good or bad for an organism really are good or bad for *it*, and not just for us. So an organism really does have a final good in a much deeper sense than an artifact does.

Yet the objection may be pursued further. Do we even want to say that plants have a final good? Many people believe that only beings who are conscious have a final good. It at least seems true that things can be good or bad for conscious animals in a *deeper* sense than they can be for a plant. And it seems extremely plausible to suppose that a conscious being’s final good has something to do with the state of her consciousness. For such a being, as I said earlier, in my discussion of hedonism, we seem to require that her good be something that she can or even does experience as a good. This gives rise to a question. Does the presence of consciousness introduce a sense of “final good” and of “good for” which is simply independent of the evaluative sense of good and the ideal of well-functioning that so naturally accompanies it? Or could some form of well-functioning still be the good for conscious beings considered as such?

VIII. What Difference Does Consciousness Make?

This leads us to the question: what difference does consciousness make? There seem to be three possible views we might take about the way in which the presence of consciousness in a sentient being might affect the character of his final good.

One view is that consciousness introduces a new sense of “final good” that has nothing *intrinsic* to do with a creature’s well-functioning at all. The hedonist’s conviction that the good just must *be* pleasure is grounded in this way of thinking. According to hedonists, when we talk about the good in the sense that is relevant to ethics, the good that utilitarians think we ought to promote, we are not talking about well-functioning at all, but about a distinct kind of final goodness made possible by the existence of consciousness. Of course, well-functioning may, as it happens, tend to the agreeable state of one’s consciousness, but that is not to say that well-functioning itself is *intrinsically* connected to the final good.

Some of the familiar protests against Aristotle’s claim that we find our good in the practice of *moral* virtue spring from the idea that the good has this ineluctably subjective element. Imagine a gentle and generous human being who, in some emergency, sacrifices his life for the sake of others at an early age. He is surely a good person, and in one sense his life was a good one, but we do not feel easy saying that he has thus attained *his own* good or the good *for himself*. His life is not enviable or choiceworthy, although given the circumstances he did

well to choose as he did. At a notorious moment in the *Nicomachean Ethics*, Aristotle claims that someone who sacrifices his life in battle “is getting a great good for himself ” because he prefers “a twelvemonth of noble life to many years of humdrum existence.”¹⁰ Aristotle is trying to establish that a morally virtuous act is always good from the agent’s own point of view, but the claim seems absurd, not least because there is no reason to suppose a virtuous person who fails to sacrifice himself will then have a humdrum existence. While the virtuous choice may indeed always be better than the alternative—for a virtuous soldier, desertion is not an attractive option—it does not seem to follow that you always attain your own good by making a virtuous choice. Although this is not the time to go into it, I do not think Aristotle really needs to draw this conclusion in order support his own theory anyway, both because the good is supposed to be well-functioning *in circumstances conducive to well-functioning*, and because the well-functioning of a human being includes more than moral virtue. That is a longer story. But the general point is that simply identifying the *summum bonum* with a life that is good only in the sense that we would approve of it morally does not satisfy the intuition that the good for a person must be something that that person experiences or can experiences as a good, as something welcome from his own point of view. And that leads people to think that the good must be some experiential thing like pleasure.

¹⁰ *Nicomachean Ethics* IX.8 1169a 25-30.

But if even if we agree that the good for a person must be something welcome from his own point of view, we may still wish to join the many philosophers in the tradition who have resisted the hedonist idea that the good *just is* a certain state of consciousness. The arguments here are familiar. It would be bad, we think, to spend your life hooked up to a so-called experience machine –one that delivers a steady stream of pleasant sensations and imaginary pleasant experiences directly into your brain—and so to live in a dream.¹¹ It would be bad, we think, to be hated by the people whom you imagine love you and despised by the people whom you imagine admire you. It would be bad to imagine that you are doing a great deal of good by actions that are actually creating havoc, or to spend your life carrying out some arduous project destined to collapse like a house of cards shortly after your death. These things are bad even if you are fated never to be cured of your delusions or to know of your failure. For, many philosophers would argue, it is not the case that it is bad to be *aware* that you are hated, or despised, or a failure, or a walking catastrophe to those around you, simply because the *consciousness* of these things is painful; rather, these conditions are objectively bad, and that is *why* the consciousness of them is painful: because it is the consciousness of something bad.

