
Essays

ON JOHN RAWLS'S *A BRIEF INQUIRY INTO THE MEANING OF SIN AND FAITH*

Paul Weithman

ABSTRACT

This essay challenges the view that John Rawls's recently published undergraduate thesis *A Brief Inquiry into the Meaning of Sin and Faith* provides little help in understanding his mature work. Two crucial strands of Rawls's *Theory of Justice* are its critique of teleology and its claims about our moral nature and its expression. These strands are brought together in a set of arguments late in *Theory* which are important but have attracted little sustained attention. I argue that the target of Rawls's undergraduate thesis is a form of Christianity which rests on assumptions Rawls later came to think were fundamental to teleological views, and that the thesis defends an alternative form of religiosity that anticipates what Rawls says in *Theory* about the expression of our nature. Those sections of *Theory* also provide resources Rawls could have used to respond to a number of prominent and recurrent criticisms of his account of moral motivation. Seeing the continuities between *Brief Inquiry* and *Theory of Justice* shows how long Rawls wrestled with problems he took up in the neglected sections of *Theory* and thereby shows their importance to Rawls's thought.

KEY WORDS: *John Rawls*, Brief Inquiry, Theory of Justice, Political Liberalism, *hedonism*, *moral motivation*

IT IS NOW COMMONLY ACCEPTED that John Rawls's undergraduate thesis *A Brief Inquiry into the Meaning of Sin and Faith*¹ shows him to have been much better informed about, and much more sympathetic to, religion than was previously thought.² It is also widely recognized that *A Brief Inquiry* anticipates some of the ideas found in *A Theory of Justice*

Paul Weithman is Professor of Philosophy at the University of Notre Dame. He is the author of a number of articles in political philosophy and is a previous contributor to the *Journal of Religious Ethics*; his most recent book is *Why Political Liberalism? On John Rawls's Political Turn* (Oxford, 2010). Paul Weithman, Department of Philosophy, 100 Malloy Hall, University of Notre Dame, Notre Dame, IN 46556, weithman.1@nd.edu.

¹ This essay was drafted for a symposium on Rawls 2009 that was convened in December 2010 at the Eastern Division meeting of the American Philosophical Association. I am grateful to Erin Kelly for the invitation to participate in the symposium and to Hilary Bok for reading the paper when weather kept me from attending. I am also grateful to members of the Moral Theology Colloquium at Notre Dame—especially Gerald McKenny, Jean Porter, and Maura Ryan—for helpful comments on an earlier draft, and to two anonymous referees for the *JRE* for their comments on a later one.

² For especially clear statements, see Gregory 2007, 183 and Wolterstorff 2010, 84.

[*TJ*]³ and beyond.⁴ There is, however, a considerable divergence of opinion about whether the thesis advances our understanding of Rawls's mature work.

Some readers have claimed that the Rawls of *TJ* failed adequately to support some of his most fundamental arguments. With the thesis in hand, they say, we can now see why Rawls thought those arguments were successful, since the arguments can be made good by religious premises which Rawls openly avowed in *Brief Inquiry* and which, these readers maintain, he tacitly continued to hold (Berkowitz 2009; Galston 2009). Jürgen Habermas has recently asserted an important connection, not between the undergraduate thesis and *TJ*, but between the thesis and Rawls's eventual presentation of justice as fairness as a political liberalism (Habermas 2010, 450–52). Many readers, however, have denied that there is any scholarly significance to the thesis at all.⁵ It is this last view which I wish to challenge.

Rawls's thesis has a substantial critical component which is directed against a "scheme of thought" that he calls "naturalism" and against the version of Christianity that he thinks is indebted to it. I shall argue that what Rawls found objectionable about the target version of Christianity was its commitment to particular claims about human nature and about the expression of human nature in human life, a commitment shared with a philosophical view targeted by Rawls in *TJ*. The sections of *TJ* in which Rawls attacks those claims are §§83–85, which deal with hedonism, dominant ends, and the unity of the self. Readers generally neglect these sections, as they do a good deal of Part III.⁶ One of the effects of *Brief Inquiry* is to suggest that they deserve attention that is much more sustained, since the thesis shows that those sections respond to concerns which Rawls regarded as important enough to wrestle with

³ All references are to the revised edition of 1999.

⁴ As noted by Thomas Nagel and Joshua Cohen in their "Introduction" to Rawls 2009, as well as in Bok 2009.

⁵ For example, in a generally sympathetic review of *Brief Inquiry*, Jonathan Harmon writes, "I will not dwell too long on the arguments of the thesis, as I believe the benefits to the Rawls scholar of reading it are mostly indirect: not necessarily in what Rawls says, but in what it says about the young Rawls" (Harmon 2011, 735). Gilbert Meilaender says of the positive views Rawls advances in *Brief Inquiry*: "One may, as Cohen and Nagel do, note certain continuities and discontinuities between them and views developed by the mature Rawls, but there's not much else to be gained from them" (Meilaender 2009, 50). And David Schaefer is dismissive of the whole, concluding his review of the thesis by saying that "the greatest value of *A Brief Inquiry* for students of Rawls's thought may lie in its revelation of how the characteristic limitations of his outlook were present from early on in his development" (Schaefer 2010, 278).

⁶ To give only one example of many, see the glancing references to §§83–85 at Sandel 1982, 151, 156, 158, and 162. Thomas Nagel vigorously criticizes Sandel for his misinterpretation of a crucial passage in *TJ* §85 in Nagel 2006, 47n3 and the associated text.

from the early 1940s, when he wrote the thesis, until the 1970s, when he published *TJ*.

Brief Inquiry also includes two lengthy constructive chapters in which Rawls sketches the form of Christianity which he thinks should replace the target view. These chapters are fascinating enough in their own right, but I shall pay less attention to them than to the critical parts of the thesis. For my purposes, what is significant about the constructive chapters is not their theological detail, but the views about human nature and its expression which underlie them. An important but neglected argument in *TJ* §86 rests on the claims that members of a just society would want to express their nature as moral beings and that they can satisfy that desire only by living ongoing social lives which are regulated by principles of right. These claims of *TJ* are anticipated in the constructive sections of *Brief Inquiry*. While the relevant claims are not well-developed in the thesis, they—like the critical parts of the thesis—point toward premises and arguments in *TJ* that are eclipsed by other parts of that work. Once we see the main points of §86, we will be able to see how Rawls would respond to prominent and recurring criticisms of his account of moral motivation. Moreover, seeing the assumptions that underlie §86 is important for understanding the development of Rawls's thought after *TJ*. Part of the significance of *Brief Inquiry* is that it reminds us of the importance of this neglected section.

To sustain my reading of the thesis and my assertion of instructive continuities between *Brief Inquiry* and *TJ*, I need both to be more specific about the version of Christianity against which the thesis is directed and to locate the target in the text of the thesis. Since the target is a version of Christianity that Rawls says has been led astray by what he calls “naturalism,” the road to a clearer understanding of the target version must go by way of a fuller account of what Rawls means by that term. I shall follow this route, first trying to understand naturalism and only then making my way to naturalist Christianity.

1. Naturalism and Natural Relations

In Rawls's hands, the term “naturalism” names a family of views. Members of this family are not materialist views or views according to which everything can be explained by natural science. Rather according to one of Rawls's characterizations, they are views “in which all relations are conceived of in natural terms” (2009, 119). As we shall see, the phrase “*all* relations” is too strong. An immediate indication of this is that when Rawls defines natural relations, he defines them as two-place. Natural relations, he says, are relations “between a person and some object insofar as personality is involved in the relation” (2009, 114). This definition seems to imply that the naturalist cannot countenance relations of more

than two places, and the implication is almost certainly false. But I shall ignore this difficulty and try to understand naturalism by understanding natural relations.

