

John Rawls's Theory of the Good

Christine M. Korsgaard

I. Introduction

In *A Theory of Justice*, John Rawls works out a theory of the good, with a view to establishing three conclusions important to his theory of justice. The first is that his chosen unit of distribution, primary goods, are indeed good, and good in the way he defines them to be—they are things that citizens have reason to want no matter what else they want. The second is that it is good-for a person to be a just person, at least if the person lives in a well-ordered society, one that is effectively regulated by a publicly accepted conception of justice.¹ Importantly, Rawls, like Plato before him, aims to show, not that being a just person promotes our interests independently defined, but that being a just person is good as an end, a good thing to be for its own sake. And the third is that a just society in Rawls's sense is also a good society for its citizens to live in—again, not because it promotes our given aims, but for its own sake. Those last two claims—that a just society is good-for its citizens and that it is good to be a just person if you live in one, are needed in order to establish what Rawls call “congruence,” the harmony of the right and the good.² That in turn is necessary to show that a society under Rawls's conception of justice would be, as he calls it “stable”: that is, it would generate its own support, in the sense that those who lived under the Rawlsian system of justice would find reason

¹ I have hyphenated good-for (and bad-for) since one of the issues treated in this essay will be the distinctness of, the relations between, the idea of “good” and the idea of “good-for.”

² Rawls, *A Theory of Justice*. Congruence is discussed in §60 on p. 347 and p. 350, and established in Chapter IX, by the arguments that show that justice is a good-for the just person in a just society and that life in a just society is good.

to support its demands on them. The idea is that a just society could not be stable if the demands of justice were in general at odds with the good of the citizens.³ But we can put what is essentially the same point in more theoretical terms: since the right and the good, according to Rawls, represent two separate domains of value, it is important to moral theory to show that they do not pull us in opposite directions, that their demands do not pull us apart.

At the present moment, there is something of a resurgence of work on the good, and my aim in this paper is to talk about how Rawls's theory, which is not often referred to in the contemporary discussion, fits into that discussion. I believe that Rawls's theory shows throws important light on how to solve some problems to which the contemporary discussion has brought attention. I will not discuss Rawls's specific arguments for the goodness of being a just person and living in a just society, but I will talk about how he thinks such arguments work, and also about some more formal relations between the just and the good. And finally I will discuss whether, or rather the extent to which, Rawls's theory of the good can be part of what Rawls calls "political" as opposed to "comprehensive" liberalism, and what we should conclude if it cannot.

II. A Problem about the Good: the Relation Between Evaluative and Final Goodness

Let's start by talking about the idea of the good. We use the word "good" in several ways. First, "good" is our most general term of evaluation, applicable to almost anything we use

³ In *Theory of Justice*, Rawls discusses the problem of stability in §69, on pp. 398-399. Later, in *Political Liberalism*, Rawls rejects the argument for stability offered in *Theory* as "unrealistic" and "inconsistent with the view as a whole." (Introduction, pp. xvii-xviii).

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, friend; and people considered just as people, among many other things. All of these things may be evaluated as good or bad.

We use the idea of the good in a similar way when we speak of some agent as being good *at* some particular sort of activity. These two uses seem closely connected: a good knife is one that is good *at* cutting, or at least makes the human being who wields it good at cutting; a good car is good *at* transporting people and their baggage around safely, or makes its driver so, and so on.⁴ So I will group those two uses together as what I will call “the ordinary evaluative use of good.”

But we also use “good” in another way, to designate ends that are worth pursuing for their own sakes, lives that are worth living, and states of affairs that are worth bringing about. Love and friendship, beauty, or the experience of beauty, worthwhile work and the sense of self-worth that it gives us, happiness, and pleasurable experiences, are all among the things that have been regarded as good in this sense. I will refer to such things as “final goods,” things that we aim at or value for their own sakes. Although I just contrasted this use of “good” to the ordinary evaluative use of good, most people think the use of “good” to designate final goods is also a form of evaluation. That seems reasonable enough, but it gives rise to a problem.

A natural thought, going all the way back to Plato and Aristotle, about the ordinary evaluative use of “good” is that it has something to do with a thing’s ability to perform its

⁴ For more on why it matters whether we say a knife is good at cutting or that it makes the person who wields it good at cutting, see Korsgaard, *Self-Constitution*, §2.2.2, pp. 36-37.

function, or, though with some admitted circularity, to perform its function well.⁵ The function of a knife is to cut; a good knife is one that is good at cutting, good at that activity. (Or one that makes its user good at that activity, but I will stop adding this rider.) We can also follow Plato and Aristotle by adding that any property that makes the knife good at serving its function counts as a *virtue* of the knife.⁶ A sharp blade, and a handle that makes the knife safe and easy to wield, are virtues in a knife. Finally, we can introduce a third important use of “good” by noticing that any condition that enables the knife to maintain its virtues is *good-for* the knife, while conditions that tend to make its virtues degenerate are *bad-for* the knife. So being sharpened is *good-for* a knife, and being used on the wrong sort of material, say something that will blunt its blade, is *bad-for* the knife. Things are also said to be good- and bad-for people, plants, and animals. How that use is related to the use of “good-for” I have just described is a point to consider. We will be coming back to the idea of good-for in the sense of good-for someone later on.

The functional account of ordinary evaluative goodness appears in Plato's *Republic* and Aristotle's *Nicomachean Ethics* and variants of it reappear in many more recent treatments of the good: In Peter Geach's famous paper “Good and Evil,”⁷ in Judy Thomson's book *Normativity*,⁸ in Philippa Foot's *Natural Goodness*,⁹ and in Rawls's own account, among other places. This is

⁵ See Plato, *Republic* I, 352d-354b; Aristotle, *Nicomachean Ethics*, I.7, 1097b 22-33.

⁶ Plato, *Republic* I, 353b-354. Aristotle, *Nicomachean Ethics* II.6 1106a14ff, *Nicomachean Ethics* VI.2 1139i.18. The Greek is “arete,” sometimes translated “excellence.”

⁷ Geach, “Good and Evil.”

