

*Sharon A. Lloyd*

## **Stepping Back\***

*Abstract:* Although Rawls insists that his argument for his theory of justice neither addresses nor requires that we settle in advance any of the deep questions of philosophy, there are nonetheless more subtle ways in which his work may bear on such questions. The article explores how Rawls's work may advance our thinking on the general philosophical question of how language affects thought, by enabling us to assess the conceptual consequences of two alternative metaphors for describing our activity when we engage in critical self-reflection. The effects on our thinking of our common use of the metaphor of 'stepping back' to describe this activity are contrasted with those of an alternative metaphor suggested by Rawls's work. The resulting conclusions are then employed to illuminate Rawls's reply to an interesting objection to his theory raised by Michael Sandel.

The importance of Rawls's work for the field of political philosophy is by now beyond dispute. But the richness of that work is perhaps best seen by examining how it may facilitate progress on some of the most basic questions of philosophy generally. I use the deliberately vague expression "facilitate progress" because, as Rawls has recently made plain, it is not his aim to address such questions, and the success of his own project in no way requires their resolution. Rawls's argument for his particular conception of justice as fairness is intended to be shallow, relying only on thin political conceptions, and aims to increase its acceptability across a wide range of deep philosophical views by neither presupposing nor promoting any. Nonetheless, a piece of work need not address nor even directly apply to a question in order to facilitate progress on it, and it may be of interest, particularly to those who think the shallowness of Rawls's work a defect, to see how that work may facilitate progress on one of the fundamental questions of philosophy.

The question I have in mind is the question of how language affects thought. It is a commonplace among philosophers since Wittgenstein that how we talk may profoundly affect what we can think. We suspect that our language largely

---

\* I am grateful to Robert Brenner, Zlatan Damjanovic, Nick Pappas, Marya Schechtman, Larry Solum, and Paul Weithman for helpful comments on earlier versions of this essay.

determines what we see as problematic, what questions we can pose and what we will count as answers to them. This seems plausible, but it is not an easy matter to show that it is true, in part because showing this requires that we be able to envisage possibilities other than those our language makes compelling, so as to see our present views as contingent and linguistically conditioned. This in turn requires that one exercise sufficient imagination and intellectual stamina to resist the imposed picture and make lively a plausible alternative picture. Because Rawls has succeeded in doing this, I propose in this paper to use Rawls's philosophy to develop a concrete example that exposes the mechanism that underlies the alleged connection between language and thought for a particular case of current philosophical interest.

The case I want to consider is the case of evaluating our desires, aims, and values. I will discuss our widespread and common use of the metaphor of "stepping back" to describe what it is we do when we engage in self-reflection, adopting a perspective appropriate for evaluating our desires, values and aims. I shall suggest that our use of "stepping back" talk leads us to adopt a puzzling picture of ourselves, and in some cases may actually distort the philosophical theories that it is intended to illuminate. But, as I shall argue, from Rawls's theory there emerges an alternative conception of our activity - along with a new metaphor - that may enable us to avoid these unhappy effects of our current linguistic practice. To begin, let us look at the context in which this talk of stepping back arises.

## I.

We normally think of ourselves as having certain beliefs and desires, as committed to certain values, and as identifying with certain aims and ends. We have, to use Charles Taylor's phrases, a *stand* on issues of importance, and a sense of *where we are coming from* on them, to which we feel we must refer in trying to describe who we are (Taylor 1990). Nonetheless, we do not view these features of ourselves as permanently fixed, immune to revision. Our sense is that these features can be altered, and that when they change, *we* change, in ways that may be for the better or for the worse. Self-improvement is possible; so is self-stunting error. We don't generally think just *any* revision is possible (and thus are fascinated, and put at a loss, by those rare apparently wholesale transformations of character from Sauls into Pauls), but we do think that limited revision of these things is possible, even usual. And we think that *we* are able to bring the change about, as a result of our assessment of our own qualities. The question is how we do it. What is it that we are doing in evaluating our desires, values and aims?