As we have just seen, the definition of natural relations says that one of the relata in such a relation is a person, but the “insofar” phrase in the definition is important. I can enter into a relation with something by bumping into it inadvertently. This is presumably a relation in which my personality is not involved. The relation in which I stand in virtue of bumping into it is therefore not a natural one. It is what Rawls calls a “causal relation,” a kind of relation he mentions only to put aside (2009, 114). I believe Rawls thinks someone can endorse naturalism in his sense while acknowledging that persons can enter into causal relations. So Rawls would have to modify his claim that naturalists conceive of *all* relations as natural. Instead he should say that “naturalism” names a family of views “in which all relations *in which personality is involved* are conceived of as natural relations, as relations between a person and some object.”

What of the other relatum in a natural relation? By “object,” Rawls cannot mean to denote just physical objects or particular metaphysical substances, since he says that objects include processes (2009, 180) and probably states of affairs (compare 2009, 150). Later we will need to look closely into what objects are. For now, it is important only to note that Rawls contrasts objects with persons, in that the former are said to lack the powers of personality. Objects are, he says, “impersonal” (2009, 115–16, 180).

It is important that the second relatum in a natural relation need not actually *be* an object. It need not actually *be* something which lacks personality. It is enough that it *be treated* as something which lacks personality by the person in the relation. This conclusion helps us to understand what Rawls means by “conceived of” when he says that naturalistic views are views “in which all relations are *conceived of* in natural terms” (emphasis added). The naturalist is not committed to the view that the only relations that persons as such enter into are with things that are in fact objects and so lack personality. Rather, what the naturalist thinks, according to Rawls, is that all the relations persons as such enter into—that is, all the non-causal relations they enter into—can be described or “conceived of” as natural relations, and hence as relations between a person and an object. They can be described as or conceived of that way because, even when persons as such enter into relations with other persons, they enter into them as if they were entering into relations with objects.

Thus we might say that according to the Rawls of the undergraduate thesis, the naturalist thinks that human beings objectify everything with which we enter into relationships, even persons. Relationships that entail

objectification are the only kind of relationship that human beings are capable of entering into insofar as our personality is involved. This is why I believe that naturalism as Rawls understands it is best interpreted as involving a core thesis about the nature of human personality—a thesis about human nature. As I read *Brief Inquiry*, the ethical mistakes of naturalism and naturalist Christianity stem from the error of this core psychological thesis. Let me now try to confirm this interpretation.

2. Naturalism and Motivation

So far, I have explicated naturalism by beginning with one of Rawls's characterizations of it, the characterization which defines naturalism in terms of the *relations* that are entered into by human beings as such, and by moving from that characterization to a conclusion about human nature. But Rawls also characterizes naturalism as committed to a thesis about motivation, for he implies that according to naturalism, intentional human action is motivated by what he calls "desire" (see Rawls 2009, 119–20). Sometimes, Rawls runs the relational and motivational characterizations of naturalism together. On the first page of the thesis, he says "Naturalism is the universe in which all relations are natural and in which spiritual life"—by which I take it Rawls means something like "activity involving the powers of personality"⁷— "is reduced to the level of desire and appetite" (2009, 107).⁸

Rawls's use of the phrase "reduced to" in this last characterization suggests that he thinks naturalism unjustifiably explains the dynamics of spiritual life in terms drawn from some other level of human experience where they can appropriately be used. Insofar as the suggestion implies that Rawls thinks there *is* a level or a subset of human experience in which we *are* moved by appetite, this suggestion is right. Just as Rawls would not object to the claim that human beings can enter into *some* natural relations, so he would not object to the claim that we are *sometimes* moved by "desire and appetite"—as both he and the naturalist understand those terms. When he describes desire, Rawls is describing a motive that both he and the naturalist think plays a role in human

⁷ See Rawls 2009, 111: "Personality is equivalent, perhaps, to what we mean by 'spirit.' When we speak of spiritual life, it seems that we mean personal life. Personality and spirit . . . we shall use interchangeably[.]"

⁸ Later, he says that "appetitional desires are the energies of all natural relations" (2009, 178). One reason the second characterization is surprising is that we might expect Rawls to say that "spiritual life is reduced to the level of desire *and aversion*." But while Rawls has much to say about naturalism and desire, he has almost nothing to say about naturalism and aversion; an exception is the passing reference on 115–16. I believe this is because Rawls is interested in the implications of naturalist psychology for the human good. Since I am interested in his critique of these implications, I shall ignore the complication here.

life. What he objects to is naturalism's extension of this analysis of human motivation beyond its proper bounds, so that "desire and appetite" are given much greater psychological prominence, and much greater ethical significance, than they merit.

It may be surprising to find Rawls using the relational and motivational characterizations of naturalism interchangeably, since the two are not obviously coextensive. But if what Rawls meant by saying that naturalism conceives all relations as natural relations is that the naturalist thinks we objectify everything with which we enter into non-causal relationships, then we can see how the two characterizations can be coextensive after all. They can be coextensive if desire and objectification are appropriately connected. And Rawls seems to imply that they are. He writes, "Appetite for us means the impulse or striving for *any* object whatsoever. The criterion of appetite is that it seeks some object" (2009, 180). Since Rawls thinks we can have a desire or an appetite for persons (2009, 123, 187–88) and since persons are not objects, his claim about the criterion of appetite must be read as saying that appetite seeks things *as* objects or, more straightforwardly, that to desire something is to objectify it. And so by desiring something, the desirer enters into a natural relationship with it.

Clearly, if we are to understand naturalism and to see where Rawls thinks it goes wrong, we have to understand desire and objectification, and to see why Rawls thinks one entails the other. Rawls seems to provide us some help near the end of the thesis, where he says that desire "is controlled by an attitude of seeking and getting. Desire leads us to acquire something" (2009, 250). This echoes a conclusion Rawls reached earlier, in which he said that for the naturalist "all love is acquisitive" (2009, 178). These passages suggest that to desire something in Rawls's sense entails treating it as something that can be acquired or that I can come to possess. They suggest, that is, that to desire something is to objectify it by treating it as a possible object of *possession*. Rawls seems to confirm the suggestion in a passage in which he adds a further element or dimension to desire: desiring something must entail treating it as a possible object of *my* possession, for Rawls says that "[d]esire is egoistic; it seeks some object *for the self*" (2009, 250, emphasis added). Since desire establishes a natural relationship, we can see how Rawls reaches a conclusion that is merely asserted early in the thesis, namely, that "natural relations are egoistic" (2009, 118).

But while much of what Rawls says about desire, acquisition, egoism and naturalism can be tied together in a satisfying way, some of his remarks about acquisitiveness remain very puzzling. For one thing, they raise the question of what we want to acquire things *for*. For another, Rawls thinks I can desire to engage in "concrete process[es]" (2009, 180) such as exercise and sleep (2009, 180, 184). But it is not at all clear how

I can want to possess or acquire exercise and sleep. Moreover, as we have seen, Rawls also thinks that persons can desire other persons and he considers the possibility that there is an appetite for God. Indeed, as we shall see, he seems to think that naturalist Christians like Augustine think there *is* such an appetite. But what could it mean to say that we want to acquire another person or that we treat God as a possible object of possession?⁹

I think these latter questions arise because we are misled—in ways Rawls himself invites or fails to anticipate and correct—by the way we ordinarily think of acquisition: as resulting in the exclusive control over something's use and disposition. The egoistic element of acquisition thought of in this way is then expressed when I say that the thing over which I have acquired control is, at least *de facto*, "mine." But when Rawls says that we desire something or want to acquire it, I think he is most charitably read as zeroing in on something else that follows from acquisition: when we acquire something, we no longer lack it and so the desire stemming from that lack is satisfied. If that is right, then—though Rawls does not say this—we have to take possession, which results from acquisition, as a weak relation. I possess something if I stand to it in a relation which brings the satisfaction of desire.

