

Thin and Full Theories of Goodness

Rawls first draws the distinction between the thin and the full theories of goodness in section 60 of *A Theory of Justice*. The first and perhaps the most important thing to note about the distinction is that it is a distinction between two *theories* or *accounts* of goodness. It should not be conflated, as it sometimes is, with a very different distinction that is drawn by some of Rawls's critics. Communitarians sometimes oppose 'thin' to 'thick' and predicate these terms of *conceptions* of the good endorsed by members of Rawls's well-ordered society or by members of the actual societies with which we are familiar. But this is not the distinction Rawls is drawing. He opposes 'thin', not to 'thick', but to 'full' and he applies both adjectives to accounts of what makes something good rather than to conceptions of the good.

The thin and full theories differ in two respects. First, they differ in the resources on which they draw. The full theory "takes the principles of justice as already secured" (*TJ*, p. 349) and so draws on those principles to account for goodness, while the thin theory does not. This difference leads to the second difference, which is one of scope. The thin theory can account for the value of some goods, such as the primary goods. Indeed, Rawls says that he introduces the thin theory precisely to account for the value of those goods (*TJ*, p. 347), though we shall see later that the thin theory can account for the value of other goods as well. Because the full theory draws on more philosophical resources than the thin theory does, it can account for values that the thin theory cannot. In particular, because the full theory draws on the principles of justice, it – unlike the thin theory (see *TJ*, p. 381) -- can account for various kinds of moral goodness. Among these is the goodness of the morally worthy person, who is good because she "has the moral sentiments that support adherence to" the principles of justice. (*TJ*, pp. 384)

Though the thin and the full theories are different accounts of goodness, they are not unrelated. Rawls says that the full theory is a "develop[ment]" of the thin theory. (*TJ*, p. 382) Indeed, he might have gone further. Instead of saying in section 60 that he was "distinguish[ing] between two theories of the good" (*TJ*, p. 347), he might have said that he was drawing an important distinction within a single overarching account of goodness. To see why he might have said this, it will be useful to recall the main point of Rawls's overarching account of goodness in *TJ*, an account he calls "goodness as rationality".

Goodness as rationality is a "rational-desire" account of goodness. The central thesis of such accounts can be put, most crudely, as the claim that the good is the object of rational desire. Something is a good instance of its kind if it has the properties it is rational to want in things of that kind, given what things of that kind "are used for or expected to do". (*TJ*, p. 351) To use one of Rawls's examples, a particular timepiece is a good watch if it has the properties it is rational to want in watches, given what watches are used for. But sometimes we speak of an object, not as good example of the kind of thing it is, but as good *for* a person or as part of his good. Goodness as rationality can be used to explicate this kind of goodness as well. Something is good for a person if (i) it has the properties it is rational for him to want in things of that kind, given his aims and (ii) his aims are organized into a plan of life that is rational given his circumstances and abilities. (*TJ*, pp. 350-51) Note that the way goodness as rationality accounts for a thing's goodness and for something's being good for a person are both "information-sensitive". What it is rational to want in things of a given kind can only be determined in light of information about what a known population normally uses of things of that kind for. What it is rational for someone to want depends upon what he knows – or perhaps what he ought to know -- about his own aims and how to achieve them.

Both the thin and the full theories of goodness use this rational-desire account of goodness, and so both give accounts of what makes something good and of what makes something good for a person that conform to its pattern. The differences between the thin and the full theories are differences between two accounts of goodness that apply the same template. The first of the differences mentioned above was that the full theory draws on greater philosophical resources than the thin theory. Since goodness as rationality is information-sensitive, that difference can now be re-described this way. When the full theory accounts for the goodness of things and of persons, and for the goodness of things for persons, it uses Rawls's rational-desire account of goodness, but it uses a richer informational base than the thin theory does. The informational base of the full

theory includes the principles of justice. The informational base of the thin theory does not. This difference between the informational bases of the two theories accounts for the second difference between them, the difference in scope. To see this, it will be useful to start by seeing how the thin theory accounts for the value of some of the goods that fall within its scope, the primary goods.

Rawls says that he introduces the thin theory of the good to secure “assumptions about the parties’ motives in the original position” and, more specifically, about their motives to choose one principle or set of principles rather than another. In asserting the need for a theory of the good to secure these assumptions, Rawls may seem to be saying that if the parties are to have any motivation at all to choose, then there must be some things which appear good to them. But on rational-desire accounts of goodness, the goodness of something, or our perception of its goodness, does not explain the desire for it or account for the rationality of desiring it. Rather, to say that something is rationally desired by someone just is to say that it is good for him. So what Rawls means is that if parties in the original position are to be motivated to make a choice, there must be some things it is rational for them to want which they can get by adopting principles. Because of the veil of ignorance, they cannot want those things in light of their circumstances, abilities, aims or plans. If they are to make a choice, there must therefore be things it is rational for them to want regardless of their aims and plans or, as Rawls says, some things “it is rational to want ... whatever else is wanted”. (*TJ*, p. 380) Those things are the primary goods. The rational desire to secure those goods motivates the parties’ choice of principles of justice.