¹¹ The problem of the experience machine was introduced by Robert Nozick in *Anarchy, State, and Utopia*, pp. 42-44.

These reflections give rise to another view we might take of the difference that consciousness makes. Perhaps consciousness does not make *any* difference to what constitutes your final good: perhaps it simply enables you to be aware of whether you have achieved that good or not.

But that does not seem right either. The arguments I just mentioned work by driving a wedge between an agreeable consciousness and a bad reality; but we can also construct arguments that drive a wedge between a disagreeable consciousness and a good reality. Perhaps you are loved by people whom you believe despise you, and perhaps your own efforts, that seem so fruitless to you, are actually setting humanity on its collective feet. Are we to say of someone who suffers from these negative delusions that he had a good life (in the final, rather than the moral sense) but failed to know it? That does not seem right. This consideration brings us back to the idea that the final goodness of a life depends on whether it is perceived or experienced as a good life after all.

I do not think we should resolve this conundrum by picking one side or the other: that is, either by deciding that the good just is agreeable consciousness after all or by deciding that it is something wholly objective of which consciousness merely makes us aware. Rather, I think we should conclude that the concept of goodness, in the final sense of goodness, has a kind of *reflexivity* about it: nothing can be a final good if it cannot be perceived as a final good, and indeed the final good is, as it were, made complete by our perception of it. Someone who is unaware of the goodness of his life therefore actually has a less

good life. Or rather, we might say, the concept of being *an entity that has a final good* is reflexive in this way. Nothing can have a final good that cannot be aware of its final good as such. In this respect, the concept of having a final good parallels the concept of having a personal identity: for as many philosophers have noted, nothing can *have* a personal identity that is at not least potentially aware of itself as having a personal identity. Having a self-conception is not merely having a conception of a self that exists independently of that conception. And in the same way, having a consciousness of the good is not merely having a consciousness of a good that exists independently of that consciousness. I think we should adopt this third view of the relationship between consciousness and the final good. If we are tempted to say that only conscious beings have a final good, it is because having a final good, like having a personal identity, is a reflexive property that only a conscious being can have.

IX. Animals and the Good

Personal identity and the goodness of your life depend on consciousness. These things can only be had by beings who can be aware that they have them. But this is certainly not to say that either the goodness of your life or the character of your identity are whatever you take them to be. They are the awareness of something, even if it is not something wholly independent of that awareness itself. According to Aristotle's theory, what a being that has a final good is aware of it is own well-functioning, of its goodness in the extended-

evaluative sense. But this suggests a very tight connection between the nature of an animal as Aristotle understands it, and the idea of a being who has a final good. For an animal is not merely aware of her own well-functioning: her awareness of her own well-functioning is itself an evaluative awareness. And that is essential to the way she functions. She functions by standing in an evaluative relation to her own well-functioning. She functions by experiencing her good as a good.

Here's what I mean: the distinctive form of life that characterizes an animal involves the maintenance of that very form of life by means of a relationship that obtains between the animal and her own functioning: the animal monitors her own functioning and has positive evaluative attitudes towards the things that promote her functioning and negative ones towards the things that will inhibit it. These evaluative attitudes motivate her to act in ways that promote her good. To put it more simply and intuitively, healthy, well-functioning animals like to eat when they are hungry, are eager to mate, fear their enemies, work assiduously to keep themselves clean and healthy, and so on. (Do not say "well, *of course* they do." Allow yourself to be struck by the fact that there are entities, *things*, that attend in this way to the goodness of their own condition.) What these phenomena show is that the function of an animal is to take care of itself—and nature made that possible by designing the animal *to care about itself*—by which I mean, to enjoy and suffer from her own extended-evaluative condition. On Aristotle's conception, that is not just a fact about

animals: that is what an animal essentially is, something that functions by caring about herself, and how she is doing. Animals have a final good because it is their nature to have evaluative attitudes about their own extended-evaluative condition. And that is what a final good is: a final good is something that constitutes or contributes to the good condition of something that can experience its own condition as a good. That, I want to say, is the Aristotelian theory of the *nature* of the final good. To put it more carefully, Aristotle's theory of the nature of the final good is that a final good is something that constitutes or contributes to the good condition of something that stands in an evaluative relationship to its own condition. Since an animal is essentially something that stands in an evaluative relationship to its own condition, to say that an animal has a final good is a kind of tautology. The two concepts—the concept of a being with a final good, and the concept of an animal—are pretty much co-extensive.