One answer attractive to philosophers is that we are metaphorically stepping back from our desires, values and aims in order to adopt an impartial perspective from which to decide whether or not they are worthy of our affirmation. We are

stepping *outside* of ourselves, in order to get a better perspective *on* ourselves. Not that we are literally able to do this, but that, metaphorically speaking, this is what evaluating these members of our motivational set is like.<sup>1</sup>

This stepping back picture is brought to bear on philosophical debate by the following sort of argument. Some philosophers seem to suggest that in deciding on the good life for ourselves we ought to consider what sort of life we would see as good if we were fully informed and perfectly rational, given our current scheme of desires, interests, and aims.<sup>2</sup> This fully informed and rational viewpoint is argued to be the most appropriate one for determining what course we ought to pursue given our motivational set. Part of the appeal of this view is that by taking the motivational set as given, it allows us to maintain an internal connection between what one judges to be one's good, and thus thinks one ought to pursue, and what it is that one can be motivated to pursue. Let us call theories of this sort Type G (G for given motivational sets).

We can imagine the critic of Type G theories charging that all such theories must be defective insofar as they take motivational sets as given, since, after all, human beings are the sorts of creatures that can "step back" from their current desires, interests and ends - this is part of what is distinctive about human beings - and any theory that presupposes some determinate fixed scheme of motivations fails to appreciate part of what it is that we are trying to decide in deciding on the good life for ourselves. Part of what we're trying to decide on is exactly *what* set of motivations to *have*, on what sort of person to *be*. We don't want to take our motivational sets as given, nor do we necessarily want to decide as a perfectly informed and rational person with *our* character would decide, since we may have motivations that we don't admire. We want to step back from our current motivational set, so as to assess *it* in terms of its contribution to the good life for ourselves. The objection comes down to the claim that humans are largely self-constituting beings, who can decide not only what goals to embrace *given* who they are, but also who to be, and that theories that do not give due attention to this capacity must be defective.

It is in this sort of objection that the stepping back metaphor naturally appears, and it is interesting to note that the most obvious reply to this sort of objection depends upon resisting the idea that revision of one's motivational set need involve anything like stepping back. Indeed the objection itself seems to arise from a failure to appreciate the resources available to the proponents of

---

<sup>1</sup> The use of this common metaphor is not confined to philosophers. Consider General Thomas W. Kelley's criticism of U.S. press reporting on the Gulf War: "I think the press needs to take a step back and reflect on how they cover things." Reported in the New York Times, March 5, 1991.

<sup>2</sup> Early versions of this view can be found in the Ideal Observer tradition of Smith and Hume, and later of Firth. One might include among more recent variants the works of Rawls and Habermas; and elements of this view appear as well in the theories of Gauthier and the Analytical Marxists.

Type G theories which is directly traceable to the picture of what goes on that is evoked by the notion of stepping back.

The obvious reply available to a Type G theorist is that the criticism rests on an importantly incomplete rendition of the facts about a person that are to be taken into account in idealized deliberation. Among the facts that must be taken as part of my motivational set, and thus carried into the idealized point of view, are precisely facts about what sort of person I wish I were - facts about, for example, my desire to be more generous than I am, my interest in being a more generous person, and my goal of becoming a more generous person. Such facts about me as that I have these desires, interests, and ends will have to be included in the idealized point of view from which I am to make judgments about my good. We should not resist adopting the idealized point of view for fear that such facts will not carry appropriate weight in my idealized deliberations, since they will appear as having whatever weight they actually *do* have. If the desires and aims I have to be a more generous person are sufficiently strong that they could, say, induce me now to resist adopting my own idealized point of view on the grounds that I would in that position decide as a greedy person would, then they are most likely also sufficiently weighty that they would carry priority in the idealized point of view, regulating the choice I would make from it.<sup>3</sup> If, on the other hand, I have *no* desire or aim to be a more generous person, then on this construal of a Type G account, the mere thought that perhaps I ought to want to become one will not bear any regulative force in the idealized point of view. But, of course, it *shouldn't* bear any regulative force there, *if* the internal connection is to be maintained. In contrast, on the 'stepping back' account, it is not clear how giving such a thought inordinate weight is to be prevented.

