

## 1 On John Rawls's *A Brief Inquiry into the Meaning of Sin and Faith*

---

It is by now commonly accepted that John Rawls's undergraduate thesis *A Brief Inquiry into the Meaning of Sin and Faith*<sup>1</sup> shows him to have been much better informed about, and much more sympathetic to, religion than was previously thought.<sup>2</sup> It is also widely recognized that *A Brief Inquiry* anticipates some of the ideas found in *A Theory of Justice* and beyond.<sup>3</sup> There is, however, a considerable divergence of opinion about whether publication of the thesis advances our understanding of Rawls's mature work.

Some readers have claimed that the Rawls of *A Theory of Justice* 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.<sup>4</sup> Jürgen Habermas has recently asserted an important connection, not between the undergraduate thesis and *A Theory of Justice*, but between the thesis and Rawls's eventual presentation of justice as fairness as a political liberalism.<sup>5</sup> Many readers, however, have denied that there is any scholarly significance to the thesis at all.<sup>6</sup> It is this last view which I wish to challenge.

Originally appeared in *Journal of Religious Ethics* 40 (2012): 557–82 and is reprinted by permission of John Wiley & Sons Publications.

<sup>1</sup> This chapter was drafted for a symposium on Rawls's undergraduate thesis 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 Jerry McKenny, Jean Porter and Maura Ryan – for helpful comments on an earlier draft, and to two anonymous referees for the *Journal of Religious Ethics* for their comments on a later one.

<sup>2</sup> For especially clear statements, see Gregory, "Before the Original Position," p. 183, and Wolterstorff, "Review of John Rawls's *A Brief Inquiry*," p. 84.

<sup>3</sup> As noted by Nagel and Cohen in their "Introduction" to *Brief Inquiry* as well as by Bok, "When Rawls Met Jesus."

<sup>4</sup> See Berkowitz, "God and John Rawls," and Galston, "Driven Up the Rawls."

<sup>5</sup> Habermas, "The 'Good Life'."

<sup>6</sup> For example, in a generally sympathetic review of *Brief Inquiry*, Jonathan Harmon writes that "I will not dwell too long on the arguments of the thesis, as I believe the benefits to the Rawls

4 The undergraduate thesis

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 claims about human nature, and the expression of human nature in human life, which it shares with a philosophical view targeted by Rawls in *A Theory of Justice*. The sections of *A Theory of Justice* 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 much of Part III.<sup>7</sup> 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 from the early 1940s, when he wrote the thesis, until the 1970s, when he published *A Theory of Justice*.

*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 *A Theory of Justice*, §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 *A Theory of Justice* 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 *A Theory of Justice* 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

scholar of reading it are mostly indirect: not necessarily in what Rawls says, but in what it says about the young Rawls." Harmon, "Review of John Rawls, *A Brief Inquiry*," p. 735. Gilbert Meilander 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." Meilander, "We Were Believers Once, and Young," p. 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, "Review of John Rawls, *A Brief Inquiry*," p. 278.

<sup>7</sup> To give only one example of many, see the glancing references to §§83–85 at Sandel, *Liberalism and the Limits of Justice*, pp. 151, 156, 158 and 162. Thomas Nagel vigorously criticizes Sandel for his misinterpretation of a crucial passage in *TJ* §85 in Nagel, "Progressive but Not Liberal," p. 47, note 3, and the associated text.

assumptions that underlie §86 is important for understanding the development of Rawls's thought after *A Theory of Justice*. 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 *A Theory of Justice*, 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.

### §1.1 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" (*BI*, p. 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" (*BI*, p. 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 (*BI*, p. 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,

6 The undergraduate thesis

since he says that objects include processes (*BI*, p. 180) and probably states of affairs (cf. *BI*, p. 150). Later we will need to look closely into what objects are. For now, it is important 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” (*BI*, pp. 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.” 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 as 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.

## §1.2 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 *BI*, pp. 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”<sup>8</sup> – “is reduced to the level of desire and appetite” (*BI*, 107).<sup>9</sup>

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 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” (*BI*, p. 180, emphasis original). Since Rawls thinks we can have a desire or an appetite for persons (*BI*, pp. 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

<sup>8</sup> See *BI*, p. 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[.]”

<sup>9</sup> Later, he says that “appetitional desires are the energies of all natural relations” (*BI*, p. 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 *BI*, pp. 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.

8 The undergraduate thesis

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” (*BI*, p. 250). This echoes a conclusion Rawls reached earlier, in which he said that for the naturalist “all love is acquisitive” (*BI*, p. 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*” (*BI*, p. 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” (*BI*, p. 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]” (*BI*, p. 180) such as exercise and sleep (*BI*, pp. 180, 184). But it is not at all clear how can I 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?<sup>10</sup>

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

<sup>10</sup> Indeed, Augustine is often said to claim that all human love is acquisitive; see Fitzgerald, *Augustine Through the Ages*, p. 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.”

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*" (*BI*, p. 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 that "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*" (*BI*, p. 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 means to *my* satisfaction or relief. It is important to note 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<sup>11</sup> (*BI*, pp. 187–88). So the state of satisfaction that results from the acquisition of an

<sup>11</sup> See *BI*, p. 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[.]"

10 The undergraduate thesis

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 that it comes with a correlative ethical view, which Rawls calls “natural ethics” (*BI*, p. 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.<sup>12</sup> Once we see what that object is, we 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 *A Theory of Justice*.

### §1.3 Naturalist Christianity

Rawls thinks that Augustine, and Aquinas following him, Christianized the naturalism of Greek philosophy. Like Aristotle and Plato, Rawls thinks, 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” (*BI*, p. 115). They think, he says, that “God is . . . a bigger and better object for our enjoyment, an object which shall . . . satisfy our various appetitions” (*BI*, p. 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

<sup>12</sup> See *BI*, p. 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.”

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. Thus 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*” (*BI*, p. 130, emphasis added). He notes that Augustine speaks of God as “our journey's end” (*BI*, p. 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*” (*BI*, p. 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.<sup>13</sup> 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 (*BI*, p. 182).

<sup>13</sup> 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, *Way of the Lord Jesus*, 34-a. For what is, in effect, an attempt to anticipate and rebut the reading of Aquinas, see McCabe, OP, *The New Creation*, p. xiii.

12 The undergraduate thesis

### §1.4 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 thoroughgoing, for we would expect him to object to any natural relations whatever on the grounds that the “energy” (*BI*, p. 118) which animates them is immoderate. 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 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” (*BI*, p. 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” (*BI*, p. 121). He thinks that 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” (*BI*, p. 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” (*BI*, p. 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.