⁸ Thomson, *Normativity*. See especially Chapter II.

⁹ Foot, *Natural Goodness*, especially Chapter 2. Foot sees such judgments as applying in particular to living things. She derives them in the first instance from what Michael Thomson calls “Aristotelian categoricals,” used to describe

hardly surprising, for the account is obvious enough to seem almost banal. But, as I said before, a problem arises when we try to transfer it to talk of final goods. Ends, lives, and states of affairs do not, or at least do not obviously, have a function. In fact to say that something is an end and not a means is to imply precisely that it is not pursued for the sake of something else. It is valued, as we say, for its own sake. When we say that an artefact is a good one, or that someone is good at an activity, the standard of evaluation we use is that they are able to perform their function, or to perform it well. But what is the standard of evaluation we are using when we say of final goods, or ends and lives, that they are “good”? How do we determine which possible ends, or lives, or states of affairs are good, in the sense of being worthy of pursuit or achievement for their own sakes?

Just how serious a problem this is perhaps did not become clear until G.E. Moore published *Principia Ethica* in 1903. With his focus mainly on the question of final goods, Moore argued that goodness is a simple unanalyzable property shared by all finally good things. He had in mind chiefly states of affairs which we might aim to bring about for their own sakes. Two people standing in a loving relationship or a person appreciating a beautiful sunset for the right reasons or someone’s experiencing pleasure—those sorts of things, Moore thought, are intrinsically good. He argued that that meant that they have the simple unanalyzable property of goodness. So to say that they are good is not merely to indicate that they may be evaluated

the life-form of a species, but then limits the categoricals that give rise to judgments of goodness to those that concern the well-functioning of the organism. See pp. 30-31. She sets aside judgments of the goodness of artefacts as “secondary,” if I understand her correctly, because their goodness is primarily goodness for the creatures who use them. See p. 26.

positively but to offer the evaluation in question: they are good.¹⁰ That we cannot simply identify goodness by definition with a natural property, such as pleasure, is made clear by what has come to be called “the open question argument.”¹¹ Someone who asks whether, for instance, pleasure is good is asking a substantial question. To say that pleasure is good is not merely to say that pleasure is pleasure, and one may reasonably wonder whether pleasure is really good, much less *the* good. So it is an open question whether pleasure is good or the good. Moore thinks we can only be asking whether pleasure has the unanalyzable property of goodness.

So here is where that leaves us. On the one hand, we have the idea of ordinary evaluative goodness, which seems to have something to do with an object's ability to perform its function, and on the other hand, we have final goodness, which appears to be a simple unanalyzable property. What makes finally good things good is just that they have the property of goodness, and there is nothing more to be said. What makes ordinary evaluatively good things good, on the other hand, is not that they have the property of goodness, but that they have certain other properties that explain how they can do what they do or do it well. A knife is a good knife not because it has the property of goodness, shared by good cars and good dogs and good perfume, but because it has the properties of being sharp and easy to handle. A friendship or an aesthetic experience, on the other hand, is supposedly good simply because it has the unanalyzable intrinsic property of goodness. Faced with this situation, we might wonder what ordinary

¹⁰ For Moore's account of goodness see *Principia Ethica*, Chapter I; for his account of which things are good see Chapter VI.

¹¹ Moore, *Principia Ethica*, Chapter I, §13, pp. 15-16.

evaluative goodness and final goodness even have in common. Why do we even use the same word, “good,” for both of them?

III. Another Problem about the Good: the Relation Between Good and Good-For

One natural recourse for philosophers faced with such a situation is to try to extend the account of one kind of goodness to cover the other. Should we say that a knife is good not just because it has a sharp blade and a certain sort of handle, but also because it has the property of goodness, or that in virtue of those things it has the property of goodness? The trouble is that something can be a good such-and-such in the ordinary evaluative sense without begin just plain good.¹² We may grant that the unscrupulous man with a steely heart and a steady aim is a good assassin, but we need not therefore concede that he is good.

A more promising line to try is the reverse project: to try to extend the account of ordinary evaluative goodness so that it covers the case of final goodness. Aristotle’s famous function argument in Book One of the *Nicomachean Ethics* was perhaps the first and is still the most famous attempt along these lines.¹³ Aristotle urged that we could discover what a good human life is, if “we could first ascertain the function of man” since “in general, for all things that have a function or activity, the good and the ‘well’ is thought to reside in the function.”¹⁴ In other words, Aristotle thought that if we could assign human beings a function, which explains

¹² Geach does not think there is any such property as “just plain good,” but he points out that you can call someone a “good burglar or a good cut-throat” without commending him. See Geach, “Good and Evil,” pp. 35-36. Rawls makes the same point at *Theory of Justice* §61, p. 354.

¹³ Aristotle *Nicomachean Ethics* I.7 1197b20-1198a18.

¹⁴ Aristotle *Nicomachean Ethics* I.7 1197b24-27.

the distinctive nature of human living, we would be able to say what counts as living well for human beings. Aristotle's argument is better than most of you think, and I have defended it elsewhere.¹⁵ But here our business is with Rawls, whose account of the good is also an attempt to extend the ordinary functional or evaluative account of goodness so that it covers the goodness of lives and other final ends.

Before I can explain what Rawls's argument achieves, however, I need to put one more notion on the table, or rather back on the table. One way that we might argue against extending Moore's account of goodness to cover cases of ordinary evaluative goodness is to point out that in the case of ordinary evaluative goodness, the property of goodness does no explanatory work. What explains the goodness of a good knife is not the property of goodness, but the properties of having a sharp blade and a handle that makes it safe and easy to wield. Recently, Richard Kraut has made a similar argument. The Moorean property of goodness—absolute goodness, as Kraut calls it—is otiose because we can more plausibly do all the explanatory work Moore proposed to do with “good” with “good-for,” in the sense of good-for someone.¹⁶ Friendship, love, and aesthetic experiences are surely good-for people, Kraut argues, and at least if we can find a way to explain that fact without rendering their value instrumental, adding that these things are also absolutely good adds nothing practically important or theoretically illuminating.

