

Human Nature Constraints
Upon the Realistic Utopianism of Rawls and Nussbaum

James M. Donovan
May 2007

Human Nature Constraints Upon the Realistic Utopianism of Rawls and Nussbaum

The maxim that “ought implies can”—frequently offered as a paraphrase of Hobbes¹ or Kant²—suggests that moral duties can accrue only to those actions that lie within the abilities of the subject. Even were a philosopher able to construct an elegantly conclusive proof that everyone would be better off if we all grew wings and flew (rather than, say, driving automobiles), no duty could possibly follow from this demonstration because the growing of wings lies outside the ability of any person. Whatever our duties may ultimately be found to be, they will be drawn from that list of things we can imagine currently constituted human beings actually performing.

The “ought implies can” standard raises a particularly challenging threshold of feasibility for the political philosophies articulated by both John Rawls and Martha Nussbaum. Although diverging on points both fundamental and superficial, each structures their argument according to the following generic outline:

1. Human nature is X;
2. X entails a particular view of the good, either in itself or when conjoined with some other mechanism (such as Rawls’s original position); and
3. Political institutions that foster the realization of the goods identified in (2) are the ones most deserving of support by citizens.

The first premise is stipulatively framed in such a way that, if (1) is granted, (2) follows self-evidently, requiring little defense (e.g., humans are animals that must eat to survive, therefore it is a human good to have enough nutritious food available to stave hunger and promote growth and development). A full argument of the transition from (1) to (2)

would require consideration of problem posed by some version of the naturalistic fallacy, which typically bars normative conclusions about what is “good” from mere facts about what “is.”

The greater philosophical challenge arises when more than one human good has been identified, requiring a institutional balancing of competing demands for often limited human goods. That task becomes especially challenging when working—as do both Rawls and Nussbaum—from a posture that rejects utilitarian justifications that allow the misery of some to be offset by a greater increase of benefit to others. When their work is complete, both writers expect people of all comprehensive doctrines to be willing to revise their ideas about the principles of distributive justice after learning the ways in which their existing political structures conflict with the appealing objectives of (2). That is to say, having granted (1) and (2), only an irrational opponent would continue to support political institutions that undermine the ends of (2) rather than preferring those described in (3). But because such persons are by that fact “unreasonable,” their objection would not constitute a meaningful critique of the proposed political arrangements.

Within this form of argument, we can see that the ultimate form of the proposed political structure is sensitive to the initial assumptions about human nature. Human beings differently defined will necessarily require different kinds of social institutions, for if all asserted human natures could thrive within a single form of political system then the logical necessity linking the three steps of the argument disappears as does its power to persuade. Rawls, for example, expressly desires his argument “from the original position to be, so far as is possible, a deductive one” (*JF* §23.4). This desideratum would

be frustrated if the initial depiction of human nature did not carry with it consequences particular to that suite of premises.

Taken as a whole, the structure of this argument can be quite compelling. When well constructed, each step follows as an implication of the preceding premise. Viewed in terms of its parts, however, each tier presents its own special challenges for the philosopher. But because of its step-wise progression, the overarching goal to describe the best political system (variously characterized as that which promotes justice, or fairness, which may or may not be the same thing) falters at the outset if the argument has begun with an insupportable account of human nature.

Given that both Rawls and Nussbaum are concerned that their respective projects end with a preferred conclusion—to justify liberal constitutional democracies—the ability of their schemes to persuade skeptics requires that they each build upon an account of human nature that will consistently support their later assertions. Both scholars have, however, put upon themselves demands that go beyond such argumentative niceties. Each also intends to articulate conditions of distributive justice that could be achieved in the real world, and would deem the enterprise a failure were it lauded for its theoretical elegance but intrinsic unattainability. They both require, therefore, that their accounts of human nature be sufficiently complete and accurate that they describe real people in the real world, while also supporting the kinds of political structures they wish to build.

Two standards can help to determine whether these authors have surmounted the challenges they have set for themselves:

1. *Criterion of External Validity*: Are the claims about human nature consistent with the findings of the empirical disciplines working to investigate that very problem (e.g., anthropology and psychology)?
2. *Criterion of Internal Consistency*: Is the philosophical scheme consistent in its assumptions about human nature, or does that nature vary depending upon the immediate point the author wishes to emphasize?

The following sections will discern how well the works of Rawls and Nussbaum perform against these measures.

The starting point for the inquiry, however, will be a review of the task that they have set before themselves, the articulation of a “realistic utopia,” one that can arise in the world as it actually exists. The attributes of human nature, both explicit and implicit, that are essential to each system are then reviewed. The results of this analysis provide the input for critique under each of the two tests, those of external validity and of internal consistency. The final conclusion shall be that both philosophers fail both tests, although not necessarily for the identical reasons or with the same impact upon their theories.