Thus the account of the value of the primary goods to parties in the original position follows the pattern of goodness as rationality. The primary goods are good for parties because it is rational for them to want them. Moreover, the account reflects the information-sensitivity of goodness as rationality, since the primary goods are good for the parties in the original position because they want them in light of what they know about their aims and plans. The original position is merely the limit case of this information-sensitivity, since the parties have no information about their own aims and plans at all. Finally, because the account of the value of the primary goods does not appeal to the content of Rawls’s principles of justice -- as it cannot if the primary goods provide the motivation for choosing among principles -- that account is clearly provided by the thin theory of goodness.

In writings after *Theory of Justice*, Rawls gave increasing prominence to the claim that free and equal moral persons are thought of as having two moral powers: a capacity for a sense of justice and a capacity for a conception of the good. At the same time, he began to cite different reasons when explaining why parties in the original position want primary goods. Whereas in *Theory of Justice*, he said that primary goods are things “it is rational to want ... whatever else is wanted”. (*TJ*, p. 380), in these later writings he described the primary goods as things which “are generally necessary as social conditions and all-purpose to enable human beings to realize and exercise their moral powers and to pursue their final ends”. (*CP*, p. 314) Parties in the original position are then said to want the primary goods because of the interest they know they will have in developing and exercising their moral powers when they step out from behind the veil.

While this revision may constitute a significant change in what Rawls says about why the primary goods are wanted, it did not require a departure from his use of goodness as rationality to account for the value of primary goods for the parties in the original position. For in later writings as in *Theory of Justice*, Rawls insisted that “the parties’ preference for primary goods is rational”. (*CP*, p. 314) This insistence allowed him to maintain, as he had in *Theory of Justice*, that the primary goods are good for the parties because it is rational for them to want them. Rawls’s account of their goodness remained information sensitive. In stressing the parties’ interest in conditions and means needed to develop and exercise the moral powers, Rawls was merely adding to their information base by adding general information about their moral psychology. Because the information he added did not include or presuppose the principles of justice, Rawls enriched the parties’ information base while staying within the thin theory.

Of course, primary goods are not good only for parties in the original position. They are good for members of the well-ordered society too. Their goodness for members of the well-ordered society can be accounted

for in the same way as their goodness for parties in the original position. But many other things are thought to be good, and are good for members of the well-ordered society. An especially important case of goodness is the goodness of the morally worthy person. We will need to see how Rawls accounts for that.

We saw earlier that according to goodness as rationality, something is a good instance of its kind if it has the properties it is rational to want in things of that kind, given what things of that kind “are used for or expected to do”. (*TJ*, p. 351) If goodness as rationality is to account for the goodness of the morally worthy person, it must follow this template. We saw then that the account of a things goodness which is provided by goodness as rationality is information sensitive because it depends upon the normal uses or expectations of a known population. The use of goodness as rationality to account for moral worth therefore depends upon the identification of the relevant population. The population Rawls chooses is, not surprisingly, the membership of a well-ordered society. And so he says that “a good person has the features of moral character that it is rational for members of a well-ordered society to want in their associates.” (*TJ*, p. 384) The goodness of the morally worthy person is accounted for by the rational desirability of those features.

Is this a plausible account of moral worth? In a well-ordered society, everyone knows she has a sense of justice and is inclined to abide by the constraints. It is therefore rational for each member of a well-ordered society to want others to abide by those constraints too rather than to take advantage of her. (*TJ*, p. 385) And so it is rational for each to want others to have a sense of justice and the other qualities of character that ease compliance with the demands of justice. That is, “the features of moral character that it is rational for members of the well-ordered society to want in their associates” are those qualities of character that we intuitively regard as the moral virtues. Of course, the contents of the principles of justice themselves must express what we regard as plausible demands, but Rawls thinks his do. Provided he is right about this, goodness as rationality can plausibly account of the goodness of the morally worthy person.

We have seen that Rawls also uses goodness as rationality to account for goodness *for persons*. If we think that it is good for members of the well-ordered society that their compatriots be just, then goodness as rationality must be able to account for that value as well. It should now be clear how it does so. It is rational for members of the well-ordered society to want their compatriots to be morally worthy because each knows that she has a sense of justice and, therefore, that her aims include the aim of being a just person herself. So the moral worth of others is something it is rational for each member of the well-ordered society to want in light of information about her aims and ends. It follows from goodness as rationality that having morally worthy compatriots is good for her.