Since no idealized picture of me would be complete without these self-critical, potentially regulatory, interests of mine, the idealized point of view will clearly have to include them among the relevant facts about me to be taken into account in idealized deliberation. Once we see that my wishes about what sort of person to be, provided these wishes carry some regulative force, will affect judgments about my good from the idealized point of view, we can see that 'stepping back' *need* not yield any result different from that of taking a thorough inventory of all of the desires, values, and aims I actually have. And if it does, it does so by *severing the internal connection*. We may not care about maintaining an internal connection, but if we do, we must resist the picture of deliberation that stepping back presses on us; for though when I step back I might well consider the thought that perhaps I ought to want to become the sort of person who would care most

---

<sup>3</sup> The philosophical literature on higher order interests attempts to give sense to this sort of weighting or prioritizing among desires which allows some to regulate our pursuit of others. This attempt involves fashioning room for revision of our desires and ends from within our given motivational set, rejecting the idea that revision requires any metaphorical 'stepping back'. In appealing to higher order interests we evoke a very different picture of what goes on in evaluation than that illicitly by the stepping back metaphor.

about becoming a championship golfer, if I adopt this as the appropriate point of view from which to determine my good, it is not the least bit clear that I as I am will be moved to act on the conclusions 'I' reach from this point of view.

## II.

Now one of Michael Sandel's objections to Rawls's theory of justice<sup>4</sup> is the mirror image of this objection to Type G theories of the good. Sandel argues of Rawls's theory for how we are fairly to regulate people's pursuit of their comprehensive conceptions of the good that Rawls's theory requires us to step back from our constitutive commitments - from the values, interests, ends and ideals that make us who we are - into the alienated, unencumbered point of view of someone who has no such commitments, and who might, for that reason, equally well be anyone. Sandel questions both the desirability and the possibility of assuming the required point of view. Even supposing it were *possible* to step back from our constitutive commitments in this way, the worry is that if Rawls's theory does insist that we step back from all we care about, and so, in the end, from all we are, then we may reasonably wonder what possible justification could be offered us for doing so. *Why* should we do this, and how could we be expected to view ourselves as bound to take principles determined from that point of view as *regulative* of our pursuit of what really *does* matter to us? The puzzle, at bottom, is to see why we should care about what we would decide from a point of view in which we ignore all that we care about. This, I think it is fair to say, is where the force of Sandel's objection lies.

I say this objection is the mirror image of the earlier objection to Type G theories of the good because whereas that earlier objection claimed that our natures as self-constituting beings who can choose who to be is not respected by Type G theories, and it should be, this objection of Sandel to Rawls is precisely that Rawls thinks we can and should adopt a view that presupposes that we are *entirely* self-constituting - that we can assess, ignore, or whimsically and willfully *choose* the interests, ends and ideals that make us the persons that we are - but that, on the contrary, we neither can nor should be required to do this. Sandel's idea is that no theory that insists that I distance myself from those commitments of mine that are constitutive of my identity can have any real claim on me. Clearly then, Sandel too seeks to maintain an internal connection between (in this case) what principles of justice I ought to choose, and what principles I can be motivated to act on.

To descend to particulars, Sandel argues that Rawls's imposition of a veil of ignorance on the parties in the original position with respect to their religious, moral, political, cultural, and affectionate attachments, and other aspects of their

---

<sup>4</sup> This objection is developed in Michael Sandel 1982.

comprehensive conceptions of the good so strips the occupants of that point of view of the features that distinguish them as individuals that the decision such disembodied subjects would reach is *not* one that the real people whom the parties are supposed to represent could see as binding on *them*. The parties cannot effectively represent real people because they have been made to step so far back from the constitutive commitments of real people that they cannot represent *them*; real people could not be moved to act on the principles their would-be representatives would choose. Once again, just as was the case with a 'stepping back' criticism of type G theories, on this 'stepping back' interpretation of Rawls's original position, the internal connection is threatened. Wherever one comes down on the question of the desirability of preserving internal connections, the interest of this point is in seeing *why* 'stepping back' accounts threaten internal connections.