In reply to Kraut we might be tempted to argue that the reason *why* these things are good-for people is that they are absolutely good, in the sense that they have Moore's unanalyzable

¹⁵ See Christine M. Korsgaard, “Aristotle's Function Argument.”

¹⁶ Richard Kraut, *Against Absolute Goodness*. See especially Chapter 9, where Kraut argues that we do not count against something bad as were twice over, once for its being bad-for someone and then again for its being absolutely bad.

property. To see why this does not quite rescue Moore's account, it is helpful to examine an argument of Moore's own. Moore himself argued that if we accept the idea that goodness is a simple unanalyzable property, we must reject the idea of something's being good-for someone in particular.¹⁷ Moore's target was egoism: he wanted to deny that a state of affairs can be good-for one person and bad-for another. A thing cannot have a simple, intrinsic, unanalyzable property relative to some particular person and not to another. It would be like saying metal is solid for you but liquid for me. To put the point metaphysically, if goodness is an intrinsic property, as Moore's view implies, then Moore supposes it is a monadic and not a relational one. Good-for would have to be a relational property.

It seems staggering to suggest that we should have to give up the familiar notion of something's being good-for someone. But according to Moore, all we can mean by saying that something is good-for someone is either that it is an intrinsically good thing and the person has it, or that his having it is an intrinsically good state of affairs.¹⁸ The first of those two options seems to be the idea we are looking at here: that friendship, for instance, is good-for you because it is a good thing in itself and you have it. But Moore does not tell us what conception of "having" he has in mind in here. In many ordinary contexts, when we say someone "has" something, we mean that he is its legal owner (he has a house in Florida) or that it is under his control (the Patriots have the ball; George has the car right now). But neither of these ideas seems to be what we have in mind when we say that something is good-for someone: that it is good and he owns or controls it. Of course, it can *in fact* be good-for you to own something good

¹⁷ Moore, *Principia Ethica*, Chapter III, §59, p. 98.

¹⁸ Moore, *Principia Ethica*, Chapter III, §59, p. 98.

or have it under your control, since then you can get the benefit of it whenever you please, but that you own it or control it is not what we *mean* by saying that the thing is good-*for* you. Rather, it is good-*for* you to own or control it, *if* it is, because it is good-*for* you, or you can use it to get something that is.

Now of course you will want to protest that Moore did not mean “having” in one of these familiar senses. Surely we can say, that, for instance, pleasure, or friendship, are the kinds of things people “have” or “don’t have,” that they are good things, and therefore that they are good-*for* you when you are the one who has them? And we can do that, because “having” is also a word we use to describe our relation to conditions which we may or may not be in. The trouble is that the use of the same word we use for the other sort of “having” may give the impression, often false, that the thing that is “had” can exist independently of the “having” relation, the way a piece of land can exist independently of being someone’s property. We speak of having experiences, having thoughts, having reasons, having friendships, having colds, and having pains and pleasures. But whatever we mean by those expressions, we do not mean that there are unclaimed experiences, thoughts, reasons, pains, pleasures, colds or friendships floating about which may come into our possession or under our control, or somehow come to be our own. What Moore must mean is that things like “having a pleasant experience” or “having a true friendship” are good conditions to be in, and it is good-*for* you when you are the one who is in them. But that leaves it open that what we mean when we say that they are good is not that they have a simple unanalyzable property of goodness, but that they have the internal relational

property of being good-for whoever happens to be in them.¹⁹ And provided we can give an account of what we mean by good-for, that is a much more plausible thing to say. To see this, you need only consider more modest uses of good-for.²⁰ When we say that a carrot is good-for a rabbit, we do not mean that carrots are intrinsically good things and it is therefore good-for a rabbit when she has a carrot. On that showing, it would also be good-for a cat to have a carrot, but it is not. What we mean rather is that carrots are in general good things for rabbits, since they answer to something in the nutritional needs of rabbits, and a carrot is therefore good-for this particular rabbit. In the same way we might say that friendship and beauty are good things for people, because they answer to something in the nature of people, and that therefore it is a good thing for you when you have them.

With that in view, now suppose that being in a true love relationship is a good thing for whoever is in it, and that I am in it, but only because I have stolen away your sweetheart. Is this good-for me but bad-for you? Moore's other possible reading of good-for is that someone's having something is an intrinsically good state of affairs. Is it an intrinsically good state of affairs that I am in a true love relationship? Moore argued that a state of affairs is intrinsically good if it is good when considered all by itself, as if, he extravagantly suggested, it were the only thing in

¹⁹ I suggest that, contrary to what Moore thinks, things may be "intrinsically" good in virtue of internal relations in Korsgaard, "Two Distinctions in Goodness." See especially pp. 270-271.

²⁰ I explore the relation between what I above call "the more modest sense of good-for"—good-for as in "healthy"—and the sense of good-for under discussion here, which means something like "constitutes some creature's good," in Korsgaard, "On Having a Good."

the universe.²¹ This appears to mean we must consider my romance independently of its deleterious effects on yours, so why not? If so, the possible intrinsic goodness of certain states of affairs has no weight against egoism. You can admit that my romantic condition is a good thing in itself, even that it is a good thing in itself because it is good-for me, and still think it is bad-for you. There is, admittedly, still an issue to be settled about where this leaves practical reason: does its badness-for you give you a good reason to oppose my romance or not? I do not mean to settle that question here. The point is that if we can give an account of final goods in terms of what is good-for people, we do not need to regard the property of goodness as unanalyzable: we can analyze it in terms of good-for.

So here we have two issues for the theory of the good. First, what is the relation between ordinary evaluative goodness and final goodness? And second, what is the relation between something's being good and its being good-for someone? With those questions in view, I now turn to Rawls's theory.

IV. Rawls: Goodness as Rationality

Rawls calls his theory "goodness as rationality." According to Rawls's theory, to say that something is good is to say that it has the properties that it is rational to want in that kind of thing. He develops his account in three stages.²²

²¹ In *Principia Ethica*, Moore argues that when we consider something's goodness we must consider it "in absolute isolation" (Chapter III, §53, p. 91) and "absolutely by itself" (Chapter VI, §112, p. 188). The idea of a universe consisting entirely of the good thing is invoked on Chapter 6, §119, p. 197.