That Rawls is concerned with the feature of acquisition that I have isolated is suggested by a remark he makes about pure appetite. "A pure appetite," Rawls says, "would be characterized by the concentration of the activity of consciousness upon the object of the desire and upon *the expected state of relief once the object was acquired*" (2009, 184, emphasis added). But Rawls does not just think that relief is one of the foci of the desiring person's consciousness. Early in the thesis, he states, "The whole phenomenon of desire, though it seems to include personality to a degree, moves along the level of biological causation, and the *end* desired *is an impersonal state which uses the object as a means only*" (2009, 117, emphases added).

If the impersonal state to which Rawls refers is "the expected state of relief" that we enjoy "once the object is acquired," then this passage implies that when I desire some object, what I really or ultimately want—my final aim in seeking it—is relief or satisfaction. Thus if I desire hot coffee, coffee is the proximate object of my desire. The desire for it is acquisitive and egoistic. What I ultimately want is something that follows from the acquisition of the coffee: satisfaction or relief of the desire to be slaked or warmed. That desire is *my* desire, and I desire the coffee as a

⁹ Indeed, Augustine is often said to claim that all human love is acquisitive (Fitzgerald 1999, 511). The accusation, and attempts by Augustine's defenders to absolve him of it, can be puzzling since not all the things Augustine includes among the objects of love are things that it is possible to acquire, at least in the usual sense of "acquire."

means to *my* satisfaction or relief. It is important that Rawls does not say desire ultimately aims at the *pleasure* of satisfaction. And so while desire is egoistic, it is not hedonistic. Rather Rawls seems to think that desire aims at a different state, one which he does not characterize with any precision: the state of relief or desire-satisfaction itself. Desire is, we might say, not hedonist but quietist.

This reading helps us to see what Rawls thinks we want to acquire things for, how he might think we can have an acquisitive desire for God and processes, and why he thinks that acquisitiveness is egoistic. Now recall Rawls's insistence that when I desire something, I enter into a natural relation with it, and that natural relations obtain between a person and an object. We saw earlier that when Rawls speaks of an object in this connection, he must mean "something objectified." I believe he thinks that the desire for something as a means to satisfaction or relief entails such objectification because it entails regarding or treating the thing desired as something without the powers of personality. Desire entails such treatment because, Rawls thinks, the powers of personality are not what enable the desired thing to bring satisfaction. That is true even if the thing desired is a person, as can happen when appetite is sexual (2009, 187–88).¹⁰ So the state of satisfaction that results from the acquisition of an object is not impersonal because the desirer is absent or impersonal. It is impersonal because the object acquired is not treated as a person.

We have seen that according to naturalism, all the relations that we enter into are natural relations, from which it follows that we are always moved by desire in the naturalist's sense of that term. Now that we have seen what it is to desire something, we can see that the naturalist is committed to a very strong psychological thesis. That thesis is that what all human beings *really* seek—and the only thing we seek as an end in itself, though Rawls does not put the thesis that way—is a certain mental state: the satisfaction or relief of our own desires.

This psychological thesis is, as I implied earlier, a thesis about what we have in common as persons: the human personality. It is therefore a thesis about our nature. And it is because naturalism is committed to this thesis about our nature that it has an ethical component, or correlates to an ethical view, which Rawls calls "natural ethics" (2009, 114). For the naturalists with whom Rawls is concerned think that our good lies in getting what our nature inclines us to seek. It lies in an object that is truly capable of satisfying our desires.¹¹ Once we see what that object is, we

¹⁰ See also 123: "The egoist," by which Rawls means "the person moved only by desire in his sense," "treats other people as so many objects to be used as instruments for his own appetitional satisfaction."

¹¹ See 128, where Rawls says: "Any ethical theory which tries to find the 'good' in some objective value, i.e. in some object, is what we call a natural ethic."

learn something important about the kind of creatures we are. If that object is God, then it follows that we are creatures whose nature is to seek and enjoy God. Thus do our desires for satisfaction reveal our nature. These claims bring us at last to the version of Christianity that is the target of Rawls's thesis and, as we shall see, to interesting and telling continuities between the thesis and *TJ*.

3. Naturalist Christianity

Rawls thinks that Augustine, and Aquinas following him, Christianized the naturalism of Greek philosophy. Like Aristotle and Plato, Augustine and Aquinas thought that our good lies in what truly satisfies our desire. Rawls then notes that "God . . . is conceived by Augustine and Aquinas as the most desirable object" (2009, 115). They think, he says, that "God is . . . a bigger and better object for our enjoyment, an object which shall . . . satisfy our various appetitions" (2009, 162). It follows immediately that our good lies in God, just as Christians have traditionally thought.

This conclusion raises a number of questions, including what it means to say that our good lies *in* God and in what relationship we must stand to God to find God satisfying. It is on just these points that I believe Rawls misreads Augustine and Aquinas, implying that they think that relationship is to be a natural rather than a personal relation. But why does Rawls misread them? Why does he think that Augustine and Aquinas believed we seek an impersonal relationship with God?

I believe the answer is that he does so because he thinks they accept the core psychological thesis of naturalism and endorse Christianized versions of natural ethics on the basis of that thesis. In the previous section, I identified that core thesis as the claim that what we ultimately want is the satisfaction or the quelling of our desires. I believe Rawls reads this thesis into naturalist Christianity. Though he does not state the imputation clearly, there are a couple of passages which suggest it if pressed. Rawls speaks of a precursor of naturalist Christianity as holding that "if man cannot save himself, he must turn to some redeemer, some savior who has provided knowledge and secret pass-words for the return trip past the wicked demons and there be united once again to the heavenly realm *where man's immortal soul is to rest in peace*" (2009, 130, emphasis added). He notes that Augustine speaks of God as "our journey's end" (2009, 175). And after noting that "For [Augustine and Aquinas], God is . . . an object which shall so satisfy our various appetitions," Rawls continues immediately "that we shall cry 'Abba, Father' and *rest contentedly*" (2009, 162, emphasis added).

I do not think Rawls is right to read the core thesis of naturalism into Augustine and Aquinas, and I shall suggest below that he later came to

see the error of this reading. But I do not think that Rawls's reading is entirely without merit. While careful parsing of key passages—as about our restless desire for God at the beginning of Augustine's *Confessions* and about the nature of beatitude in Aquinas's *Summa Theologiae*—might allow them to slip out of Rawls's target area entirely, a twenty-one-year-old approaching the texts on his own could hardly be blamed for finding the thesis in these passages.¹² Moreover, if Rawls did read that thesis into Augustine and Aquinas, then we can see why he misunderstood the relationship in which they thought we must stand to God to be satisfied. For if they did accept the core thesis, that relationship would have to be one in which we regard God as the means to the ultimate end of rest or desire-satisfaction. As we have seen, Rawls reads naturalists as saying that even when the means is a person, that person is regarded or treated as an object rather than as a being with the powers of personality. That is why Rawls reads Augustine and Aquinas, incorrectly I think, as denying that we seek a personal relationship with God and as saying that we seek a natural one instead (2009, 182).

4. The Problem with Naturalist Christianity

What exactly did Rawls think was wrong with naturalist Christianity?

One possible line of criticism begins with the claim, which Rawls thinks he shares with the naturalist, that desire is egoistic and acquisitive. It might be inferred from this that all desires, because acquisitive, have an element of selfishness, graspingness, or rapacity that is ethically troubling. The problem with any kind of naturalism might then seem to be that naturalism locates our good in the satisfaction of desires which are troublingly immoderate if not vicious. But if this were the problem Rawls found with naturalism, then we would expect him to criticize Augustine and Aquinas for failing to notice the troubling elements of desire. We would also expect his criticism of naturalism to be thorough-going, for we would expect him to object to any natural relations whatever on the grounds that the "energy" which animates them is immoderate (2009, 118). Finally, we would expect Rawls to consider the possibility that naturalist Christianity can be salvaged by the availability of grace, which—it might be thought—can purify human desire by restraining its immoderation so that desire, thus purified, points to our good after all.