X. The Puzzle Resolved

Now I return to my puzzle. What are we evaluating, when we say that something is good, in the final sense, and in what respect are we evaluating it? We are evaluating the condition of a being who stands in an evaluative relationship to his or her own condition. We might say that the judgment that something is good for someone (some person or animal) in the final sense is *essentially* empathetic, because when we do it we must be viewing the person's or

animal's condition the way the person (or animal) herself *must* view it, namely: evaluatively.

And what gives this evaluation content—the reason why this formula is not empty—is that we are using the extended-evaluative notion of the good to appraise the being's condition. Furthermore, this knowledge of what we are doing when we use the concept of the good may make it possible for us to determine what the content of the good for that being will be, without recourse to mysterious intuitions. We just have to learn how the being functions.

To achieve the good in the *final* sense, then, is to be aware of oneself as being in a good condition in the *evaluative* sense—to be aware of oneself as well-functioning, as the kind of thing that one is. All of this is just an overly analytic way of saying that the final good for an animal is to be conscious of her own healthy life, or more correctly to be conscious of *herself* as healthily alive. In the human case, that includes functioning well in the life of rational choice, with all that that entails. If Aristotle and Kant are right, being well-functioning in the life of rational choice entails being morally well-functioning, so that is part of our good. According to this view, people who express their sense of being in a finally good condition by saying that they *really feel alive* are saying something literal. They feel *their life*, and they feel it as a good to them: and that is something that it is in the nature of an animal in a good condition to do.

XI. Conclusion

The view I have just described is essentially Aristotle's, with a modification in favor of an emphasis on consciousness. This modification seems to me to be necessary to capture the essential subjectivity of the idea of the final good, the reflexive element in the concept of a being that has a final good. I think that the theory so modified captures the element of truth in hedonism, without falling into the characteristic error of utilitarianism. On this theory, everything that is a good is a good to or for some sentient being. The final good is essentially relational, not intrinsic, because it is derived from the evaluative relation in which an animal, by her very nature, stands to herself. Because it is essentially relational, it cannot be added across the boundaries between persons, or other animals.

Another advantage of this theory, when combined with the Kantian theory of value, is that it enables us to explain the existence of value in a naturalistic way. On the intrinsic value theory, various objects, activities, and experiences simply *have* intrinsic value. There is no explanation of why that should be so: it is just a fact. The value of the objects and activities comes first, and there are valuing beings—beings who have a final good—because there are beings who are equipped to get in touch with those values. On the Aristotelian-Kantian theory I propose, the order of dependence between valuing and values goes the other way. Values exist because there are valuing beings, beings who have evaluative attitudes towards their own condition, beings for whom things

can therefore be good or bad. It is because there are such beings—animals—that there is such a thing as the final good.

REFERENCES

Aristotle. *The Nicomachean Ethics*, Translated by W. D. Ross, revised by J. O.

Urmson. *The Complete Works of Aristotle: The Revised Oxford Translation*,
Volume 2. Edited by Jonathan Barnes. Princeton: Princeton University
Press, 1984.

Aristotle. *Physics*. Translated by R. P. Hardie and R. K. Gaye. *The Complete
Works of Aristotle: The Revised Oxford Translation*, Volume 1. Edited by
Jonathan Barnes. Princeton: Princeton University Press, 1984

Korsgaard, Christine M. "Aristotle's Function Argument." Korsgaard, Christine
M. *The Constitution of Agency: Essays on Practical Reason and Moral
Psychology*. Oxford: Oxford University Press, 2008.

Korsgaard, "Kant's Formula of Humanity." Korsgaard, Christine M. *Creating the
Kingdom of Ends*. Cambridge: Cambridge University Press, 1996.

Korsgaard, Christine M. *The Sources of Normativity*. Cambridge: Cambridge
University Press, 1996.

Moore, G.E. *Principia Ethica*. Cambridge: Cambridge University Press, 1903.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974, 2013.