²² Rawls, *A Theory of Justice*, §61, pp. 350-351.

At the first stage, we consider the everyday conversational use of the term “good,” in expressions such as “good knife” or “good car.” The everyday use presupposes that the person who is doing the wanting wants the thing for the reasons for which people usually want things of that kind. When we say that a certain make of car is a good one, for example, we mean that it has the properties it is usually rational to want in a car, given what cars are usually used for, say a safe means of transport for people and their various forms of baggage. A Honda, someone says in general conversation, is a good car, safe and efficient.

At the second stage, we relativize the notion to a person’s particular circumstances, including any special interests the person might have. As Rawls understands it, this involves relativizing the notion to the person’s plan of life. Given that a certain person’s plan of life involves spending a lot of time traveling during which he essentially lives out of his car, a minivan, we say, is a good car for him. Notice that at this point, the special status of primary goods falls out of the theory pretty much by definition: they are things that it is rational to want whatever else you want, that is, relativized to anyone’s life plan. So they are good-for anyone.

At the third stage, we add that the person’s plan of life is itself rational. After all, it seems clear enough that we should not say something is rational for you to want just because it fits well with your plans, if your plans themselves are irrational.

This move requires Rawls to define what rationality means for plans of life.²³ He offers a two-part account of this. First of all, your plan of life must be one of the plans open to you that is consistent with the principles of rational choice, or as Rawls also calls them “the counting

²³ Rawls, *A Theory of Justice*, §63, pp. 358-365.

principles.”²⁴ Second, of those plans, it is the one that would be chosen by you with what Rawls calls “full deliberative rationality.”²⁵ Let me briefly say what all this means. Rawls mentions three principles of rational choice or counting principles, though he suggests there may be others. The first is a principle of taking effective means: your plan must effectively promote your ends. The second, rather surprising, principle directs us to prefer more inclusive plans, that is, ones that achieve more different ends. If one plan effectively promotes ends A through F and another just as effectively promotes A through F and also G, you should choose the second plan. What makes this principle surprising is that the additional end “G” need not be among the things you already, independently, wanted: rather, the principle tells us that it is rational to want more different ends—or, as Rawls puts it, to prefer “the development of wider and more varied interests”²⁶—as long as we have just as good a chance at achieving the additional ends. Rawls defends the principle of inclusiveness by appeal to a psychological principle he calls “the Aristotelian principle,” according to which human beings enjoy the exercise of our realized or developed capacities.²⁷ Evidence for the Aristotelian principle comes in the form of its ability to explain certain facts about human beings. It explains why people enjoy learning. It explains why those who have mastered a more inclusive and developed skill normally prefer exercising it over less inclusive and developed skills—why once you learn chess, you lose interest in checkers, say. The

²⁴ Rawls, *A Theory of Justice*, §63, p. 364.

²⁵ Rawls, *A Theory of Justice*, §63, p. 359.

²⁶ Rawls, *A Theory of Justice*, §63, p. 363.

²⁷ Rawls, *A Theory of Justice*, §63, p. 364. Rawls also appeals to the idea that our happiness depends on the proportion of our aims that are realized; this proportion is likely to go up if we effectively pursue a more varied plan. See 363-364.

Aristotelian principle favors the principle of inclusiveness because having more ends will develop more of our capacities more fully, and given our psychology, this will make us happier. A third principle Rawls mentions is that of greater likelihood: if one plan is more likely to achieve several of the ends it includes and not less likely to achieve others, it is to be preferred.

These principles of rational choice, Rawls thinks, will not single out any one possible plan as the most rational. For that, we must add that the plan is the one that would be chosen with “full deliberative rationality,” which Rawls defines this way: “It is the plan that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his most fundamental desires.”²⁸ Deliberative rationality is a way of uncovering what we care about most, by imagining, as best we can, in light of true information, what it would be like to lead the various possible lives that are open to us.

There is of course a great deal that one could question about all this. Rawls does not give any argument for his principles of rational choice. He thinks they are fairly intuitive as principles governing short-term plans and can be defended as principles governing plans for a whole life. The defenses he gives sometimes depend on speculative claims about human psychology, such as his claims about the Aristotelian principle, which he says gives us a “higher-order desire” to follow the principle of inclusiveness. He is apparently not concerned about the oddity of making what is supposed to be a principle of rationality dependent on a psychological tendency. (For comparison, imagine making the principle of instrumental rationality depend on whether human beings have a psychological tendency to *enjoy* achieving our ends.) Rawls does not offer to

²⁸ Rawls, *A Theory of Justice*, §64, p. 366. Rawls attributes the idea to Sidgwick.

derive these principles from some more general account of what rationality is. Instead he proposes that “these principles are to be given by enumeration so that eventually they *replace* the concept of rationality.”²⁹ (I have no idea why that seemed to Rawls to be a desideratum.) The principles of rational choice, Rawls tells us, are not, like the principles of justice, chosen—or to put it in later terms, they are not constructed.³⁰ This is not because the account of the good is dogmatic, but rather because it is not: for political purposes, at least, Rawls points out, there is no need for all of us to agree on what exactly constitutes rational choice, because we need not all have the same conception of the good.³¹ So there is no need for all of us to agree on what the principles of rational choice are.

Of course, if we are viewing Rawls’s conception of the good as part of what he later called a comprehensive philosophical conception—that is, as a genuine metaphysics of the good, not merely as part of a political conception, this seems rather unsatisfactory.³² As philosophers we want to know what the good is so we can work out what is good, and for that purpose we do not want to be told there is no fixed conception. Notice that Rawls is not saying merely that different things are good-for different people, but that people may legitimately have different

²⁹ Rawls, *A Theory of Justice*, §63, p. 361.

³⁰ On constructivism, see *Political Liberalism*, Lecture III, §1, pp. 90-99.