Rawls’s accounts of the goodness of the morally worthy person and of the value to members of the well-ordered society of having morally worthy compatriots – unlike his accounts of the goodness of the primary goods and of their value to the parties in the original position – appeal to the contents of the two principles of justice. They are therefore provided by the full rather than the thin account of goodness. But despite this difference in their informational bases, the former accounts – like the latter -- follow the template of goodness of rationality and reflect its information-sensitivity. It seems that the full theory can therefore be described as one part of goodness as rationality. As if to confirm this, Rawls speaks at one point in *Theory of Justice* of “the full theory of goodness as rationality”. (*TJ*, p. 381)

Later, in *Political Liberalism*, Rawls equated goodness as rationality with the thin theory (*PL*, p. 178) We might take this remark to express what Rawls thought all along and conclude that he never thought the full theory was part of goodness as rationality. But another possibility is that as Rawls’s thought developed, he came to see that the distinction between the thin and the full theories was one he no longer needed, and that he dropped the full theory altogether while retaining goodness as rationality to account for the value of the primary goods. To see whether this is so, it will helpful to ask why the full theory and the distinction between it and the thin theory were important to Rawls when he wrote *Theory of Justice*.

We have already seen that with an informational base that includes the principles of justice, Rawls thinks he can account for the goodness of the morally worth person and the value to members of the well-ordered

society of having morally worthy compatriots. He also thinks that with the principles of justice on hand, he can explicate other important ethical concepts such as ‘beneficence’, ‘benevolence’ and ‘supererogation’ and show why persons and deeds that answer to them are good. (*TJ*, pp. 381-85) Because Rawls’s treatments of beneficence, benevolence and supererogation appeals to the principles of justice, his account of their goodness – like his account of moral worth -- are provided by the full theory.

As with moral worth, so with persons and deeds that are beneficent, benevolent and supererogatory, Rawls needs the full theory to give a plausible account of their moral value. (See *TJ*, p. 381) If he did not appeal to principles of right and drew only on the thin theory, he would not be able to say what kinds of acts the benevolent or beneficent person actually performs or to claim plausibly that the actions classified as supererogatory are morally good ones. Thus “[d]eveloping the thin theory into the full theory via the original position is,” Rawls says, “the essential step” in providing an account of moral goodness. (*TJ*, p. 382) Accounting for the moral value of beneficence, benevolence, supererogation and moral worth is something Rawls wants to do, since showing that the contract view can give an account of their moral goodness lends the contract view necessary credence (see *TJ*, p. 349) and helps to make good on the promise that it can be extended beyond a theory of justice to “an entire ethical system”. (*TJ*, p. 15)

There is another task that the distinction between the thin and full theory is needed for, a task that is more central to the project of *Theory of Justice*. Rawls says that the “central aim” of much of part III “is to prepare the way to settle the questions of stability and congruence.” (*TJ*, p. 347) Both the full and the thin theories of the good are used to help settle those questions.

Rawls gives a two-part answer to the question of whether justice as fairness would be stable. (*TJ*, p. 397) In the first part, laid out in *Theory of Justice*, chapter VIII, Rawls argues that members of the well-ordered society would acquire a sense of justice. Each would therefore have a strong desire to act from principles chosen in the original position. Those principles are, Rawls says, “collectively rational”. (*TJ*, p. 497) But what is collectively rational might not be individually rational. If it were rational for individual members of the well-ordered society to defect from the agreement reached in the original position or to uproot their sense of justice, then the well-ordered society will be undermined by well-known collective action problems. So in the second part of his treatment of stability, brought together in section 86 of *Theory of Justice*, Rawls argues that members of the well-ordered society would have and realize that they have strong rational desires to preserve their sense of justice. Preserving their sense of justice is, he argues, something that must be included in their rational plans. And since according to goodness as rationality, each person’s good consists in the fulfillment of his rational plan, preserving the sense of justice would be – and would generally known to be -- good for each. Rawls concludes that “the right and the good are congruent” (*TJ*, p. 450). Then “no reasonable and rational person in the well-ordered society is moved by rational considerations of the good not to honor what justice requires” (*CP*, p. 487 note 30) and the well-ordered society would be stably just.

So described, Rawls’s line of thought leaves open the question of what informational base members of the well-ordered society draw on when they judge that it is rational to preserve their sense of justice.