The Realistic Utopia

Of the two, John Rawls has been especially explicit in his wish to craft what he terms a “realistic utopia.” An envisioned state is utopian in that it does not currently exist, but realistic to the extent that it could describe a possible condition of actual citizens. This idea was elaborated in *Justice as Fairness* in the following terms:

We view political philosophy as realistically utopian: that is, as probing the limits of practicable political possibility. Our hope for the future of our society rests on the belief that the social world allows at least a decent political order, so that a reasonably just, though not perfect, democratic regime is possible. So we ask: What would a just democratic society be

like under reasonably favorable but still possible historical conditions allowed by the laws and tendencies of the social world? (*JF* §1.4)

This desire to situate the most just society possible within the “historical conditions allowed by the laws and tendencies of the social world” (similar language appears at *LP* §1.1) places upon Rawls a heightened demand that he fairly represent those social realities.

Accordingly, he identifies two conditions that must be satisfied if a political philosophy is to be realistic:

[First] it must rely on the actual laws of nature and achieve the kind of stability those laws allow, that is, stability for the right reasons. It takes people as they are (by the laws of nature), and constitutional and civil laws as they might be, that is, as they would be in a reasonably just and well-ordered democratic society.... The second condition for a liberal conception of justice to be realistic is that its first principles and precepts be workable and applicable to ongoing political and social arrangements. (*LP* §1.2)

Here Rawls himself requires that his political philosophy comport with “people as they are (by the laws of nature),” meaning that if he has not accurately characterized that human nature, then by his own terms his project is fundamentally flawed.

In contrast to Rawls’s direct grappling with the criteria and conditions for philosophical realism, Martha Nussbaum’s position is less easily discerned. Certainly she is less open to allowing her philosophy to be hampered by empirical details. The relationship between intellectual tasks is decidedly tiered rather than mutually informed, with philosophy accorded pride of place: “Philosophy seems best at articulating basic political principles at a rather high level of abstraction, leaving it to other disciplines to think how, as institutions and their configurations change, those principles can be made reality” (*FJ* p. 307).

Although not explicitly defending a political realism, she sprinkles throughout her text an implied assumption that her thinking nonetheless heads towards that realistic utopia. Her exposition leaves the reader with an impression that she only performs minor tinkering upon Rawls, and therefore imports into her own theory any Rawlsian premise she does not directly challenge. This incorporation extends presumably also to his quest for a realistic utopia. Examples of her indirect endorsement of this goal include repeated invocations of anecdotes touching on the lives of three disabled children (Sesha, Arthur, and Jamie). These narratives are intended to both illustrate the problems to be solved as they manifest in the lives of real people, and also as a touchstone of correctness for her own views: how would her proposals positively impact the lives of these children?

Nussbaum further demonstrates that the realistic utopia represents her goal through her criticisms of alternative views that she reads as failing to rise to that standard. For example, she describes the position shared by Hobbes and Locke that society is bound by ties of mutual advantage as “a particular picture of who we are and what political society is [that] has for some time imprisoned us, preventing us from imagining other ways in which people might get together and decide to live together” (*FJ* p. 414). This picture is historically particular, and only one among many possibilities. “We should by now be able to see that it is a picture...rather than a realistic description of what people are and must be.” This criticism of the prevailing perspective has substance only to the extent to which her own position is not vulnerable to the same objection. Tellingly, we again see that this hoped-for realism must be grounded in premises about “what people are and must be,” which is to say, in premises about human nature.

In her final chapter Nussbaum voices a concern whether her capabilities approach may be “hopelessly unrealistic” (*FJ* p. 410), reminding us that realism has not been deliberately built into the weave of her theory to the same extent that Rawls achieved. Her attempts to diffuse that worry, consequently, are ad hoc, but expressive nonetheless of a concern to rebut the doubt. She wants her ideas to be realistic, even if she has not systematically ensured that that will be the final outcome.

Both writers under discussion, therefore, submit their work for judgment against the standard of realism, and both understand the realism at issue to involve descriptive rather than solely prescriptive statements about the subjects of political philosophy. Both Rawls and Nussbaum intend their characterizations of human nature to be at least accurate as far as they go, and so much the better if they are also uncontroversial.

Human Nature

Pinning down a philosopher’s assumptions about human nature can be a tricky feat. While explicit statements constitute low-hanging fruit, Rawls also points out that “We can see what we have assumed by looking at the way the parties and their situation have been described” (*JF* §23.1). Granting that both Rawls and Nussbaum intend their political philosophies to describe social institutions suited to real people as they are found in the “here and now” (to use Rawls’s frequently-repeated phrase), we have warrant to consider both direct and indirect evidence to ascertain the contours of human nature operating within the philosopher’s ethical thinking.