But Rawls never mentions this way of salvaging his target view, even to refute it. And he does not seem to think that there is anything ethically

¹² For a treatment that identifies the temptation to read Augustine in this way, and that offers a critique that is surprisingly similar to the one I shall impute to Rawls, see Grisez 2008, 34a. For what is, in effect, an attempt to anticipate and rebut the reading of Aquinas, see McCabe 2010, xiii.

troubling about desire or appetite as such. Rather, what Rawls objects to about naturalism is its exclusivity. For as we saw, "naturalism" names a family of views "in which *all* relations are conceived of in natural terms" (2009, 119, emphasis added). The naturalist's mistake, Rawls thinks, is to give her analysis of relations and desire, and the ethical conclusions she draws from that analysis, a cosmic reach. Thus Rawls says of naturalism that "the error lies, not in accepting nature, but in extending natural relations to include all of those in the cosmos" (2009, 121). He thinks the error of naturalist Christianity lies not in "accepting nature," but in concluding that our relation with other persons, including God, should be a natural relation.

This makes it tempting to suppose instead that Rawls thinks the error of naturalism is theological. That is, it is tempting to suppose that he thinks naturalist Christianity errs in thinking of God as the kind of being who can satisfy desire as Rawls and the naturalist understand it. There are a couple of passages which lend some credence to the supposition. Rawls says in one place that "We are mistaken if we think of God as another object of desire. We thereby make him part of nature" (2009, 121). Later, as if to raise the stakes, he says that "To speak of God as the most beautiful object, the most satisfying object, the most desired of all objects is to sin" (2009, 182). But while part of Rawls's objection to naturalist Christianity may be that naturalists think God is an object, I want to suggest that he thinks its fundamental mistake lies elsewhere.

Recall that Rawls defines naturalism as a way of *conceiving* of relations. The naturalist *conceives* all relations as relations between persons and objects, not because she thinks that persons can only enter into relations with things that really are objects, but because she thinks that when persons enter into relations, they thereby *treat them or regard them* as objects. And so the naturalist Christian is not committed to the view that, because we can only enter into natural relations with God, God must *be* an object. Rather, the naturalist Christian is committed only to the weaker view that, because we can only enter into natural relations with God, we must *treat God* as an object. The naturalist Christian thinks and speaks of God as an object, not because she denies that God is a person, but because she thinks that our good lies in God and she thinks our nature is such that God can be good for us but only if we objectify God. She therefore stresses the qualities of God in virtue of which God satisfies our desires, without denying that God has other qualities as well.

Thus on the reading I want to propose, Rawls thinks the fundamental mistake of naturalist Christianity is not theological but psychological. It lies in the supposition that we are moved only by desires as Rawls and the naturalist understand them—as acquisitive appetitions the ultimate object of which is a mental state of satisfaction, relief or rest. We have seen that Rawls's aims in his thesis are "to enter a strong protest" against

naturalism generally and against naturalist Christianity in particular, and to propose an alternative (2009, 107). That those aims have to be fulfilled by correcting the psychology rather than the theology of naturalism is suggested by the question which Rawls takes to be primary when he introduces the constructive part of the thesis. He begins, not with the theological question “What is God?” but with the question “First of all, what is man?” (2009, 121).

Rawls sees two related problems with naturalist psychology in its Christian and pre-Christian forms. Both of these problems stem, as I have suggested, from the cosmic reach or extension of naturalism. One problem is that the motive that the naturalist takes to be universal and exclusive is in fact quite unusual. Though the naturalist thinks human beings are moved only by desires, Rawls thinks that this is very rarely so. He says “Our natures are such that we seldom, if ever, are in a state of pure appetite” (2009, 185). And he has to go to considerable lengths to concoct an example of someone who is in such a state (2009, 184–85). The second and related problem Rawls sees with naturalism is that it ignores the mixture that he thinks is pervasively characteristic of human motivation. Indeed, the reason he has to go to such lengths to isolate the state of pure appetite is, he says, that “man’s creatureliness is always mixed up with and involved in his spirituality” (2009, 184). The mixture of our creatureliness and our spirituality—the latter of which Rawls equates with our personality—affects our motivation. He takes one of the strengths of his own analysis to be that it, unlike his target view, is true to what he calls the “dualism” of our nature and experience.¹³

In contrast to the clarity and economy with which Rawls characterizes appetite, his own view of what he thinks characteristically moves us is complex and would have to be teased out of his treatments of a number of examples. As in his treatment of naturalism, so in his own view, he thinks that the motive identified as characteristic is correlated with our participation in a distinctive kind of relation. In the case of Rawls’s own view, that relation is what he calls “personal.” In contrast to natural relations, into which one of the relata enters as a person who treats the other relatum as an object, personal relations are relations in which each of the relata is and treats the other as a being with the powers of personality. Treating one another properly as persons, Rawls thinks, establishes community between them. Because of this connection between personal relations and community, Rawls is emphatic in stressing that those who think all relations are natural miss what is essential to ethics.

¹³ Thus on 119, he says: “It is our hope that the dualism which we all feel in experience is explained by the two types of relations analyzed above. Personal relations open us to the realm of spirit; natural relations to the realm of nature; and because man participates in both realms, he is the peculiar creature that he is.”

For the naturalist misunderstands the ways in which persons can fail properly to relate to other persons—including God, and she misunderstands the ways in which persons should treat one another. That is why she misunderstands the nature of sin and faith. Rawls puts the failure of naturalism in strong terms. He says, for example, that “The result of a complete naturalization of the cosmos is to lose community, personality, and of course the real nature of God. . . . The approach of the natural cosmos is false; it misses the heart of the matter” (2009, 232).

Interestingly, Rawls never says that the naturalist philosophers and theologians offered faulty arguments for naturalism. Instead, he seems to think that they simply did not see the personal element to human motivation. Thus at the end of thesis, in an implied contrast with naturalism, he writes, “Therefore love, although it is giving, does not *overlook* the personality of the giver” (2009, 250, emphasis added). Elsewhere, he says that “[a]ll naturalistic thinkers have *completely missed* the spiritual and personal element which forms the deep inner core of the universe” (2009, 120, emphasis added). Rawls’s clear implication is that these thinkers never seriously entertained alternatives to naturalism because the possibility that naturalism might be wrong never occurred to them. They were—to take a phrase from Wittgenstein, whose work the young Rawls studied with Norman Malcolm—in the grip of a naturalist picture. In the next section, we shall see why Rawls came to think philosophers have found that picture so gripping.

5. Naturalism and Hedonism

By the time he wrote *TJ*, Rawls had ceased to have an academic interest in naturalist Christianity, but he continued to target views which were committed to the core thesis of naturalism that he identified as an undergraduate. This is evident in *TJ*’s diagnosis of teleological theory, found near the end of the book. There Rawls says that while teleological views are not necessarily committed to any form of hedonism, hedonism is their “symptomatic drift” (Rawls 1999, 490). Hedonistic utilitarianism, like the version of Christianity that Rawls targeted in the thesis, says that we ultimately aim at a certain kind of mental state. As Rawls puts it, hedonistic utilitarians, like the Augustine and Aquinas of the thesis, “have supposed that characteristic experiences exist and guide our mental lives” (Rawls 1999, 490). The naturalist of the thesis thinks we ultimately desire something called “satisfaction” while the hedonistic utilitarian of *TJ* thinks we aim at pleasure. Though this may seem to be a significant difference, the Rawls of *TJ* clearly thinks that teleological doctrine which drifts toward hedonism goes wrong for the same reason that naturalist Christianity does. It, like naturalist Christianity, supposes that there is some one kind of experience that we ultimately desire. And it, like

naturalist Christianity, supposes that the end at which that desire aims reveals the kind of beings we are (Rawls 1999, 491).