³¹ Rawls, *A Theory of Justice*, §68, p. 392. While Rawls later identified *A Theory of Justice* as giving a “comprehensive” rather than a “political” conception of the good (*Political Liberalism*, Introduction, xviii), this remark seems to belong to a political conception. See note below.

³² For the distinction between political and comprehensive doctrines, see Rawls, *Political Liberalism*, Lecture 1, §2, pp. 11-15.

conceptions of the principles that determine what is good-for them, which is a much more radically relativistic claim.

Once we add in the idea of deliberative rationality, worries also arise about whether conformity to the principles of rational choice or counting principles is constitutive of the rationality of plans or merely a heuristic for arriving at possible rational plans. Rawls makes it constitutive—he says that the rational plan for a person is the one “among those consistent with the counting principles and other principles of rational choice once these are established” that the agent would choose with full deliberative rationality.³³ But suppose, for instance, that when you imagined a possible life in light of the facts as clearly as you could, you found that you would not enjoy a more inclusive life as much as a less inclusive one? Or, more radically, suppose the life that most appealed to you from the standpoint of deliberative rationality turned out to have little to do with the desires from which you started? Rawls seems aware of this worry, for he stipulates that the agent who is to engage in deliberative rationality “is under no misconceptions as to what he really wants.”³⁴ But why should we suppose that we need deliberative rationality in order to get a clear fix on how much we want things, but not in order to get a clear fix on what we really want?

V. How Rawls Solves the Problems about Goodness

For all these reasons, I think Rawls's account of the goodness of rational plans has some serious problems as it stands. But for now I invite you to leave these worries aside. For what is

³³ Rawls, *A Theory of Justice*, §64, p. 366.

³⁴ Rawls, *A Theory of Justice*, §64, p. 366.

interesting and powerful about Rawls's account is the way that it enables him to address the problems we have been discussing. Start with the puzzle of how ordinary evaluative goodness and final goodness are connected. In ordinary cases, the idea of something's being rational to want captures the same content that the idea of something's being good at its function does, so Rawls's theory easily captures the ordinary evaluative sense of good. Since knives are ordinarily wanted for cutting, a good knife is a sharp one, and if you want a knife for the usual reasons, sharpness is a property it is rational to want in your knife. So instrumental or functional properties, broadly speaking, coincide with the properties it is rational to want. But, with Rawls's idea of a rational life plan in hand, we can extend the idea of "rational to want" to a person's ends. A person's ends are rational to want when they form part of his most rational plan of life. Earlier, when I asked by what standard we are to judge final ends good, I denied that things we want for their own sakes have a function. But in effect, Rawls's theory *does* assign our ends a kind of role or function. Their role or function is to serve as an element in a person's rational plan; and some of them are better than others at playing that role. It is therefore possible to evaluate the goodness of final ends in a way that is continuous with our judgments of the goodness of ordinary objects. Although I won't repeat the argument here, it is this that enables Rawls to argue that, at least if you live in a well-ordered society, being a just person would be something to be valued as an end. Rawls shows this by showing how it fits in with the other ends we are likely to have.³⁵ We can also judge a person's life to be good when it is successfully executed in accordance with a genuinely rational plan.

³⁵ Rawls, *A Theory of Justice*, §61, pp. 353-354.

But there's another way to characterize Rawls's move here that I think is even more illuminating. I have now described the ordinary evaluative notion of the good in two different ways: first, as invoking the Platonic or Aristotelian idea that something has the properties that enable it to serve its function or to serve it well; second, as invoking the Rawlsian idea that something has the properties it is rational to want in that kind of thing. These two ideas are almost the same, but there is a slight but important difference of emphasis. When we think of the object's goodness in terms of its being one that is rational for some agent to want, it becomes clear that we are considering its functional properties *from the point of view* of someone who wants that sort of thing.³⁶ This is what enables Rawls to establish a continuity between the ordinary evaluative notion of the good and the final good, since the final good as Rawls conceives it is also characterized from the point of view of the one whose good it is. That is, it is characterized as what it is rational *for that* person to want, given his rational plan. Rawls's theory of goodness is, in the first instance, a theory of good-for.

³⁶ Speaking of the ordinary evaluative use of "good," Rawls says, "There always stands in the background a point of view from which an artifact, functional part, or role is being appraised..." (*A Theory of Justice*, §61, p. 353). My point is that this carries over to the idea of deliberative rationality—it is meant to capture what is good from the point of view of the person whose good it is. Stephen Darwall and T. M. Scanlon have both pointed out to me that for this point to be a substantial one, we need some criterion for distinguishing what "is good from X's point of view" from "what X thinks is good." In my own view, to make the judgment that X is good, you have to empathize with the person whose good is in question, and see it through her eyes, while for her it appears, directly, as desirable or in some way appropriately motivating. So the "being desirable (or in some way appropriately motivating) from a point of view" is, in a way, prior to the judgment of goodness. (See Korsgaard, *Fellow Creatures*, §2.1.9, p. 22.) I am not sure whether Rawls would agree with that.

Earlier I pointed out that we say something is good-for X when it enables X to maintain its virtues, and I asked whether good-for in the sense of good-for some person or animal is the same notion. I think Rawls's theory shows that they are not exactly the same. Rawls's view captures an essential element of *subjectivity* in our notion of the good: that what is good-for a person must also seem good to her, at least when she is considering it correctly. This does not mean that *whatever* seems good to an agent is good-for that agent, for it is a condition of deliberative rationality that the agent is knowledgeable and has imagined things correctly. Recall that Rawls says he conceives the person who engages in deliberative rationality as not being confused about what he really wants—what Rawls elsewhere calls his “fundamental desires.”³⁷ Rawls does not say what he means by “fundamental desires.” Although this is a stretch, if we suppose that a person's fundamental desires in some way reflect his objective needs, then what Rawls's conception brings out is something like this. Our notion of good-for involves a harmony between an agent's objective needs and the way things seem to her. Someone for whom these were regularly at odds would not be capable of having a good.³⁸ To say that what is good-for you is good from your point of view is to say that it is answerable both to your objective needs and to the way things seem to you. I think his account, though suggestive, leaves it unsettled how exactly this is related to the other notion of good-for, good-for in the sense of sustaining a thing's virtues.