The principal feature distinguishing Rawls’s foundational characterization of human nature from that of Nussbaum concerns the respective place of benevolence

within each thinker's approach. Rawls explicitly eschews a role for any third-party interests within the original position:

[The] circumstances of justice obtain whenever mutually disinterested persons put forward conflicting claims to the division of social advantages under conditions of moderate scarcity.... A further assumption is that the parties try to advance their conception of the good as best they can, and that in attempting to do this they are not bound by prior moral ties to each other. (*TJ* p. 128)

These claims represent what Nussbaum called Rawls's commitment to "methodological simplicity and economy" (*FJ* p. 107):

It should be noted that I make no restrictive assumptions about the parties' conceptions of the good except that they are rational long-term plans. While these plans determine the aims and interests of a self, the aims and interests are not presumed to be egoistic or selfish. Whether this is the case depends upon the kinds of ends which a person pursues.... But the postulate of mutual disinterest in the original position is made to insure that the principles of justice do not depend upon strong assumptions. Recall that the original position is meant to incorporate widely shared and yet weak conditions. A conception of justice should not presuppose, then, extensive ties of natural sentiment [i.e., benevolence]. At the basis of the theory, one tries to assume as little as possible. (*TJ* p. 129; see similar language at *JF* §24.2)

Rawls does not demand that human beings be solely interested in securing their own advantage. Once the veil of ignorance has been lifted, he says, it may well be the case that "the parties find that they have ties of sentiment and affection, and want to advance the interests of others and to see their ends attained," perhaps even at cost to themselves. But the circumstances of justice require that there be competing claims for scarce resources. If one party is willing to sacrifice her own interests to advance another's, such altruism removes the episode from one properly analyzed from the perspective of political justice. Justice requires competing claims, which foregrounds motives of self-interest at the expense of other-directed sentiments: "In an association of saints agreeing

on a common ideal, if such a community could exist, disputes about justice would not occur” (*TJ* p. 129).

While therefore not denying the (possible) role of altruism or benevolence in human lives, Rawls does exclude such feelings from the original position and thus from the moral psychology operable when specifying the requirements of the basic institutions of society. Persons outside the original position may tend toward benevolent consideration of others, but the institutions in which they will live will not depend upon that fact, but will be fully just even if they are in real life as self-concerned as they must be in the original position. The basic political institutions will be designed to contain selfish beings, although the actual inhabitants may be something better.

In contrast Nussbaum rearranges priorities to effect a dramatic alternative. Benevolence is not merely a possible moral style operative after the institutions of justice have been established; rather, it fundamentally shapes those institutions at the outset. Her capabilities approach “requires people to have very great sympathy and benevolence, and to sustain these sentiments over time” (*FJ* p. 409). “The benevolence that that full inclusion of people with impairments requires is extensive and deep, requiring the willingness to sacrifice not only one’s own advantage, but also the advantage of the group” (*FJ* p. 122). Human beings are of course not motivated or tied only by benevolence, but this is the moral sentiment that works as the prime mover within her political philosophy. Human beings, she has written, “are born for mutual aid and mutual accord” (Harpham 2002 p. 68, quoting Nussbaum’s *Therapy of Desire*).

She is able to demand a priority of fundamental benevolence because her conception of the circumstances of justice differs from Rawls’s. Instead of viewing

justice as arising between parties of roughly equal power in situations of moderate scarcity (ideas taken from Hume), Nussbaum follows Grotius in that demands of justice arise from “the twin ideas of dignity (the human being as an end) and sociability... [which indicate] that advantage is not the only reason for which human beings act justly” (*FJ* pp. 36, 37).

Nussbaum’s approach generates a wholly different result from that offered by Rawls. Rawls had defended the conclusion that the original position yields his famous two principles of justice, including the Difference Principle with its exclusive attention upon economic welfare. Nussbaum on the other hand offers her readers a list of capabilities, all of which “are implicit in the idea of a life worth of human dignity” (*FJ* p. 70), and thus expressive of what it means to be a well-functioning human being. To have the claimed relationship to human dignity requires that each itemized capability offer expression to some intrinsic need within the nature of the human animal itself—otherwise its inclusion on the list would be contingent and not necessary, and its absence would be unfortunate but not determinative of the kind of dignity worthy of a human being *qua* human.

Rawls has his own list of primary goods, but importantly these do not flow from the minimalist account of human nature Rawls has employed. They follow instead from his characterization of the person *qua* citizen: The “account of primary goods does not rest solely on psychological, social, or historical facts. While the list of primary goods rests in part on the general facts and requirements of social life, it does so only together with a political conception of the person as free and equal, endowed with the moral powers, and capable of being a fully cooperating member of society” (*JF* §17.1).