One of the great unanswered questions of *Brief Inquiry* is why so many powerful thinkers accepted naturalism's core thesis and believed that we ultimately aim at mental states. Clearly Rawls thinks that naturalist Christianity was misled by naturalist philosophy. But why did Plato, Aristotle, Augustine, and Aquinas find naturalist philosophy so bewitching that they simply overlooked claims about human nature and its expression that Rawls found so obvious?

Though the undergraduate Rawls seems not to have been interested enough in this question to ask it, the mature Rawls found the question intensely interesting. I think he came to see that those whom he had described as naturalists in *Brief Inquiry* did not endorse the core thesis of naturalism. But he also came to see that the philosophers who *did* endorse that thesis did so because it promised a solution to certain important philosophical problems.

Rawls argues in *TJ* that teleological theorists tend toward the view that human life has a single dominant end for the sake of which all else is to be sought. For without a dominant end, they think, there would be no standard for choosing among and ordering other ends into a conception of our good. Our good would then be arrived at by what Rawls calls "purely preferential choice." This rational indeterminacy in the individual's good would itself be troubling, since it would leave open the possibility that individuals are enjoined to maximize an indeterminate quantity. An even more serious difficulty, Rawls says, is that "[i]n teleological theory, any vagueness or ambiguity in the conception of the good is transferred to that of the right" (1999, 490). For if individuals just decide what is good for them, and if—as teleological theory says—the right is what maximizes the good, then the contents of our obligations to one another will ultimately depend upon our purely preferential choices. This implication, Rawls says, is counterintuitive. Rejecting it, teleological theorists reject rational indeterminacy and gravitate toward "dominant end theory," an ethical theory according to which we seek a single end to which all else is subordinate. Pleasure seems the only plausible candidate for the single end of all that each of us does. Moreover, views which identify the right with the maximization of the good need some way of commensurating and combining individuals' goods so as to arrive at what is to be maximized. If desire-satisfaction is always accompanied by pleasure, then pleasure can serve as the requisite "common denominator" of individual goods, and hence as the *maximandum*.¹⁴

At the end of the previous section, I noted that naturalism as Rawls describes it in the thesis might seem to be importantly different than

¹⁴ This paragraph summarizes the argument of *TJ* (1999, 490).

hedonist utilitarianism as Rawls describes it in *TJ*. Though both hold that we ultimately aim at a mental state which serves as a dominant end, the states at which we are said to aim are different. The naturalist thinks we aim at the quieting of desire while the hedonist thinks we aim at pleasure. But pleasure is notoriously difficult to define, either verbally or ostensibly, since pleasures seem to be so various. If it is understood broadly as the satisfaction of desire, then differences between the naturalism of *Brief Inquiry* and the hedonism of *TJ* are narrowed considerably, and the claim that the thesis and the relevant sections of *TJ* have a common target looks even more compelling.

If Rawls thinks that the fundamental mistakes of teleological doctrines and naturalist Christianity are similar, so, too, are the ways that he attacks them.

In Part I of *TJ*, Rawls argues that utilitarianism clashes with our considered judgments of justice and he offers lengthy arguments that neither average nor classical utilitarianism would be chosen in the original position. But the force of the latter arguments, at least, depends upon the willingness to address questions of justice using Rawls's theoretical machinery. It is only in the late sections of *TJ* on hedonism, dominant ends and the unity of the self that Rawls finally unearths and engages the deep motivations of teleological theories. The diagnosis of teleological theory offered there, and the arguments offered against it, are quite developed. But despite the development of Rawls's argument, at a crucial moment he pulls back from saying that he has *refuted* teleological doctrine. He implies that teleological doctrine—like naturalist Christianity, though he does not say that—supposes that it is “our aims that primarily reveal our nature” (Rawls 1999, 491). He then says that the difficulties that result from identifying that aim with pleasure “*suggest*” rather than imply “that the structure of teleological doctrines is radically misconceived” (Rawls 1999, 491). What is needed is not a different solution to the philosophical problems that pushed teleological doctrine toward hedonism but a new way of looking at those problems. What saves us from the radical misconception of teleological theory is “the complete reversal of perspective” we experience when that theory is contrasted with a developed version of the contract doctrine (Rawls 1999, 493).

The Rawls of the undergraduate thesis seems to want to win over his readership by effecting a similar reversal of perspective. As his title suggests, he thinks naturalist Christianity is unable to give adequate accounts of sin and faith. While he offers arguments to this effect (2009, especially at 189–92), the arguments depend upon Rawls's own definitions of “sin” and “faith,” definitions for which he does not offer any extensive defense. With one exception, he does not allege that naturalist Christianity clashes with our intuitions, or what he would later call our “considered judgments,” about sin and faith (2009, 182). Nor does he identify tensions

or incoherencies internal to his target view. Rather, the thesis conveys the impression that Rawls thinks readers will just see that his view of sin is right, and that naturalism “is radically misconceived,” once we are exposed to his view. At the beginning of the thesis, Rawls says that he is “proposing more or less a ‘revolution’ by repudiating” naturalist Christianity (2009, 107). In the thesis, the revolutionary’s chief weapon seems to be the telling juxtaposition, the contrast that revolutionizes by reversing our perspective on sin and faith.

6. The Aristotelian Principle

What are we to see when our perspective is reversed?

An important but under-appreciated contention of *TJ*, Part III is that members of a just society would want to express their nature as free and equal rational beings (Rawls 1999, 503). Indeed, according to *TJ*, members of the well-ordered society regard the expression of their nature as a very great good. Obviously the desire to express one’s nature requires considerable explanation. Let me make just three points about it. First, the desire to express one’s nature is a desire that is satisfied in action. Second, that action is not one-off. The desire to express our nature is not, for example, satisfied by action performed at life-defining moments at which someone shows what kind of person she is or where her deepest commitments lie. Rather, members of the well-ordered society express their nature when they conduct themselves as free and equal rational beings over the course of life, framing and executing their plans of life in a certain way. Thus we might say that the object of the desire is *adverbial*. Third, the lives in which members of the well-ordered society express their nature are, of course, social lives. They are lives lived among others who are “self-originating sources of valid claims” (Rawls 1980, 543) and who reveal their aims in social and political life by pressing those claims.

How do we express our nature as free and equal rational beings living among other such beings? In the section of *TJ* entitled “The Kantian Interpretation of Justice as Fairness,” Rawls says “to express one’s nature as a being of a particular kind is to act on the principles that would be chosen if this nature were the decisive determining element” (Rawls 1999, 222). The original position is structured so that our nature as free, equal and rational is the decisive determining element of the choice there. So the Rawls of *TJ* thinks members of the well-ordered society express their nature when they regulate their lives by principles that would be chosen in the original position.

Thus according to the Rawls of *TJ*, when our perspective is reversed and we see the error of teleological theories, we see that what reveals our nature is not our aims. It is “the principles that we would acknowledge to govern the background conditions under which these aims are to be

formed and the manner in which they are to be pursued" (Rawls 1999, 491). Once we see *that*, we shall see that it is the sense of justice that "reveals what the person is" and that we *express* our nature in action when we live a certain kind of social life (Rawls 1999, 503). In *TJ*, that life is the ongoing life of a good citizen in a well-ordered society. According to the Rawls of *TJ*, the nature we express is our nature as free and equal rational beings. The reason we express it by acting from principles of justice is because the life of a well-ordered society is a life lived among persons who are, as noted earlier, "self-originating sources of valid claims" on the rights, liberties, income, wealth, and opportunity that result from social cooperation.