³⁷ Rawls, *A Theory of Justice*, §64, p. 366.

³⁸ I do not mean to suggest that this is an impossible condition. At the conference for which this paper was written, a member of the audience asked about the case of an addict, to whom his drug seems good. If the addiction were incurable, we might well say that the addict had become incapable of having a good. There is no condition that both meets his objective needs and is satisfying to him.

VI. Rawls and Geach

In any case, since Rawls's theory is a theory of good-for, it follows from Rawls' theory that everything that is good is good-for someone, for some creature whose nature gives it a point of view. Notice that this claim is not the same as the claim, made by Peter Geach and others, that "good" is used, in Geach's terms, "attributively" rather than "predicatively."³⁹ Geach famously argued that "good" functions like, say, "big," in the sense that you cannot conclude from the fact that something is a big mouse that it is big. Nothing is "big" simply; "big" is attributive because it always invokes a standard of comparison. In the same way, Geach thinks that "good" must always be used in a way that invokes some sort of standard of evaluation, a standard normally supplied by the thing's function. Yet in fact Geach's own examples include both things that are "good-for" in the evaluative or functional sense (knives are good for cutting) and things that are "good-for" in this person-relative (or creature-relative) sense. For instance, Geach mentions ordinary examples of evaluative goodness like "good car" and "good book," but he also tells us that Caesar's murder was "a bad thing to happen to a living organism" but "a good fate for a man who wanted divine worship."⁴⁰ But what standard of evaluation does Geach think is invoked by the idea of good-for in the person-relative sense? Geach does not say. Rawls gives us the standard—consistency with someone's plan, where that plan is rational from the agent's own informed point of view.

³⁹ Geach, "Good and Evil," p. 33.

⁴⁰ Geach, "Good and Evil," p. 41

Followers of Geach often argue, as against G. E. Moore, that nothing is good simply—in Geach's language, nothing is good in the predicative sense, or absolutely good.⁴¹ Relatedly, they often argue that we can make no sense of ideas like “good thing” or “good event” or “good state of affairs” because the terms “thing” “event” or “state of affairs” are too general to convey any specific standard of evaluation. Things, states of affairs, and events do not have some function they can serve well. So we cannot say that something is a good event or a good thing. Since saying that something is a good thing or a good event is as close as we can come to saying that something is just plain good or absolutely good, followers of Geach propose to reject the idea that there is any such property as being just plain good or absolutely good.⁴²

But I think Rawls's theory does makes it possible to say that some events and states of affairs are something very like good absolutely—at least, and I am afraid this is an important qualification—if we limit our judgments of goodness to what is good-for human or rational beings.⁴³ This is where it becomes important that Rawls distinguishes the what he calls the “thin” theory of the good, which is what I have been laying out so far, from a “full” theory of the good,

⁴¹ Judy Thomson, for example, says “...there is no such property as goodness just as there is no such property as bigness.” *Normativity*, p. 11. Foot gives a critical discussion of the consequentialist's use of “good state of affairs” in her “Utilitarianism and the Virtues.” Kraut does not quite deny that what he calls “absolute goodness” exists; he rather denies that we need appeal to such a property in practical reasoning. He does not consider the possibility I mentioned earlier and am here canvassing again—that something might be absolutely good because it is good-for someone. This is because he takes the idea of “absolute goodness” to imply that the goodness is impersonal. See *Against Absolute Goodness*, p. 10.

⁴² Geach, “Good and Evil,” p. 41

⁴³ See Korsgaard, *Fellow Creatures*, §8.8.3, pp. 152-154; §12.1.1, pp. 215-216.

in which the good is restricted by the principles of the right.⁴⁴ In the full theory, your plan does not count as a good one if involves a violation of the principles of justice, which, it is important to remember, we are taking to be principles upon which everyone agrees. But if someone's rational plan of life falls within this restriction, then there seems to be no reason not to say that things that are good-for her are good absolutely, because the judgment that they are good can be supported from every human being's point of view.

Unfortunately, Rawls cannot draw this conclusion in what I at least, as a Kantian, would regard as the most satisfying way. Earlier I mentioned that Rawls proposes that an enumeration of what he takes to be intuitive principles of rational choice should replace the concept of rationality. It is also true, of course, that Rawls separates the ideas of the "reasonable" and the "rational," where the reasonable involves the willingness to propose and abide by fair terms of cooperation, and the rational is a matter, roughly speaking, of pursuing our ends intelligently.⁴⁵ Somewhat oddly, Rawls traces the distinction to Kant, to whom it seems to me to be rather alien.⁴⁶ In any case, Rawls does not attempt to derive the "reasonable" and the "rational" from a common source in reason. If we take the "reasonable" to be roughly equivalent to the categorical imperative and the "rational" to be roughly equivalent to the hypothetical imperative, then we could say that Kant does trace them to a common source in reason. And if we could do that—if we can say, as Kant does, that being moral, or being just, is a *way* of being "rational,"—then we could say that the definition of goodness in terms of the rationality of wanting gives us

⁴⁴ Rawls, *A Theory of Justice*, §60, pp. 347-348

⁴⁵ Rawls, *Political Liberalism*, Lecture 2, §1, pp. 48-54.

⁴⁶ Rawls, *Political Liberalism*, Lecture 2, §1, pp. 48-49 fn. 1.

something close to the absolute sense of goodness. We could say that a state of affairs is good absolutely if it is one that it is rational for anyone, or at least any rational being, to want.