In sum, Rawls's initial assertions about human nature required by his theory are truly minimal: Whatever else they may be, persons are self-interested and seek to maximize mutual advantage in the original position; the resulting basic social institutions will be designed accordingly. Nussbaum favors a more elaborate picture of human nature, one that accords special privilege to the sentiment of benevolence which occasions the demands of justice as expressed in the ten capabilities.

Criterion of External Validity

The "then" clause of an if-then statement becomes intrinsically more interesting to the extent that the conditions described in the "if" clause are more accurate. That observation often leads philosophers to ground their theorizing in appropriate conclusions of the empirical sciences. To avoid asserting "oughts" that transgress the "can" limitation, it indeed seems prudent to begin with a firm grasp of the nature and abilities of the entity under discussion.

John Finnis has offered a forthright description of this strategy. Identification of the basic values—those "indemonstrable but self-evident principles shaping our practical reason"—

is, in a way, an attempt to understand one's own character, or nature. The attempt thus parallels attempts made, in quite another way, by those anthropologists and psychologists who ask (in effect) whether there is a human nature and what are its characteristics. The anthropological and psychological studies ought to be regarded as an aid in answering our own present question—not, indeed, by way of any 'inference' from universality or 'human nature' to values (an inference that would be merely fallacious), but by way of an assemblage of reminders of the range of possibly worthwhile activities and orientations open to one. (1980 p. 81)

The introduction of anthropological and psychological data, Finnis suggests, is not to serve as a micromanaged replacement for ethical thinking about how people should

live—there remains, after all, the barrier of the naturalistic fallacy that prevents any immediate translation of the former into the latter. These data, however, do serve to delimit the range of ways of living that human beings have found. They do not tell us what people ought to do, but do tell us what people are capable of doing.

Finnis’s strategy can be favorably contrasted with the considerably looser approach applied by Nussbaum, who concedes that her list “does indeed rely on intuition—although not on uncriticized preferences.... That is, some deep moral intuitions and considered judgments about human dignity do play a fundamental role in the theory” (*FJ* p. 83). Basing her capabilities approach upon “sympathetic imagining” rather than ethnography or ethology suggests a repugnance by Nussbaum for what others would consider the best evidence by which to evaluate her claims. Accordingly her conclusory statements about human nature are not preceded by an informed review of the relevant scholarship culled from interdisciplinary sources. While there perhaps exists no uniform requirement within a work of philosophy for such wide-ranging control of extra-disciplinary material, it does mark a curious lack in a work aspiring toward realism.

The challenge for Nussbaum is that if benevolence—a moral sentiment operationally defined as the “virtue that disposes us to care about other people’s flourishing” (Feldman 2000, p. 1436)—is to serve the role she requires, it must be an evolved interactional default at the group level. The question is *not* whether humans are self-interested or altruistic: both Rawls and Nussbaum would argue that they are both. The issue is whether one offers a more reliable predictor for the way people will behave—what, lacking any other information, would offer the “best guess” about

behaviors in a given situation? What kind of people should just political institutions be prepared to accommodate?

Infamous examples such as the murder of Kitty Genovese, witnessed by thirty-eight persons who neither rushed to her aid nor even called the police (Rosenthal 1999), would seem to argue against any position that people are “born for mutual aid.” Without denying that perhaps all persons are able to act benevolently at some times, if benevolence exists only as a personal relationship style that is either sporadic in its display (I will care about the flourishing of the world’s unfortunate during the Christmas season, but give it little notice during the rest of the year) or limited in its extent (I will care about the flourishing of my family and friends above all others and let the others fend for themselves), then it cannot do the work she requires. Human nature is best documented by identification of human universals (Brown 1991, p. 148), not from behaviors at the tail of the curve. Even if higher primates can be taught to use human language, it does not follow that they are a language-using species but only that some exemplars of the species, when exposed to environments that are extreme and unusual from their own habitats, can display linguistic behavior. So too, even if irregular or individual acts of benevolence to strangers can be identified, it does not follow that benevolence is meaningfully part of our nature *without also* admitting that that same evidence supports the related conclusion that such acts are rare, or at least drastically less common than Nussbaum requires and thus an insufficient basis for the design of political institutions.

The question we must ask is whether a fundamentally benevolent nature justifying the identification of the ten capabilities contradicts anything established by the social

sciences. As Nussbaum has articulated her ideas, benevolence is indistinguishable from the related, more technical term of *altruism*. Altruism denotes “behavior that benefits an unrelated individual(s) while being detrimental (cost > benefit) to the actor in the short term” (Moore 1984). Thus, when I donate money to children in Africa out of my food budget, I am acting altruistically; when I donate the same money to the same cause, but instead out of my vacation Starbucks allowance, or give my own food money to help my children pay a bill, my acts are generous without rising to the higher threshold of altruism or benevolence.