The Rawls of the undergraduate thesis also proposes an alternative view according to which our nature is not fulfilled by the satisfaction of desire, but is expressed by living a certain kind of social life. In the thesis, that life is a life of personal rather than natural relations. Properly conducting a life of personal relations, like conducting a life of good citizenship in the well-ordered society of *TJ*, poses challenges that grow out of the nature of persons. In personal relations, Rawls says, we are always held accountable by those with whom we enter into such relations (2009, 116). Moreover, the persons with whom we enter into personal relations are not automatically and completely self-disclosing, as objects are said to be (2009, 115–16). Rather, Rawls says suggestively that "personal relations proceed on the basis of mutual revelation" among individuals each of whom is unique (2009, 124; see also 117–18, 124, 153, and 224). The notion of mutual revelation is unfortunately underdeveloped in the thesis. But I think Rawls means to imply that the dynamics of mutual revelation and accountability are what make for the distinctive character of personal relations, in which neither of the relata is passive, in which both of the relata are unique and in which both engage in what he calls the "activity" of fellowship (2009, 180).

Rawls says in *TJ* that the goodness of expressing our nature "follows from the Aristotelian Principle" (Rawls 1999, 390). The Aristotelian Principle is a psychological generalization according to which human beings prefer more complex and challenging activities rather than less. It explains, for example, why people with knowledge of both prefer chess to checkers, and why we complicate activities to alleviate boredom. In an important passage, Rawls says that "conveys" two points: "enjoyment and pleasure are not always . . . the result of returning to a healthy or a normal state, or of making up for deficiencies; rather many kinds of pleasure and enjoyment arise when we exercise our faculties," and "the exercise of our natural powers is a leading human good" (Rawls 1999, 374n20).

From these remarks, I take it that the views of our nature and its expression that Rawls opposes in *TJ* are those according to which the

human good is realized by our return to “a healthy or a normal” condition, by our satisfying serious deficiencies so as to attain that condition, or by stasis in a healthy, normal or completed condition once that condition is attained. The quoted remarks also suggest that Rawls will propose an alternative view of our good. That view is one which takes account of the three observations I made about the desire to express our nature. For in *TJ*, the good of expressing our nature is realized in the ongoing activity of living as just members of a well-ordered political society the citizens of which are equally possessed of moral personality.

I believe that one of the young Rawls’s objections to naturalist Christianity was, in effect, that it overlooks the implications that the Rawls of *TJ* would draw out of the Aristotelian Principle. The points expressed by the principle imply his opposition to precisely the mistake about our good that he thinks the target view of Christianity makes. For Rawls thinks that according to that view of Christianity, our good is found in restoration to a state in which we are completed or our deficiencies overcome. That is the state naturalist Christianity says we reach at the end of “return” trip to a realm where we will find complete satisfaction in the presence of God and where we shall enjoy “rest” and “contentment.” Rawls thinks the target view makes that mistake because it takes our good to lie in the satisfaction of our desires and because it accepts naturalism’s core thesis about what we really desire. It makes that mistake because, like hedonistic utilitarianism, it takes us to be creatures whose nature is revealed by our common aim or ultimate object of desire.

Thus once the reversal of perspective is brought about, we see that in the thesis as in *TJ*, our nature is realized in our on-going conduct toward others who have the powers of personality. In *Brief Inquiry*, that life is not a life among others who are thought of as self-originating sources of valid claims. Nor is it a life among those who make demands on a collectively produced social surplus. And so the thesis never says that we express our nature by acting from principles of political justice. But it *is* a life lived among those to whom we have obligations. For Rawls says that “[s]ince each man bears the image of God, he is a responsible being, and a personality with obligations” (2009, 121–22). The image of God, in virtue of which we have obligations to others, “provides the basis of man’s moral being,” and since we are moral beings, we are communal (2009, 122). Thus I think it is fair to say that in the thesis as in *TJ*, we express our moral nature by living on-going social lives in which we honor the demands of right.

7. Rawls on Moral Motivation: Some Objections Considered

The Rawls of *TJ* is often criticized for offering an inadequate account of human motivation, specifically, of the motivation to do justice. The sec-

tions of *TJ* to which *Brief Inquiry* draws attention provide Rawls the resources to answer these objections. To see that they do, we need to see how the arguments for the objections go.

Those arguments, when fully spelled out, would vary in their specifics. But the arguments for them begin with roughly the same steps, which I take to be the following. The objectors assume that being just is or is part of leading a fully human life; it is or is part of fully realizing our human nature. Since what moves us to be just must be desire of some kind, that motive must be a desire for what fully realizes our nature. What fully realizes our nature is identified with a common human end, and the desire to be just is then identified with a desire or love for that end. With these background assumptions in place, the objections to Rawls's account of moral motivation follow straightforwardly. Rawls tries to arrive at principles of justice without reference to a common human end, and he denies the presence or the relevance of a desire for it. He must therefore identify moral motivation with some other human desire. It is sometimes objected that Rawls identifies the motive to be just with self-interest. More often, it is objected that the desire with which Rawls identifies the motive to be just cannot really be morally motivational. And, it will be said, his misidentification will show itself in the politics of a well-ordered society—a society in which, Rawls stipulates, everyone would act from that motive. To confirm that criticisms of Rawls's account of moral motivation do indeed follow this pattern of argument, let us look at objections lodged by some prominent critics.

An article of Timothy Jackson's seems to follow the pattern closely. Jackson implies that if we fail to act from the desires he himself identifies as morally motivational, we are "less than fully human" (Jackson 1997, 60). I take it that in saying this, Jackson means to take what I said is the first step in the common line of argument I sketched, for I take it he thinks that in acting from moral motives, we *are* "fully human" and fully realize our human nature. I believe he takes the second step elsewhere by identifying an end in which he thinks our nature is fully realized, the end of "participat[ing] in the life of God" (Jackson 2007, 14). He takes the third by identifying moral motivation with a desire for that end, for the desires he identifies as morally motivational include love for God and neighbor (Jackson 1997, 50).

Rawls arrives at principles of justice without appealing to the claim that our good consists in participation in the divine life, and so he identifies the desire to act justly with a different desire than Jackson does. The question is what desire Jackson thinks he identifies it with. Rawls's principles of justice are, of course, chosen in the original position. On Jackson's reading—and here he takes the penultimate step—Rawls is forced to the conclusion that our interest in acting from principles chosen in the original position is "rational self-interest" —as opposed,

presumably, to a desire to live as fully human by acting from charity (Jackson 1997, 59). Thus Jackson writes that

those who enter the original position take the veil of ignorance and agree to Rawls's two principles of justice, not because they empathize with those potentially less well off than themselves, but because they fear that they themselves will be among the unfortunate or victimized. This is a fundamental avoidance, not to say denial, of charity as other-regard; it must deeply trouble an orthodox Christian who would see the social order ideally delivered from anxiety and self-seeking. (Jackson 1997, 59)

As we would expect, Jackson concludes that Rawls's misidentification of the desire for justice has unwelcome political implications, for he says that if citizens were generally to act on the desire Rawls identifies with as morally motivational, "justice itself [would] atroph[y]" (Jackson 1997, 59).