VII. Rawls and Aggregation

In light of that argument, I want to make one more point about the implications of Rawls's approach. One of the main targets of the arguments by Geach and others against the idea of a good state of affairs or a good event is utilitarianism. Utilitarians have to believe that certain events or states of affairs are good since they think it is our duty to promote those. Although I have argued that Rawls's theory does make it possible to characterize states of affairs or events as good, I certainly don't mean to suggest that Rawls's theory of the good makes it possible to defend utilitarianism. On the contrary, it follows from Rawls's theory that everything that is good must be good-for someone, and the idea that everything that is good must be good-for someone blocks utilitarianism, because it blocks the idea of aggregation. What is good-for me plus what is good-for you is not, taken just like that, good-for anyone in particular, and therefore is not good. As Rawls himself famously pointed out, "utilitarianism does not take seriously the distinction between persons."⁴⁷

This gives rise to a question of what we are to do with those of our intuitions that seem to favor the idea of aggregation. Aren't there cases in which we plainly choose actions because they do more good, even when more than one person's good is involved? Obviously, I cannot give a complete treatment of this difficult subject here. But it is worth noting that even if we accept the idea that what is good or bad must be good- or bad-for someone in particular, there are two kinds

⁴⁷ Rawls, *A Theory of Justice*, §5, p. 24.

of cases in which it still seems to be intelligible to claim that we “do more good,” even though more than one person is involved.

First of all, we can say that we “do more good” by choosing a course of action that benefits additional subjects, *so long as no one is harmed* by that course of action. This thought covers the simple case of saving two lives in preference to saving one, for example. Second, and a little more controversially, it seems intelligible to claim that we “do more good” by giving a resource to the subject who will benefit from it the most, so long as no one has a prior claim on the resource and therefore we are not harming the other subjects among whom we are choosing. This thought would cover the case, for example, where we think we “do more good” by giving a painkiller to the subject who is in the most pain. (This second point is more controversial because it assumes we can make intelligible interpersonal comparisons between how much good something does for one person and the same thing does for another.) What most obviously becomes unintelligible on the view that good and bad are always goodness- and badness-for someone is the idea that we can “do more good” by balancing the good of one subject against the good of another, say by taking pleasure away from Jack because that way we can give an even greater pleasure to Jill. That is good-for Jill but bad-for Jack, and if the goodness or badness must be goodness- or badness-for someone, that is all there is to say: there is no one for whom the situation is better overall, and therefore no sense in which it is better.

The interesting thing about these points is that they coincide with some ideas that play an important role in Rawls's account of the right. The first is the legitimacy of making moves that are, in economists' jargon, “Pareto optimal”: moves that leave everyone better off and no one

worse off.⁴⁸ This is the ground on which we move from the first thought that Rawls thinks would naturally occur to us in the original position—namely, distribute everything equally—to the idea that an unequal distribution is preferable if it would make everyone better off. Notice that the move from equality to a position that makes someone better off without making anyone worse off is licensed by the first of the two thoughts I mentioned: that it makes sense to “aggregate” goods across the boundaries between persons when we are just adding goods for more people and not subtracting. And—this is the second idea—among all the possible Pareto optimal positions, the choice of that of the worst off as the one with respect to which we should maximize is licensed by the second thought I mentioned—that when we are distributing a resource to which no one has a prior claim, we do more good by giving it to the one to whom it will do the most good. Thus in both of the two senses of “doing more good” that remain to a theory of the good that insists that all good must be goodness-for someone, the distribution that accords with the difference principle is *better*—that is, it does more good—than any other.

Two points of warning. First of all, the plausibility of these claims depends on our keeping in view something that Rawls believed the friends of aggregation tend to forget—namely, that whether a distribution counts as harming someone is relative to the position in which they begin. You are not made worse off by a way of distributing things simply because there is some other imaginable distribution of things under which you would be better off. In the original position, it is clear what the starting point or benchmark is—it is an equal distribution. So those whom a certain distribution will place in the middle and upper ranks cannot claim to be harmed merely by the fact that we decide to maximize with respect to the position of the worst

⁴⁸ Rawls, *A Theory of Justice*, §12, pp. 58-63.

off rather than with respect to those other possible positions. And for the same reason, a distribution that left those who are worst off under it with less than an equal distribution would have done *can* claim to be harmed by it. It is a theme in Rawls's work that utilitarians tend to reason as if distribution were an act from on high, the act of an administrator trying to maximize the amount of utility he can produce by distributing resources to which no one has a prior claim. As Rawls puts it in "Justice as Reciprocity," this image of distribution compares it to that of "an entrepreneur deciding how much to produce of this or that commodity in view of its marginal revenue."⁴⁹ And, as Rawls goes on to point out, it envisions social distribution as if "the individuals receiving these benefits are not...related in any way: they represent so many different directions in which limited resources may be allocated."⁵⁰ Begging many questions, we might suppose that someone who has decided to give a certain sum of money to charity is situated this way: no one has a prior claim on her money and she is free to allocate it in what she regards as the most effective way.⁵¹ But actual distributions are almost never like that: they involve people who are related in some way and who have some claims going into the process. In the main case for Rawls, the case of the distribution of primary goods in society, they are related as fellow citizens with a basic claim to be treated equally, and the distributive scheme is not something someone else imposes on them, but something they choose together.

Now I am claiming that there is a clear sense in which we can claim that a distribution in accordance with the difference principle is better, that is, does more good, in Rawls's sense of

⁴⁹ Rawls, "Justice as Reciprocity," p. 216

⁵⁰ Rawls, "Justice as Reciprocity," p. 217.

⁵¹ Actually, I don't think this is the right way to think of charity. See Korsgaard, "The Claims of Animals and the Need of Strangers: Two Cases of Imperfect Right," especially §9.

good, than other possible distributions. But, importantly, I am not claiming that this is what makes it *right*. What makes it right is not that it does the most good—what makes it right is that it is the distribution that the parties in the original position would agree to. Of course, the original position is designed so that the parties in it choose what is best for themselves. And the veil of ignorance guarantees that each of them cannot but choose on behalf of the representative citizen considered as such. So there is nothing surprising in the fact that the choice of a distributive scheme turns out to be the best one from the point of view of the representative citizen. The original position is set up to yield that result. My point, though, is that even when the veil of ignorance is lifted, and we know what sort of position we occupy in society, someone who accepts Rawls's theory of the good, or indeed any theory that insists that everything good must be good-for someone, can still see a clear sense in which distributions in accordance with the difference principle are better than others. In that sense, there is a formal harmony between the right and the good in Rawls's theory.