Nussbaum requires altruism and not simple generosity because she admits repeatedly that to raise everyone to the thresholds for her capabilities will be extremely expensive, and must thus be paid for by diverting money and resources from other projects to improve society. If the level of special education she demands for the disabled necessitates the closing of libraries, museums, and universities, she gives no indication that this should not be done (Bobonich 1993 p. S88). Those who view such institutions as necessary to the fully human life are therefore being asked to make an altruistic, or net cost, sacrifice.

While much can be debated about the state of various anthropological or evolutionary psychological findings, a well-established result has been that true altruism would evolve with only great difficulty. Most observed cases can be successfully analyzed in terms of *reciprocal altruism*, or altruistic acts that accrue a short term cost but with an expectation that over the long term the exchange will be balanced in some way or another (e.g., I give money to the beggar because, since I may be begging next year, I am better off living in a society in which passersby give money to random

beggars). *Genetic altruism* involves other acts of apparent benevolence that fall short of what Nussbaum describes. It can make genetic sense to sacrifice my own life if doing so preserves the lives of relatives of sufficient number and relatedness.

Reciprocal and genetic altruisms are species of the mutual advantage described by Rawls, and rejected by Nussbaum. Again, she requires that people be willing to sacrifice personal resources to strangers with no expectation of ever receiving a return gift or personal advantage of any kind because the intended recipients (the disabled, animals) are not able to make the return.

The conditions for an evolved altruistic sensibility are unlikely to have pertained during the evolutionary significant segments of human development, leading one theorist to conclude that “evolutionary models of ‘altruistic’ behavior should be treated with caution” (Moore 1984). Thus, for example, the kind of altruism Nussbaum requires to be part of our human stock “cannot evolve in a species or be stably sustained in a social group unless the cognitive machinery of the participants allows a potential cooperator to detect individuals who cheat, so that they can be excluded from future interactions in which they would exploit cooperators.... In this context, a cheater is an individual who accepts a benefit without satisfying the requirements that provision of that benefit was made contingent upon” (Cosmides & Tooby 1997). While some persuasive modeling for evolutionarily sustainable cooperative behavior may be feasible (e.g., Bergstrom & Stark 1993), these remain limited to those within the circle of family, neighbors, or acquaintances. Because such rationales do not extend easily to strangers, they cannot support the institutional edifice Nussbaum envisions.

The implication for Nussbaum of this research corpus would seem to be that altruism is either not the type of psychological motivation that could come to characterize the broad default posture of human interactions, or that, to the extent ideologies arise that would argue the contrary (e.g., Christianity), these represent exaggerations of minor and parochial strategies into frequently maladaptive or at least disadvantageous moral imperatives (e.g., Laura Nader [1990] demonstrates that the Spanish employed Christian teachings of peace and non-quarrelsomeness in order to pacify the Native Americans and prevent them from objecting to their forcible colonization by the Europeans). In either case, altruism does not arise in the situations or at the rates Nussbaum requires for her capabilities theory to have prima facie plausibility much less realism.

It would exaggerate the state of knowledge on this problem to conclude that Nussbaum is wrong in her assumption that altruism/benevolence offers an adequately reliable and pervasive platform upon which to build realistically utopian political institutions. The literature is such, however, that the burden of proof would fall upon her to demonstrate why she believes this is a reasonable initial position for her arguments. Failing as she does to situate her assumptions within the relevant contrary literature, her thesis about human nature does not contain the needed external validity at the necessary points. While she has shown us why her positions may be desirable, she fails to demonstrate that they can be real.

These same arguments seem to ratify the Spartan assumptions Rawls employs in his own thinking. Again, the literature from the relevant sciences do not say that he is “right,” but do support a burden-shifting conclusion upon a critic who would contradict his minimal assumption that human beings are self-interested, which expresses itself in

the search for mutual advantage when negotiating with parties of similar powers amidst moderate scarcity. His problems arise elsewhere.

Rawls's difficulties arise with the psychology he depicts as operative within the original position itself. This can be a serious limitation in his exposition, for he expects the original position to be a thought experiment that a person can run by an act of imagination. While we can perhaps imagine ourselves as lacking information we ordinarily have available (i.e., the veil of ignorance), the entity in the original position must still be "us" in some fundamental sense. The psychology of that entity (the second of the two social science disciplines named by Finnis as relevant to the philosophical account of human nature) must be recognizable as being "ours," thus making the outcome of the experiment binding upon us.