L. Gregory Jones's highly compressed critique of Rawls on moral motivation, like Jackson's more expansive one, follows the line of argument I sketched. Jones suggests that he himself endorses an Aristotelian view, according to which being just is part of "attaining *eudaimonia*." He also suggests that he thinks moral motivation is properly understood as, or as part of, "the person's particular commitment to her flourishing" (Jones 1988, 258). But, Jones objects, Rawls cannot endorse the right account of moral motivation because of his "commitment to the original position." Here I take Jones to be objecting to Rawls on just the ground that Jackson did: Rawls cannot identify our interest in principles chosen in the original position with an interest in attaining our end—in Jones's case, *eudaimonia*—because the principles are chosen behind the veil of ignorance and without reference to an account of human flourishing. Jones does not infer that Rawls thinks our interest in justice must ultimately be self-interest. But he, like Jackson, thinks that that interest must be "weak."¹⁵

The criticisms of Rawls in Eric Gregory's recent book draw on elements of the pattern of argument I have traced in Jackson and Jones. Gregory thinks that a fully human life is one in which we love God and love others "in God" (Gregory 2008, 338), and that treating others justly is part of living such a life (Gregory 2008, 110, 157). He says that he "sees moral motivation in terms of . . . non-possessive *eros*" (Gregory, 240), which—as I read his discussion of non-possessive *eros*—implies that the motivation to treat others justly is love for them "in God" (see Gregory 2008, 46–47). Of course this is not the desire Rawls identifies with the desire to be just. While Gregory takes issue with Jackson's claim that Rawls equates the latter desire with a kind of self-interest (Gregory 2008, 109), he does not say explicitly what he thinks Rawls does take that desire to be. But he

¹⁵ This paragraph summarizes the argument at Jones 1988, 258.

seems to suggest that Rawls identifies it with a "formal subscription to liberal principles and procedures" (Gregory 2008, 318). Like Jackson and Jones, he insists that moral motivation as he thinks Rawls understands it is too weak to sustain "a healthy liberal society" (Gregory 2008, 318).

The young Rawls might have dismissed all three authors' criticisms as premised on the naturalist Christianity that he read into Augustine and Aquinas. The mature Rawls—who I believe realized that he had misread Augustine and Aquinas in *Brief Inquiry*—would have recognized that doing so would be unfair, since it is unlikely that Jackson, Jones, or Gregory subscribes to the core thesis of naturalism.

Instead of dismissing their criticisms, he would have begun by reiterating the critique of religious ethics found in one of the sections to which *Brief Inquiry* directs our attention, *TJ* §83. There he noted that the end of loving and serving God "is left unspecified to the extent that the divine intentions are not clear from revelation, or evident from natural reason" (Rawls 1999, 485). The demands of charity must therefore be made clear by appeal to scripture, tradition or ecclesiastical authority. Since specifications so arrived at are not publicly justifiable, a conception of the person as oriented toward the end of loving and serving God is not a useful one for arriving at a public conception of justice.¹⁶ To construct such a conception of justice, it is necessary to begin with a very different conception of the person than those that are implicit in Jackson's, Jones's, and Gregory's criticisms. With this alternative ethical conception of the person in hand, Rawls would say, it is possible to reply to their objections.

Consider first Jackson's claim that according to Rawls, the interest we take in acting from the demands of justice is self-interest rather than love for God and neighbor. The fact that Rawls does not identify the desire to

¹⁶ Some of Rawls's remarks in *TJ* §83 have led readers to think Rawls regards Aquinas and Ignatius as fanatical (Hittinger 1994, 595n23; Jackson 2007, 14n35). *TJ* §83 is a discussion of what Rawls calls "dominant ends" and it is largely devoted to hedonism. The Rawls of *TJ* also considers the possibility that Aquinas was a dominant-end theorist, but one who took the dominant end of human life to be, now not pleasure or desire-satisfaction, but the vision or the service of God.

Though Rawls does not put it this way, I believe he thinks that views which purport to be dominant end views face a dilemma. If the ends they claim are dominant are specific enough to solve the problems of indeterminacy, then they are procrustean, "irrational, or more likely mad" (Rawls 1999, 486). If those ends are not specific enough, as Rawls seems to think *divine service dominant ends* like Aquinas's are not, then the charges of irrationality and madness are not intended to apply to them. Instead, they suffer from the problems that Rawls says plague intuitionism (Rawls 1999, 30–36). They are ethical views which allow that a range of considerations bear on our choices because the verdict of the divine will in any given situation depends upon the relative weights of those considerations. But since they fail to provide any publicly defensible way of balancing those considerations even when fundamental questions of right are at issue, they leave our conceptions of the good and the content of our obligations rationally indeterminate.

be just with concern for “those potentially less well-off than [ourselves]” does not imply, he would say, that he identifies it with self-interest. To assume that it does is to assume that our desire to treat others justly must have as its object either their good or our own. It is to assume, crudely put, that we must be moved to treat others justly by a love of *someone’s* good and that the only interesting motivational question is whose good we love. But to make this assumption is to ignore the fact that there are many kinds of desires. More particularly, it is to ignore the kind of desire that the Rawls of *TJ* takes to be central to our sense of justice: what he would later call “principle-dependent desires” (Rawls 1993, 82–83). These are not desires to promote anyone’s good. They are desires to act from principles for their own sake. Rawls argues at some length that citizens of a well-ordered society would have a settled disposition to be just because they would develop such desires, focused on the principles of justice (Rawls 1999, 405–19). Thus the first thing to note about Jackson’s conclusion is that it suggests a false dichotomy, one which excludes precisely the kind of desire that Rawls takes the desire to act justly to be.

Both Jackson and Jones maintain that the inadequacy of Rawls’s account of moral motivation stems from his use of the original position. We saw that when Jackson argues that Rawls thinks our interest in justice is self-interest, he does so by making a point about the interest we have in “tak[ing] the veil” and acting from principles adopted behind it.

Rawls has a good deal to say about why we should take an interest in the principles. One of his arguments speaks to an assumption that seems to be shared by Jackson, Jones, and Gregory—the assumption that using the original position to arrive at principles of justice is an alternative to basing them on claims about our common human nature. The Rawls of *TJ* thinks that we are free and equal moral persons, that the original position represents us as such persons (Rawls 1999, 221) and, as we have seen, that we express our nature as free and equal by acting from principles chosen there. Thus Rawls says that the desires to express our nature as such persons and to act justly “are both dispositions to act from precisely the same principles: namely, those that would be chosen in the original position” (Rawls 1999, 501). It is because we express our nature by acting from principles chosen in the original position that, as I noted earlier, Rawls says that a sense of justice “reveals what the person is” (Rawls 1999, 503).

These remarks show that Rawls does not appeal to the original position *rather than* to claims about our nature. In the section of *TJ* on the Kantian interpretation, on which Rawls builds in §86, Rawls says that original position *represents* our common nature and brings it to bear on the choice of the principles. “Since all are similarly free and rational,” Rawls says there, “each must have an equal say in adopting the public principles of the ethical commonwealth” (Rawls 1999, 226). And so among

the reasons he thinks we should “take the veil” is that by doing so, we can express our nature as beings who live social lives among others who are our equals.¹⁷

How do we act on principles chosen in the original position? We saw at the beginning of section six that the desire to express our nature is not satisfied by one-off or episodic action and, as we would expect, the same is true of the desire to act on the principles. Rawls says that the just person treats her principle-dependent desire to be just as “finally regulative” (Rawls 1999, 503). What he means is that in drawing up her plans, in deliberating about what to do, and in pursuing all her ends, she treats her desire to be just as higher-order and as dispositive. It is in planning, deliberating, and acting this way that she expresses her nature. And so Rawls would maintain, against Jones, that realizing our nature is not a matter of “attaining” anything. Rather, we express our nature in an ongoing life of practical deliberation and activity, governing ourselves by the demands of right. This view is suggested in Rawls’s senior thesis. We have seen how the Rawls of *TJ* appeals to the Aristotelian Principle to give it a clearer expression and a firmer foundation.

I said earlier that Gregory seems to think Rawls identifies the desire to be just with “formal subscription to liberal principles and procedures.” Since the Rawls of *TJ* thinks that the desire to be just is principle-dependent, the reading I attributed to Gregory is right. But Gregory’s use of the word “formal” betrays a critical edge, and I noted that he—like Jackson and Jones—argues that the motive to be just as he thinks Rawls describes it is “weak.” Like Jackson, he seems to think that citizens of a well-ordered society must act from mutual love if they are to sustain a “healthy liberal democracy.”