VIII. Conclusion: Goodness as Rationality: Political or Metaphysical?

Goodness as Rationality is spelled out in detail in a *Theory of Justice* and for the most part merely referred to in Rawls's later works. I would therefore like to end with some reflections on whether, or to what extent, the theory survives as a part of political liberalism when Rawls begins to distinguish a political from a comprehensive form of liberalism. Rawls himself seemed to assume that for the most part it would, since in *Political Liberalism* he says:

The ... idea... of goodness as rationality ...is, in some variant, taken for granted by almost any political conception of justice.⁵²

At first glance this claim may seem surprising. Philosophers in modern times have an ongoing argument about whether the good rests, substantively speaking, in the satisfaction of desire, or in pleasure and the absence of pain, or in perfection, or in a so-called “objective list” of intrinsically valuable conditions and activities. Rawls’s account seems straightforwardly to fall in the “satisfaction of desire” camp. Indeed Rawls seems to admit as much when he says: “To be sure, there is one formal principle that seems to provide a general answer. [He means a general answer to the question which plan is the best.] This is the principle to adopt that plan which maximizes the expected net balance of satisfaction.”⁵³ So, substantively, Rawls’s account seems controversial. Furthermore, in this paper I have portrayed Rawls as taking a stand on some deep and controversial issues concerning the metaphysics of the good: in particular that the idea of good-for is prior to the idea of the good. I have suggested that it is this idea that stands behind Rawls’s opposition to aggregation and explains the goodness of the distributive scheme that the parties in the original position would choose. But if Rawls’s “goodness as rationality” is a controversial theory, both substantively and metaphysically, even among philosophers, how can we suppose that all reasonable comprehensive conceptions can accept “some variant” of it?

The answer lies in what Rawls goes on to say to illustrate his point:

“This idea [he means goodness as rationality] supposes that the members of a democratic society have, at least in an intuitive way, a rational plan of life in the

⁵² Rawls, *Political Liberalism*, Lecture V, §2, pp. 176-177.

⁵³ Rawls, *A Theory of Justice*, §64, p. 365.

light of which they schedule their more important endeavors and allocate their various resources...so as to pursue their conceptions of the good over a complete life, if not in the most rational, then at least in a sensible (or satisfactory), way.”⁵⁴

In other words, the uncontroversial part of the theory is just the idea that we all share a conception of the good life as involving some way of assembling our aims into a plan that will allow us to achieve the most important of them to the fullest possible extent. We can take “desire” and “want” to refer, not to some item in our natural psychology, but to that pro-attitude, whatever it might be, that we have towards our aims. And we can imagine getting those aims from whatever source we like, including our natural psychology, or our expectations of pleasure, or a list of objective values derived from Moorean intuitions. In that case, goodness as rationality is sufficiently uncontroversial for use in political liberalism.

I am a fan of political liberalism, so far as the theory of liberalism and its commitments goes. I think Rawls was right to make the turn to political liberalism that he did. But that is no reason to reject the project of doing “comprehensive” philosophy about deep and old-fashioned questions about the substance and metaphysics of the good. We may be interested in the results of such philosophical explorations even if we do not believe we are justified in insisting that every member of a liberal society must accept them. And for those of us who are interested in substantive and metaphysical questions about the good, Rawls’s theory offers some powerful insights: about the relation between the goodness of ends and the goodness of ordinary objects, about how we might go about showing that something is good as an end, and most of all about

⁵⁴ Rawls, *Political Liberalism*, Lecture V, §2, p. 177.

the important and neglected relation between something's being good and something's being good-for someone.⁵⁵

References

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

Urmson) in *The Complete Works of Aristotle: The Revised Oxford Translation*.

Edited by Jonathan Barnes. Princeton: Princeton University Press, 1984. References to Aristotle's works are given by the standard Bekker page, column, and line numbers. In this translation, the word I translate as "function" (ergon) is translated "work."

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

Foot, Philippa. "Utilitarianism and the Virtues." in Foot, *Moral Dilemmas*. Clarendon Press, Oxford: 2002, pp. 59-77.

Geach, P. T. "Good and Evil." *Analysis*, Vol. 17, No. 2 (December 1956), pp. 33-42.

Korsgaard, Christine M. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press, 2018.

Korsgaard, Christine M. "On Having a Good." in *Philosophers of Our Times*, ed. Ted Honderich. Oxford: Oxford University Press, 2015.

⁵⁵ An earlier version of some of the material in this paper appeared in "The Unity of the Right and the Good in John Rawls's Thought," presented at a symposium on "The Legacy of John Rawls," at the meetings of the Eastern Division of the American Philosophical Association in December 2012. I would like to thank the organizers of both that session and the conference for which the present paper was written, the audiences on both occasions, and Charlotte Brown for comments on the manuscript of this version.

Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009

Korsgaard, Christine M. "The Claims of Animals and the Needs of Strangers: Two Cases of Imperfect Right." *The Journal of Practical Ethics*, Vol. 6, No. 1, June 2018.

Korsgaard, Christine M. "Two Distinctions in Goodness: in *Creating the Kingdom of Ends*. New York: Cambridge University Press, 1996, pp. 249-274.

Kraut, Richard. *Against Absolute Goodness*. Oxford: Oxford University Press, 2011.

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

Plato. *Republic*. Translated by G.M.A. Grube, revised by C.D.C. Reeve, in *Plato: Complete Works*. Edited by John M. Cooper. Indianapolis: Hackett Publishing Company, 1997. References to the Republic are given using the standard Stephanus numbers which appear in the margins of most translations of Plato's works.

Rawls, "Justice as Reciprocity" in Samuel Freeman (ed.), *John Rawls: Collected Papers*. Cambridge, MA: Harvard University Press, 1999, pp. 190-224,

Rawls, John. *Political Liberalism*. New York: Columbia University Press, 1993.

Rawls, John. *A Theory of Justice*. First edition 1971; Revised edition, 1999. Cambridge, MA: Harvard University Press, 1999. Page references are to the revised edition.

Thomson, Judith Jarvis. *Normativity*. Chicago: Open Court Publishing, 2008.