The conceptual difficulty comes when seeking to strip the negotiator of all specific knowledge about personal attributes, and thereby necessarily also all personal history, and the elements one expects to comprise personal identity and sense of self. Simultaneously Rawls expects this cognitively naked creature to still have the psychological motivation to engage in the search for mutual *self*-advantage that is the reason for entering the original position in the first place. The parties are expected to negotiate their self-advantage, while lacking any concept of the self.

Such a curious condition requires specification of the psychological theory Rawls assumes for human beings. He must describe more fully how beings would behave in the original position, in a way that we would recognize ourselves in those beings and thus feel bound by their conclusions (cf. Levin & Levin 1979). As Rawls leaves the matter, however, inertia seems the more likely outcome than negotiated resolution. On the other

hand, if the parties have enough self-interest to engage in the negotiations at all, they are likely also to have enough personality to accept something other than Rawls's principles of justice. Some might be very willing to gamble on a better outcome rather than accept Rawls's assumption that human beings are inherently risk adverse (another psychological claim he does not support).

Without a properly articulated psychology, the options for Rawls are bleak: either there is more variety of motivations within the original position than he would allow, throwing into doubt the "deductive" nature of the resulting two principles of justice, or the negotiators are some inhuman creatures with whom we could not possibly identify, rendering their agreement irrelevant to us.

The political philosophies of both Nussbaum and Rawls are the weaker for their failure to employ sufficient methods to ensure their external validity. Nussbaum does not bear her burden of showing how, despite the preponderance of anthropological data to the contrary, it is still reasonable to regard human beings as fundamentally benevolent, and Rawls lacks an adequate psychology that would both assure our imaginative participation in the original position while preserving the certitude of the outcome of that negotiation.

Criterion of Internal Consistency

While a description of a realistic utopia must be evaluated by the constraints of empirical sciences, philosophers are perhaps more comfortable critiquing such theories in terms of their internal consistency. The commitment to a given picture of human nature should extend throughout the writer's discussion, and not fundamentally contradict itself at disparate points throughout the text.

Not surprisingly, Nussbaum appears to be particularly vulnerable to this criticism. Her requirements of human nature are more complicated than Rawls needs for his own work, and thus there arise more opportunities for some contradictory claims to slip into her thinking. In the first part of *Frontiers of Justice* she describes how humans are fundamentally, and not merely incidentally benevolent, and that this concern for the flourishing of others imposes upon us the duty to assure that all persons enjoy the ten capabilities at the threshold levels at least. This arrangement, she argues, would “gather broad cross-cultural agreement” (*FJ* p.79), building as it does upon “an intuitive argument...that a life without a sufficient level of each of these entitlements is a life so reduced that it is not compatible with human dignity.... The idea of what human beings need for fully human living is among the most vivid intuitive ideas we share” (*FJ* pp. 278-279).

Because she has based them upon such firm intuitions, Nussbaum does not see her political arrangement as needing argument but only commitment. The needs she identifies are overwhelmingly intuitive and thus difficult if not impossible to deny. Indeed, because we have foregrounded the moral sentiment of benevolence, we cannot deny them, although we could still quibble about details (what are the capabilities, what is an appropriate threshold for each, etc.).

Within these early pages, her argument may strike the reader as sustainable on its own terms. The pertinent difficulties arise when she seeks to expand her thesis to encompass others of the frontiers of justice, especially the relations between species and between nation-states. When discussing the former, she acknowledges that “Nature is not just, and species are not all nice. We cannot expect that they will become nice, or

supportive of the good of their enemies” (FJ p. 390). Many would find this apt summary to extend to the human species as well, not least because Nussbaum, speaking from her vantage point as a neo-Aristotelian has emphasized the animality of human beings. She seems to expect something more from humans than she does from other animals, without fully explaining why that might be.

The deeper conflict arises when she is speaking about the divide within our species itself. Whereas one might anticipate that she sees no problems of scale for her political institutions, she comes down decisively against anything resembling a world government, calling such a development “dangerous” (FJ p. 313).

The threat she sees is not philosophical, but practical, and arising from the moral weakness of people in government: “Since corruption is one of the problems, in modern nations, that most severely threaten human capabilities, *mechanisms to detect and prevent corruption*, both in government and in business, are absolutely essential to the stability of the capabilities and the conception based on them” (FJ p. 312). The flaws of world government, she insists, are inherent to the political organization and not a flaw of execution: “[Any] world state is *ipso facto* tyrannical” (FJ p. 314).

This claim is curious, given her earlier description of people—and assumedly therefore also the people who constitute governments and hold political power—as being basically benevolent. She expects such benevolence to command the field and dominate the design of political institutions at the local and state level, but then implies that something qualitatively different occurs in the context of the “world state.” Yet, if her earlier argument has convinced us, we should believe that benevolent politicians would have no greater difficulty resisting their baser impulses once they have world power than

they will when they hold national prominence. Because that is not the case, we seem to be dealing with a different kind of human being in these passages than that described previously. At the very least, we are entitled to demand that Nussbaum explain why the human nature sufficient at one level yields to something less admirable on another.