### III.

Stepping back is a way of distancing ourselves. As we distance ourselves from things, their effect on us diminishes. As I step back from the fire, I feel its heat less; and if I step back far enough, I will cease to feel it at all. One might step back from an object in order to get a simply different perspective for say, viewing it (and I will talk more about this idea later), but if mere *alteration* of perspective were all the stepping back metaphor were intended to convey, the idea of stepping up, or standing closer, would do just as well. The central idea of stepping back is one of putting distance between oneself and the objects of one's consideration. And the point of this distancing is that at a distance, (in time, it seems, as well as in space), things have less influence on us. We feel less connected with them.

The picture of oneself the notion of stepping back summons up is one of viewing any of one's desires, aims, and values at a distance; and not just at a distance, but at such a distance that they no longer appear as particularly one's own.<sup>5</sup> In stepping back we lessen our connection with these desires, so that we might view them, as it were, impartially, or from anyone else's vantage point. It is a picture of adopting a perspective independent of our desires, and in this sense, outside of our own mental life. This diminution of the force our desires exert on us, to the limiting point that we cannot feel them at all, and they might as well be someone else's, is what threatens the internal connection. The thought that these desires are not particularly ours may lead us to see them as external to ourselves, issuing in the fiction that they can be somehow up for grabs, to be either affirmed or rejected from the outside; a fiction, since they are at least provisionally given along with the other members of our motivational set. And this

---

<sup>5</sup> This picture is developed and discussed in Thomas Nagel's suggestively named book, *The View From Nowhere*, 1986.

fiction is what ushers in the apparent possibility that *I* should see it as good for me to do or be something that I could not possibly be moved to do or to be.

Once we think of our activity as stepping back, we may imagine that if we stepped far enough back we could step outside of our entire motivational set, to evaluate even the crucial elements of it upon which the more dispensable elements hinge. It is this thought that leads to a deceptive polarization of positions: *Either* I operate from within my own perspective, in which case my self-transformative capacities are utterly excluded - shown to be illusory - or I operate from the unconstrained vantage point of the observer at a distance, in which case most any change is within the realm of possibility (whether I could be motivated to effect it or not), and I might be or become most anyone. Depending upon *where* I step, so goes the logic of stepping back, either I'm in or I'm out; it has to be one or the other.

This *polarization* of possibilities pressed on us by the picture of stepping back is one unfortunate consequence for our thinking of talk that employs this stepping back metaphor. To see that it is a consequence, rather than a cause, of our use of the stepping back metaphor, let us contrast stepping back with an alternative metaphor, suggested by Rawls's work, for what it is that goes on when we evaluate our desires, aims and values.

#### IV.

If stepping back leads us to a problematic picture of radical disconnection from any or all of the elements of our motivational set, then we may prefer an alternative metaphor that better reflects our sense that when we evaluate our particular desires, aims and values there must be something that stands fast as constituting some perspective that can be called ours (though perhaps *what* stands fast may vary from case to case).

This perspective will not involve entirely disclaiming the desire, aim or value under consideration. It will not be like viewing the desire from such a distance that we can scarcely sense it anymore, eventually losing any perceptual grip on it at all. It is more like arranging our desires, aims and values in a neat array in front of our mind's eye and then cocking our heads, and narrowing our eyes - call it, for lack of a more elevated term, *mental squinting* - in order to get a slightly different perspective on the very same set of things. The things remain, and remain equally, ours. But we squint at them with a mind to discerning some feature common to the whole range of desires, aims and values, or some new aspect of the individual elements, in the way that we might squint at a painting with an eye to seeing just the play of light, or just the colors, or just the shapes of the painting. We don't squint willy nilly; we squint purposefully, in a *cognitively guided fashion*, aiming to see some previously specified aspect or feature of the raw material already present to us. And, of course, we may have good

reasons for thinking one way of squinting better *for a purpose* than another: if, for example, our purpose is to understand the technique employed in fashioning a painting, we may do better to attend to the brush-stroke than to hunt for duck-rabbits.

The point about understanding evaluating our desires, aims and values on analogy with squinting is that here, we alter *our own* perspective, rather than trying to so distance ourselves from the desires, etc., as to adopt a perspective that is not particular to ourselves. The metaphor of squinting suggests that the judgments rendered in this way are no less one's own than those made straight on, eyes opened. Nor is there any presumption that one sees less clearly in this way than in the ordinary. It is a different way of seeing that may be better for a purpose than the ordinary, and that is *equally* one's own. But it *is* different, and it is this that makes revision of the elements of one's motivational set possible.