Why does Rawls identify the desire to be just with a principle-dependent desire rather than with some other desire that has a less abstract or formal object? Sustaining a just society under modern conditions requires us to support just institutions for their own sake, to comply with legitimate laws and to act justly toward our fellow citizens in a large, pluralistic and anonymous society. Rawls would argue that while a desire to act from mutually justifiable principles can move us to act in these ways, desires with less abstract objects—such as love for particular persons, associations or communities—cannot. Recall Gregory’s claim that moral motivation is non-possessive *eros* (Gregory 2008, 240). If this is right, then non-possessive *eros* must be capable of moving us to comply with the demands of just institutions, and of maintaining just relations toward those we do not know, including other participants—whom we will

¹⁷ Thus Rawls says “when we knowingly act on the principles of justice in the ordinary course of events, we deliberately assume the limitations of the original position. One reason for doing this . . . is to give expression to one’s nature” (Rawls 1999, 222).

never meet—in large schemes of social cooperation. In that case, we must be capable of developing rightly ordered erotic attachments to all of our fellow citizens, even those we do not know and will never meet. It is not at all obvious what it would mean to say that we do without stretching the meanings of “erotic” and “attachment” quite far.

It is possible that Gregory thinks *eros* is dispositional: we would treat them rightly if we did meet them. He says that “Augustine suggests that a good lover will not love *every* neighbor as much as she will love *any* neighbor who happens across her way” (Gregory 2008, 295). This remark seems to assert that rightly ordered love is strong enough to motivate, not just in the usual range of special relationships, but also in the special relationship created by chance proximity. But it does not say how that love can motivate us to act justly toward those with whom we do not stand in even that relationship, such as most of our fellow citizens. Gregory’s account of motivation, though intended to be stronger than Rawls’s, leaves open the question to which Rawls’s account is an answer.¹⁸

I said at the outset that Rawls’s undergraduate thesis is important in part because it draws our attention to Rawls’s claims about our nature and its expression and to some of the sections of *TJ* in which the mature Rawls states those claims most forcefully. I have located those claims in the sections of *TJ* devoted to hedonism, dominant ends, the unity of the self, and the good of justice. Those sections, though neglected in the voluminous secondary literature on Rawls, are very important for understanding *TJ*. I have tried to demonstrate their importance by drawing on them to show how the Rawls of *TJ* would respond to critics of his account of the desire to be just.

The sections of *TJ* to which *Brief Inquiry* draws our attention are also important for understanding Rawls’s transition to political liberalism. For the Rawls of *TJ* was concerned, not only to identify the principles of justice, but also to show that they would be stable because—roughly—those who live under them would judge that it is good to be just. In *TJ*, the good of being just is connected with the goods of expressing our nature and unifying our practical reason. Rawls’s treatment of those goods depends upon an ethical conception of the person. That conception is highly Kantian. In this it is reminiscent of *Brief Inquiry* which, as we saw at the end of section six, stresses the good of a community of persons who are bound by ties of obligation and who express their nature by observing them. Indeed, though the young Rawls does not seem to have known Kant’s works at all well, the moral and religious sensibility manifested in the thesis matured naturally into the Kantianism of Rawls’s later work.¹⁹

¹⁸ This paragraph draws on Weithman 2011.

¹⁹ See the remarks of Cohen and Nagel at Rawls 2009, 11.

In the years after he published *TJ*, Rawls came to realize that the Kantian conception of the person might not be universally shared in a pluralistic society. His dissatisfaction with *TJ*'s treatment of the good of justice led him to recast justice as fairness so that it depended upon a political rather than an ethical conception. And he came to think that in acting from the principles of justice we express our nature, not as moral persons, but as free and equal citizens. Unfortunately I cannot take up Rawls's political turn here (see Weithman 2010). If what I have said about it is right, then the importance of the sections of *TJ* on which I have drawn will have to be much better recognized if we are to track the development of Rawls's thought beyond *TJ*. I hope that Rawls's undergraduate thesis will convince readers of their importance by showing that they respond to concerns, and betray an outlook, that moved Rawls from the very beginning of his working life.

REFERENCES

- Berkowitz, Peter
2009 "God and John Rawls." *Policy Review* 155 (May 29). <http://www.hoover.org/publications/policy-review/article/5573> (accessed July 11, 2011).
- Bok, Hilary
2009 "When Rawls Met Jesus: How the Philosopher's Early Religious Beliefs Guided His Secular Thinking." *Democracy: A Journal of Ideas* 14 (Fall). <http://www.democracyjournal.org/article.php?ID=6717> (accessed July 23, 2010).
- Cohen, Joshua, and Thomas Nagel
2009 "Introduction." See Rawls 2009, 1–23.
- Fitzgerald, Allan D.
1999 *Augustine Through the Ages: An Encyclopedia*. Grand Rapids, Mich.: Eerdmann's.
- Galston, William
2009 "Driven Up the Rawls." *The New Republic*. <http://www.tnr.com/blog/william-galston/rawls> (accessed July 23, 2011).
- Gregory, Eric
2007 "Before the Original Position: The Neo-Orthodox Theology of the Young Rawls." *Journal of Religious Ethics* 35.2 (June): 179–206.
2008 *Politics & the Order of Love: An Augustinian Ethic of Democratic Citizenship*. Chicago: University of Chicago Press.
- Grisez, Germain
2008 *The Way of the Lord Jesus, Volume 1: Christian Moral Principles*. Staten Island, N.Y.: St. Paul's/Alba House.
- Habermas, Jürgen
2010 "The 'Good Life'—A 'Detestable Phrase': The Significance of the Young Rawls's Religious Ethics for His Political Theory." *European Journal of Philosophy* 18.4: 443–54.

- Harmon, Jonathan
 2011 "Review of John Rawls, *A Brief Inquiry into the Meaning of Sin and Faith*." *Philosophy and Social Criticism* 37.6: 735–39.
- Hittinger, Russell
 1994 "John Rawls, *Political Liberalism*." *Review of Metaphysics* 47.3: 585–602.
- Jackson, Timothy
 1997 "*Prima Caritas, Inde Jus*: Why Christians Shouldn't Baptize John Rawls." *The Journal for Peace and Justice Studies* 8.2: 49–62.
 2007 *The Priority of Love*. Princeton, N.J.: Princeton University Press.
- Jones, L. Gregory
 1988 "Should Christians Affirm Rawls's Justice as Fairness? A Response to Professor Beckley." *Journal of Religious Ethics* 16.2: 251–71.
- McCabe, O.P., Herbert
 2010 *The New Creation*. London: Continuum.
- Meilaender, Gilbert
 2009 "We Were Believers Once, and Young." *First Things* 196 (October): 49–53.
- Nagel, Thomas
 2006 "Progressive but Not Liberal," *The New York Review of Books*. 53.9 (May 25): 45–49.
- Rawls, John
 1980 "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77.9: 515–72.
 1993 *Political Liberalism*. New York: Columbia University Press.
 1999 *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
 2009 *A Brief Inquiry into the Meaning of Sin & Faith*. Edited by Thomas Nagel. Cambridge, Mass.: Harvard University Press.
- Sandel, Michael
 1982 *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Schaefer, David Lewis
 2010 "Review of John Rawls, *A Brief Inquiry into the Meaning of Sin and Faith*." *Society* 47.3: 276–78.
- Weithman, Paul
 2010 *Why Political Liberalism? On John Rawls's Political Turn*. New York: Oxford University Press.
 2011 "Review of Eric Gregory, *Politics & The Order of Love*." *The Philosophical Review* 120.1: 130–34.
- Wolterstorff, Nicholas
 2010 "Review of John Rawls's *A Brief Inquiry into the Meaning of Sin and Faith*." *Hedgehog Review*, 12.1: 84–85.