Rawls's minimal assumptions can actually handle this task quite nicely. The principles of justice require contesting interests with parties of roughly equal power. In a world state, there is no opposition of equal power, and thus the circumstances for justice fail to arise. Rawls, in other words, easily accommodates a claim that a single human nature would generate concerns for justice at lower levels, but result in possible tyranny in the context of a world government. Nussbaum, however, has rejected that minimal description of human nature, while trying to retain the result that follows from it. This she cannot do, without at least more justification than she offers.

The flaw in Nussbaum's argument is that she would like to sever, when they prove inconvenient, the links of logical necessity required in the argument outline described above in the introduction. In that framework assertions about human nature require specific conclusions about the good, which in turn justify the description of political institutions of a particular kind. Without this structure, the final proposal has only the merit of one person's private preferences, rather than an outcome binding on all reasonable persons. Rawls makes this same error, although at a different point in his argument.

A central theme within Rawls's exposition is the conviction that all persons should be regarded as free and equal. In one sense this is assumed in his adaptation of Hume's circumstances of justice. If the negotiating parties within the original position

are not free and equal vis-à-vis one another, then one party would have an inherent political advantage over the other, obviating any need to negotiate just terms. Accordingly, this idea, “given by our moral and political thought and practice” (*JF* §7.6), forms one of “our firmest considered convictions about the nature of a democratic society as a fair system of cooperation between free and equal citizens” (*JF* §12.4). Later, however, what was first offered as a foundational insight of such necessity and certitude that it is able to function as a premise going into the original position receives a starkly different description: “If citizens of a well-ordered society are to recognize one another as free and equal, basic institutions must educate them to this conception of themselves...a conception which, if left to their own reflections, they would most likely never form, much less accept and desire to realize” (*JF* §16.2).

The dilemma Rawls creates for himself should be evident: a well-ordered society must intervene to create the idea of the free and equal citizen, which is a necessary precondition to enter the original position from which will spring the well-ordered society. This circularity should be read as deliberate. Rawls is keenly interested to build into his system the conditions for its long-term stability (*see JF* §§54-60). Instead of the linear progression of the argument model, a more stable form is the hermeneutic circle because it does not admit of external influences. The original position is not meant by Rawls to be a one-time experiment, but rather an ongoing exchange wherein the decisions of one cycle of negotiation in the original position are tested in the real world, the results of which are taken back into the original position for further negotiation (this is the process of *reflective equilibrium*, *JF* §10).

A tension therefore exists between the linear form of the logic Rawls applies to argue his case, and the circularity of the process he actually envisions. From this view there is a certain intelligibility to the claim that the idea of persons as free and equal is both found and manufactured, but at the cost of his other goal to make the outcome of the original position as deductive as possible. Processes can be circular, but arguments should not. Rawls appears to at times confuse the two, with the consequence that his characterization of the initial features of the human condition become inconsistently described.

Both Rawls and Nussbaum incorporate threads that contradict their earlier, better considered claims about the human condition. For Nussbaum, if people are benevolent to the degree she requires, and lacking a feasible explanation to the contrary, that account must hold whatever the scale of the political institution she is considering. If the internal inconsistencies within Nussbaum appear to arise from inadequate commitment to the earlier account, for Rawls the inconsistencies appear to be an inevitable by-product of the kind of thesis he intends to assert.

Conclusion

As Christopher Bobonich (1993, p. S92) reminds us, “The idea of basing an ethical theory on human nature has attracted Western philosophers from the very beginning of philosophical reflection on ethics.” Unfortunately many philosophers have not kept pace with the best research in the empirical disciplines exploring that topic, with the result that their works contain more impractical idealism than they intend. In order to ensure the soundness and persuasiveness of their ethical conclusions, philosophers should routinely confront the relevant social scientific literature and situate their initial

assumptions within that corpus. Before expounding on how humans ought to live, the writer must first understand how they do live.

Rawls and Nussbaum have offered their own contributions to this deeply complex project to ground political philosophy in assumptions of human nature. While both fall short of the goal of complete success, the implications of that shortfall differ for their respective ethical systems.