Despite its unnaturally silly name, "mental squinting" is a natural notion that corresponds to an array of activities familiar from ordinary experience. We engage in this activity when we listen to a complex piece of music with an ear to hearing a particularly subtle leitmotif, or just the strings, or just the melody, harmony, phrasing, or even story line or sheer emotion of the piece. We do it when we attend to a wine aiming to discern its woodiness, or fruitiness, or acidity, or sample a beer with a taste for its distinctive hoppiness. Those with the stomach can handle a snake with a feel for its weight, or its movement, or texture, temperature, density, or any of a number of other snakey qualities. One can even attend to kinesthetic sensation in this way, by metaphorically squinting at one's own breathing, or heartbeat, or discrete muscles, with a mind to experiencing the sensation of falling, or floating, or being smashed under a crushing weight. We even engage in mental squinting in the absence of any sensory stimulation. In examining a text, say, we may attend to it a number of times, each time with a different purpose in mind, enriching our view of it with each successive pass. We can, for example, read Jane Austen's *Pride and Prejudice*, and then calling it before our mind's eye, attend to it with a purpose to understand it as a commentary on self-knowledge, or on love and courtship in the early 19th Century, or as an exposé of the folly of the idle rich, or as a work on women's oppression. With each cognitively-guided consideration of 'the stuff' before our mind's eye - the passages and structural features of the text, the story, the words used in telling it, the style, voice, and tempo of the writing, and all the rest - we can *revise* our understanding of the novel. When we consider it with the purpose of reading it as a commentary on self-deception we may discover to our surprise that the seemingly reckless Lydia is actually less self-deceived than thoughtful Jane whom we had formerly admired. But notice that when we do this 'mental squinting', we revise from within the material already given. The revision is not due to adopting an 'external' perspective, as it might be if, say, we were to change our view of *Pride and Prejudice* by considering it in light of *The Great Gatsby*.



Squinting then, is an alternative way of exploring, or of elaborating, amending, or assessing what is given to us in consciousness. But, unlike stepping back, it does not involve any attempt to diminish allegiance to our own perspective by distancing ourselves from the elements of that perspective.

This is not to say that there are never occasions on which it would be appropriate or useful to 'step back' from one of one's desires or aims - perhaps when a desire is habitual or an aim obsessive it would be appropriate to try to distance oneself from it. But these occasions are quite limited. It is not in general true that increasing our distance from the objects of our scrutiny is the proper way to gain a perspective appropriate for evaluating them. Using again an analogy with the visual case, there is usually a more or less determinate best distance for viewing things that is given by the things themselves, in conjunction with such variables as background and the perceiver's faculties. I don't gain a better perspective for evaluating a painting by moving to within three inches of it, or by stepping back the length of a football field. Where I should stand depends in part on the size of the painting, and on its many internal features such as contrast and brightness, in part on the character of the background against which I am viewing it and the medium through which I am viewing it, in part on the settled peculiarities of my own perceptual apparatus, and in part on what it is for something to be a painting seen. If I get too close, parts of the painting escape my notice or my perception altogether, I may lose all sense of light and shapes, and see *not a painting*, but merely blotches of color. If I step back to a great distance, I may see *not a painting*, but one of many furry rectangular patches of gray in an ornate picture of a room. To see *a painting*, one has to stand at *the proper* distance. And when I do stand nose to the glass in front of a Seurat, it is not to gain a *better* perspective for evaluating it, but is rather precisely to marvel at the fact that anyone could have seen from the proper distance while painting at arm's length.

If a painting has its own proper viewing distance, given in part by what it is for our perception of it to count as perception of a painting, perhaps something similar is true of our scrutiny of such things as our aims and desires. Pursuing this line of thought we might consider whether, e.g., for our scrutiny to be of our goals, we need to be close enough to them that we still feel ourselves aiming, or for our scrutiny to be of our desires, we must not be so distanced from them as to be unable to feel ourselves desiring, while nonetheless being sufficiently distanced from them that we experience ourselves as having, rather than say being, the goal or desire. If there is a proper distance for considering our desires and aims that is partly given by what it is for our consideration to count as of a desire or aim, then there would not, on the face of it, be any reason to suppose that stepping back yields a superior perspective for evaluating them, as least in the ordinary case.