Suppose that that base includes their knowledge that they have a strong and rational desire to act from the principles that would be chosen in the original position. Then they would know that being just persons is something they strongly wish to do. Since that desire is strong and rational, it is a desire that should be included in their plans and its fulfillment would be good for them in the full sense ‘good’. But in that case, the conclusions that they should plan to preserve their sense of justice and that doing so belongs to their good would be “obvious” and the argument for those conclusions would be “trivial”. (*TJ*, p. 498)

If Rawls is to provide a non-trivial argument for those conclusions, the notion of goodness at work in the second part of the stability argument cannot be that accounted for by full theory. That is why he says in *Theory of Justice*, section 60:

...when we ask whether the sense of justice is a good, *the important question is clearly that defined by the thin theory*. We want to know whether having and maintaining a sense of justice is good (*in the thin sense*) for persons who are members of a well-ordered society. (*TJ* p. 350, emphases added)

To see how Rawls uses the thin theory for this purpose, note first that objects of desire are intentional objects, individuated by their descriptions. If, for example, the same referent is desired under three different descriptions, then there are three desires and three objects of desire. It may be that the description of one of those objects appeals to the content of the principles of justice, so that the rationality of desiring it depends upon the rationality of wanting to satisfy the principles. In that case, the value of that object is accounted for by the full theory. But if the other two objects are given by descriptions which do not appeal to the principles of justice, then their value – like the value of the primary goods – can be accounted for by the thin theory.

This is the case with a sense of justice. For according to Rawls's contract view, the phrase "a sense of justice" picks out a referent which can also be described as "a desire to act from mutually justifiable principles" and "a desire to act from principles that would be chosen in a contract the outcome of which is determined by our nature as free and equal rational beings". There are objects of desire answering to each of the three descriptions. The rationality of wanting the first, "a sense of justice", depends upon the rationality of wanting to honor Rawls's principles. Indeed, to desire a sense of justice just is to want to be the kind of person who wants to honor them. That is why the good of satisfying the desire is accounted for by the full theory. But the other two descriptions do not appeal to the content of the principles, and the rationality of wanting the objects they pick out does not depend on rationality of honoring or of wanting to honor the principles of justice. Their value can therefore be accounted for by the thin theory and they fall within its scope .

Rawls argues that members of his well-ordered society would want to act from mutually justifiable principles and would want to act from principles which they would agree to as free and equal rational beings. The objects of those desires are given by descriptions which have the same referent as the phrase 'sense of justice'. What Rawls calls a "practical identity" therefore holds among the objects of all three desires. (*TJ*, p. 501) Members of the well-ordered society would therefore have what are, in effect, desires to maintain their sense of justice. Moreover, since members of the well-ordered "have a lucid grasp of the public conception of justice" (*TJ*, p. 501), they would be aware of the practical identity and aware that what they want – albeit under a different description – is to maintain their sense of justice. So even "when they assess their situation independently from the constraints of justice" (*TJ*, 350), they would include the satisfaction of that desire in their rational plans. By the central claim of goodness as rationality, satisfying that desire therefore belongs to their good.

This argument – which greatly compresses Rawls's reasoning in *Theory of Justice*, section 86 -- shows that the right and the good are congruent, and hence that the well-ordered society would be stably just. Yet the question of whether someone who wants to act from mutually justifiable principles and who wants to act from principles chosen in the right sort of social contract also wants to have and maintain a sense of justice is not a "trivial" question and it does not have "an obvious answer". Rather, it is a problem the solution to which depends upon the details of Rawls's theory. Clearly, his argument that a "practical identity" holds between the objects of the first two desires and the object of third is crucial to the provision of an answer which is philosophically interesting. Because the value of those two objects is accounted for by the thin theory, the question of whether attaining them – and for that reason having and maintaining a sense of justice -- is good for members of a well-ordered society is a question that uses 'good' "in the thin sense". It is a question "defined by the thin theory".

Thus the distinction between the thin and the full theories is a distinction between two theories that follow the template of goodness as rationality, but that differ in their informational bases and scope. In *A Theory of Justice*, Rawls uses the full theory to explicate various kinds of moral goodness, showing how moral worthiness and other moral values fall within its scope. He used the distinction between thin and full, and the

scope of the thin theory, to argue for congruence and stability. In the years following the publication of *Theory of Justice*, his concern to present justice as fairness as a political conception of justice eclipsed his interest in showing how contract theory could be developed into “an entire ethical system”. (*TJ*, p.15) The full theory went into eclipse along with this interest. Rawls also offered a different account of stability, one which did not rely upon the distinction between the thin and the full theories. Perhaps these changes explain why in later work, Rawls identified goodness as rationality with the thin theory and restricted its employment to just one of its uses in *Theory of Justice*: that of accounting for the value of the primary goods to the parties in the original position.

Paul Weithman
Department of Philosophy
University of Notre Dame