Nussbaum's failure to satisfy the criteria of both external validity and internal consistency arise from her failure to include necessary arguments to defend her positions (e.g., that it is reasonable to insist that human beings are basically benevolent rather than self-interested) or to accept the implications of those positions even when they contradict her other intuitive preferences, such as an opposition to world government. The limitations within the work of Rawls, however, strike the reader as being contained within the structure of this theory itself. Both his unconvincing psychology within the original position that requires ignorance of and interest in the self, and confusion as to whether the conviction that fellow citizens are free and equal is an initial intuition or a learned social belief, each claim—the thesis and its antithesis—are critical at different points within his justice as fairness argument. These are not gaps in the discussion such as Nussbaum commits, but rather serious tensions within Rawls's conceptualization itself. While Nussbaum's shortcomings are thus of the quality of an error, Rawls's are a problem that may or may not be remediable without a significant reconceptualization of his larger project.

The present review has shown that Nussbaum's work contains promissory notes on matters that should have been more completely elaborated before asking her readers to

commit to her vision of political justice. These missing links in her argument are all the more surprising in light of the “gross flood of words appearing under her name” (Harpham 2002 p. 52). While it may be the case that given time she can fill in the gaps, it is also possible that fuller details on these fundamental issues will force significant rethinking about later features of her scheme. In this sense Nussbaum has rushed prematurely into print with her account, but supporters can still hope that when the full argument is finally available it will deliver the hoped-for impact.

In contrast the present discussion has identified no gaps within Rawls, but instead genuine tensions. Again, it may be the case that further work can explain why these seeming inconsistencies are more apparent than real, promising real work for philosophers, as compared to the imagining in which one might indulge to anticipate Nussbaum’s next move—the difference, perhaps, between the need for a literary analysis of a Dickens novel versus the speculation as to the contents of the final Harry Potter volume. Rawls’s offering therefore emerges as a more mature philosophical text that rises or falls upon our understanding of what he has actually written within its pages, without recourse to unpublished materials that may or may not satisfactorily address the questions.

While the Nussbaum’s capabilities approach may one day eclipse the justice as fairness model of Rawls, today is not that day.

ENDNOTES

¹ “[A]s much as is possible...to more no man can be obliged” (Hobbes 1886, p. 70).

² “The command ‘we *ought* to become better men,’ resounds with undiminished force in our soul; consequently, we must be able to do so” (Kant 1898, p. 353).

REFERENCES

- Bergstrom, Theodore C., & Oded Stark. (1993). “How Altruism Can Prevail in an Evolutionary Environment.” 83(2) *American Economic Review* 149-155.
- Bobonich, Christopher. (1993). *Internal Realism, Human Nature, and Distributive Justice: A Response to Martha Nussbaum*. 90(Suppl.) *Modern Philology* S74-S92.
- Brown, Donald E. (1991). *Human Universals* (Philadelphia: Temple University Press).
- Cosmides, Leda, & John Tooby. (1997). *Evolutionary Psychology: A Primer*, at <http://www.psych.ucsb.edu/research/cep/primer.html>.
- Feldman, Heidi Li. (2000). “Prudence, Benevolence, and Negligence: Virtue Ethics and Tort Law,” 74 *Chicago-Kent Law Review* 1431-1466.
- Finnis, John. (1980). *Natural Law and Natural Rights* (Oxford: Clarendon Press).
- Harpham, Geoffrey Galt. (2002). *The Hunger of Martha Nussbaum*. 77 *Representations* 52-81.
- Hobbes, Thomas. (1886). *Leviathan, or, The Matter, Form, and Power of a Commonwealth, Ecclesiastical and Civil* (London: George Rutledge and Sons).
- Kant, Immanuel. (1898). “First Part of the Philosophical Theory of Religion,” In *Kant’s Critique of Practical Reason and Other Works on the Theory of Ethics*, translated by Thomas Kingsmill Abbot (London: Longman, Green).
- Levin, Michael E., & Margarita Levin. (1979). “The Modal Confusion in Rawls’ Original Position,” 39(2) *Analysis* 82-87.
- Moore, Jim. (1984). “The Evolution of Reciprocal Sharing,” 5 *Ethology and Sociobiology* 5-14.
- Nader, Laura. (1990). *Harmony Ideology: Justice and Control in a Zapotec Mountain Village* (Stanford: Stanford University Press).
- Nussbaum, Martha C. (2006). *Frontiers of Justice: Disability, Nationality, Species Membership* (Cambridge, MA: Belknap Press of the Harvard University Press) [FJ].

Rawls, John. (1971). *A Theory of Justice* (Cambridge, MA: Belknap Press of the Harvard University Press) [TJ].

Rawls, John. (1999). *The Law of Peoples* (Cambridge, MA: Harvard University Press) [LP].

Rawls, John. (2001). *Justice as Fairness: A Restatement* (Cambridge, MA: Belknap Press of the Harvard University Press) [JF].

Rosenthal, A.M. (1999). *Thirty-eight Witnesses: The Kitty Genovese Case* (Berkeley : University of California Press).