Now even if we are attracted to this revision of the picture of what goes on when we assume a point of view useful for evaluating our desires, aims and values, and suspect that what we do is *more like* squinting than like stepping

back, we may still desire a more straightforward and rigorous philosophical characterization of the activity that our alternative metaphor picks out. The operation in this case that corresponds to 'seeing with an eye to something', or 'listening with an ear to something', or any of the other sensory forms of selective attention, could be generally termed 'cognitively-guided attending'. We specify a purpose in considering some element or set of elements in our mental life, and then attend to them with this purpose in mind. Our purpose may be to determine their consistency, or simplicity, or commonality, or origins, or to ascertain their relation to some further belief or value - our sense of the openness of possibilities for revision stems from the fact that there are indefinitely many purposes with which we may scrutinize them, and each may alter our understanding of them, our attitude toward them, or even the things themselves. Nonetheless, since the objects of our selective attention are initially given, and since the range of purposes that may *profitably* be brought to bear in examining them is probably quite limited, our sense of the possibility for revision is of a constrained openness - room to maneuver - but not that just anything might go.

The crucial point of difference between the squinting metaphor and the stepping back metaphor is that the former conveys the idea that we can alter our perspective in profitable ways - in cognitively guided, reasoned, and defensible ways given a purpose - without *distancing* ourselves from the objects of our evaluation. Because it does not urge that we strive, in a quest, perhaps, for objectivity, to step 'outside' of our motivational set, this alternative metaphor of squinting eliminates the push to polarize positions forced on us by the notion of stepping back. Possibilities for self-revision and for self-constitution can be explained as effects of the qualitative transformation of the elements of our motivational set that occurs when we attend to these with a purpose in mind. We can choose from among many ways of selectively attending to our desires, aims and values, and each way may alter these. We are not compelled to accept our current perspective as exhausting what we can bring to deliberation, or to imagine our motivations fixed in a determinate array with specified relative weights that must determine some unique deliberative outcome. Nor need we, on the other hand, resign ourselves to the unsettling view that we might wholly recreate ourselves in any imaginable form, that any perspective might be equally appropriate for determining our good, or for determining principles of justice to regulate our pursuit of our good, and accept the possibility that the consideration of what is good for us, or of what are the proper constraints on our pursuing our good, might be something on which we, from our perspective, could never be moved to act.

## V.

With this alternative metaphor in hand, let us return to the dispute between Rawls and Sandel to see what difference to our thinking this new manner of speaking

can make. Recall that the underlying problem, as Sandel presents it, is of how we might adopt a perspective that allows us to articulate the relation of the self to its ends that presupposes neither an unencumbered self (as Rawls's is said to) nor (the alternative) a radically situated self that is wholly identified with whatever particular set of commitments it happens to have, with no possibility for self-criticism or revision short of self-annihilation. But why does Sandel set up the problem this way? How do we come to be presented with the choice between unencumbered selves and radically situated selves, and why do these two unsavory possibilities seem to exhaust our options?

The answer, I think, is that it is because the notion of stepping back foists this picture on Sandel. He speaks throughout of the problem of "distancing" the self from its ends, of "standing apart" from one's commitments, of a self "independent" of its values. The picture invoked is one of stepping back from my values and ends, we hope far enough back that I can see them as mine rather than as me, but not so far back that any connection I could have with them must be purely voluntaristic. The picture of stepping back is the picture *Sandel* calls up to understand what goes on when we adopt a perspective that allows us to evaluate (or in the present case, to adopt principles to regulate our pursuit of) our commitments. He characterizes Rawls's view as one in which "the self is distinguished from its ends - it stands beyond them, at a distance" (Sandel 1982, 59). The problem becomes how we might conceive of any intermediate possibility between utter alienation and radical situation; but the problem is *set* by the notion of distancing given in the idea of stepping back. Talk of stepping back, with its danger of stepping all the way outside of oneself into a perspective that is not one's own, dichotomizes the possibilities that Sandel can imagine; but our alternative metaphor shows that this limitation on imagination is merely linguistically, and not, as it were, ontologically, conditioned.

Rawls's reply to Sandel<sup>6</sup> centers around an elaboration of the relationship within his system between the partial, political conception of the person as citizen embedded in the original position, and the full characterization of persons complete with their comprehensive conceptions of the good to which appeal is made in carrying out the project of developing an overlapping consensus on Rawls's principles of justice. Of course, Rawls needn't enter the fray over moral phenomenology in order to elaborate this relationship. But one way of understanding what this elaboration does is as imposing a different picture of our activity in assuming a perspective from which to settle on principles for regulating our pursuit of our comprehensive conceptions of the good - a picture much closer to the squinting picture we have been considering than to the stepping back picture upon which Sandel's criticism is relying.

Rawls's reply is that we, in modern industrial democratic societies where pluralism is a fact, share a conception of citizenship and a conception of society that can be used to determine principles of justice (to regulate our pursuit of our com-

---

<sup>6</sup> This reply appears in Rawls 1985; 1987, and is further elaborated in 1991.

prehensive conceptions of the good) that could gain the support of an overlapping consensus. The first and last portions of this thesis - that we share these political conceptions, and that an overlapping consensus on Rawls's principles is possible - are partly empirical claims that cannot be evaluated until the facts are determined and the appropriate arguments produced. The middle portion of the thesis - that we can lay out an appropriate selection device for principles of justice from a philosophical examination of our shared conceptions of citizenship and society - is at once a more theoretical and a more analytic task whose success must be judged on the coherence, logical rigor, and plausibility of the actual argument. But the crucial features of the thesis Rawls draws attention to in his reply are these: First, the conceptions of citizenship and of society that determine the contours of the original position, specifying the assumptions and constraints on reasoning embedded in it, *belong to us* - they are conceptions that *we* affirm, and think, for good reason, defensible. Secondly, our *motivation* for affirming the principles of justice they yield as regulative, while partly provided by our higher order interests in exercising our two moral powers, (which are, again, *our* interests), cannot be fully described without reference to the possibility that these principles could be affirmed from within our comprehensive conceptions of the good in an overlapping consensus. Rawls's conception of justice is built up from commitments and interests that belong to us, and it is a conception that we can be brought to give regulative status from within our own comprehensive conceptions of the good. The picture is one, not of stepping back from our commitments, but of (1) focussing on one subset of our commitments - our political commitments - and attending to them with the purpose of determining principles of justice for the basic structure of our society (this is the argument from the original position); and then (2) with the principles in mind, focussing on our diverse comprehensive conceptions of the good and attending to them with the purpose of forging connections between them and our principles of justice (this is the project of developing an overlapping consensus, in which we strive to create at least a coherence, and better, a derivation from each of the comprehensive conceptions to the principles). This latter step makes it possible for adherence to the principles of justice to be motivated in a full way, by their connection to our most cherished values and beliefs.

Already one can see that nothing in this process requires us to adopt an alien perspective. We are *never* required to proceed from values that we do not in fact hold, nor does the process *as a whole* require us to neglect consideration of *any* of our values or other constitutive commitments. We do adopt two distinct perspectives as we carry it out, but both of these perspectives belong quite naturally *to us*.

Now if we consider this process from the point of view of what it is *like* to engage in it, we find that squinting talk makes better sense of it than does stepping back talk. According to Rawls, we believe that all that is relevant for purposes of citizenship is possession of the capacities (what he calls the two moral

powers) that enable us to participate as free and equal persons in a system of social cooperation on fair terms, and for our mutual benefit, over a complete life. Phenomenologically speaking, (though certainly Rawls himself does not intend to speak phenomenologically) we first 'squint' at these political commitments of ours with the purpose of selecting principles of justice that most appropriately express them. Our concentrated attention to them allows us to see what will have to be constraints on our selection of principles of justice, e.g., that equality is the benchmark against which alternative principles are to be measured, that departures from equality will have to be acceptable to all, and that the principles should not privilege some over others on the basis of such things as skin color, gender, religious affiliation, or moral or metaphysical views. Squinting at the political commitments 'blurs' the precise contours of our comprehensive conceptions of the good, constraining the way in which they can influence us. It enables us to maintain awareness of our comprehensive conceptions without attending to their particular characteristics, but it would be misleading to say that we step back from them, for *we still feel their force, and desire to protect our pursuit of them*. Part, after all, of what moves us to seek principles of justice in the first place is our desire to secure the political bases for pursuing our comprehensive conceptions of the good (as well as for exercising our capacity to form and revise them). As we concentrate on our political commitments with the purpose of establishing these fair terms of social cooperation among citizens as we conceive them, we arrive, by means of Rawls's meticulous argument for and from the original position, at principles of justice that best reflect these political commitments. We then turn our attention to our comprehensive conceptions of the good, 'squinting' at them in all their detail and particularity, with the singleminded purpose of interpreting them in ways that support the conception of justice arrived at from within our first perspective. Perhaps we can so interpret them, and perhaps we can't, (this cannot be known surely in advance, though Rawls provides us with good reasons for being hopeful), but this project of developing an overlapping consensus on our conception of justice is carried out from within the second of our own perspectives, this time fully attending to our constitutive commitments. Phenomenologically speaking, this is what our activity is like.

Now on *this* picture of what goes on in Rawls's system Sandel's criticism should seem misplaced. It would be a distortion to say that Rawls requires us to distance ourselves from our constitutive commitments, and thus offers us a conception of justice that we could not reasonably be expected to be motivated to affirm. Our political commitments probably *are* among our constitutive commitments; aspects of our comprehensive conceptions surely are; and Rawls's system requires guided attention to *both* of these. If we join Sandel in viewing Rawls's veil of ignorance as a distancing device - as a metaphorical stepping back - then the distorted assessment is inescapable. Rawls must either make appeal, or not make appeal, to all of our constitutive commitments in all their situated particularity: if he does, argues Sandel, then they will not yield Rawls's conception of

justice; if he does not, then we could not possibly be moved to take his conception of justice as regulative of them. But if we employ our alternative understanding of what goes on in this process of adopting the perspectives appropriate for determining principles of justice, we can see what Sandel's stepping back picture does not allow him to see: we can see how the political commitments to which Rawls makes appeal can be simultaneously 1) ours, 2) regulative of our pursuit of other aspects of our comprehensive conceptions of the good, and 3) not, in the end, the most important thing to us - not the ultimate source of our motivation.

## VI.

It seems highly unlikely that Rawls could have conceived of this new possibility had he succumbed to the distancing picture of self-reflection urged by the stepping back metaphor, with its concomitant polarizing effect. In resisting that picture he develops a plausible alternative that is better expressed by a different, yet no less natural, way of talking.

Rawls's work, and Sandel's criticism of it, allow us to fashion a concrete example of the way in which language may affect thought. What conclusion, if any, we should draw from this particular example of the effect on our thinking of the use of one metaphor rather than another seems to me to be an open question. We might resolve to avoid metaphors, similies, and other such devices, in order to minimize the danger that unacknowledged or unwanted pictures will shape our philosophizing. But this may seem a bit too monastic, since we have, after all, fruitfully employed many such devices as heuristics. A less conservative moral to draw would be that though we may profitably use these devices, we ought not to let them bear too much of a philosophical argument's burden. But that would, of course, be only a manner of speaking.

## Bibliography

- Nagel, Thomas (1986), *The View From Nowhere*, Cambridge  
 Rawls, John (1985), Justice as Fairness: Political Not Metaphysical, in: *Philosophy and Public Affairs* 14, 223-251  
 - (1987), The Idea of an Overlapping Consensus, in: *Oxford Journal of Legal Studies* 7, 1-25  
 - (1991), *Justice as Fairness. A Briefer Restatement*, typescript  
 Sandel, Michael (1982), *Liberalism and the Limits of Justice*, Cambridge  
 Taylor, Charles (1990), *The Dialogical Self*, typescript