14

Autonomy, Self-Knowledge, and Liberal Legitimacy

John Christman

In the Enlightenment tradition of the justification of political authority, institutions of state power are seen as legitimate only if such institutions can be freely supported by those living under them. Liberal legitimacy, then, assumes that autonomous citizens can endorse the principles that shape the institutions of political power. The conception of autonomy functioning in such a picture, moreover, requires that such citizens uniformly enjoy the capacity to rationally reflect upon and critically appraise their own values, moral commitments, and political convictions. In this way, political power is an outgrowth of autonomous personhood and choice.

This traditional understanding of political legitimacy has been challenged from any number of directions, most notably from those who charge that the picture of the autonomous person underlying the mechanism of authority is parochial, exclusionary, and in tension with the sought-for legitimacy it is used to support.[1] In this last vein, it can be charged that the requirements of general support for principles of justice in a modern, pluralistic society, are in tension with the assumptions concerning individual autonomy underlying that concept. For the problem facing liberal conceptions of justice and legitimacy is that political power can be seen as justified only when supported by autonomous citizens, but the requirements of autonomy, in many construals of that term, are too stringent to be met by the majority of citizens bound by political institutions. Or, in other versions of this critique, the conditions set out for autonomy refer best only to some in the population and not others, thereby valorizing certain personality types, value perspectives, and social positions over others. So modern institutions fail to

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

achieve the desired legitimacy. It will be this concern that we will deal with here.

More specifically, the difficulty I want to examine here is that for political institutions to be legitimate, citizens living under them must achieve, for example, a level of self-knowledge and reflective self-endorsement that most fail to meet and that, in fact, would run counter to other processes of value commitment and moral obligation that motivate our moral choices. Two parallel questions arise at this point: what exactly are the conditions of autonomy that best support the role that concept plays in principles of justice and legitimacy? And what reasons are there for assuming that citizens expressing endorsement of political institutions (of the sort required by liberal legitimacy) be autonomous in this way, especially when the conditions of such autonomy do not obtain universally for all in the population?

I will approach these issues by first focusing on the concept of autonomy, where I will examine the general pattern that theorists of that notion have followed, and propose a particular view on the concept's meaning, at least as it might be used in the context of liberal political theory. The problems that have been raised about seeing autonomy in this way – in particular that it would demand certain capacities and practices that are at once difficult to achieve for most of us as well as being disruptive of our most basic value commitments – will be noted. Indeed, I will add to the usual chorus of complaints on this score, pointing out the ways that some understandings of autonomy may require levels of self-understanding and reflection that few of us ever achieve (or would want to achieve). Nevertheless, I want to suggest that the process of legitimating principles of justice in the liberal tradition require seeing autonomy in this way. That is, despite the fact that people generally do not exhibit levels of self-knowledge that some conceptions of autonomy assume, it is nonetheless important to treat them as the fundamental representatives of their own values and commitments, and it is correspondingly important to ask them to reflectively appraise those commitments as part of the process of giving reasons that political legitimacy demands.

To keep track of the rather circuitous route I will be taking through these issues, let me lay out the plan: I will first discuss the concept of autonomy; in doing so, I will propose a version of that concept that takes competence and the capacity for self-reflection as central. Then I will consider problems with such requirements, in that understanding autonomy this way appears to assume a level of self-knowledge that most people cannot achieve. Moreover, acts of reflection can in some ways disturb moral

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

commitment and manifest aspects of personhood that are not definitive of the most settled aspects of the self. With these challenges laid out, I then turn to political theory, in particular to the requirements for the legitimacy of political authority in the liberal tradition. In doing so, I will make some general claims about the nature of liberalism, in particular its commitment to pluralism, rejecting certain forms of perfectionism, and in requiring citizen endorsement for all legitimate state institutions and the principles that guide them (the so-called "endorsement constraint"). I then distinguish two importantly different strains in liberal thinking – one in which legitimacy is established as a result of self-interested bargaining for the purposes of establishing stable social environments (within which citizens can pursue valued projects), and the other in which legitimacy is seen as grounded in a *moral* commitment to political institutions resting on mutual respect and reciprocity. And I support the latter view of political justification over the former. I then return to the question of the nature of autonomy, where I will claim that the mechanisms for establishing legitimacy in the strand of liberalism worth defending need not attribute levels of self-knowledge to citizens that they are unable systematically to meet (or if they are, they must be treated as meeting them nonetheless). And reflective self-appraisal of the sort demanded by liberal legitimacy is not problematic in the ways that our earlier concerns pointed to.

In the end, then, the kind of autonomy assumed in the mechanisms of liberal legitimacy does not assume levels of self-knowledge or capacities of reflection that citizens either cannot or would not want generally to exercise.

I The Conditions of Autonomy

Various conceptualizations of autonomy have been put forward, and the contrasts among these highlight differences in the way that this concept operates in different theoretical terrains.[2] In certain contexts, stress has been placed on the way that autonomy has traditionally rested on a single and parochial conception of the self – one, for example, that assumed a "true" or "core" self residing inside of us like an "inner citadel."[3] But as many have pointed out, there are several reasons to avoid reference to a singly conceived notion of a self in models of autonomy. For there are far too many contrasting conceptualizations of our selves relevant in various settings and relative to various needs for any one of them to unproblematically count as our authentic core. Our embodiment, for

instance, is sometimes the most prominent aspect of our person (in medical settings, for example), whereas in others our identification as a member of a particular group, religion, culture, or ethnicity is salient. Moreover, as communitarian critics of liberalism have repeatedly stressed, our identities are often constituted by our deepest value commitments.[4] But these foci of selfhood vary from context to context and hence cannot, singly, play the role of the "true self" of which autonomy is meant to be an expression. So insofar as a conception of autonomy assumes a model of selfhood that features one of these aspects to the exclusion of the others, it can rightly be labeled as overly narrow and hence problematic.

Some theorists have therefore approached autonomy, not as the operation of a core set of identity-creating characteristics, but rather as a range of capacities, competences, and functions. This "functional" account of autonomy may in a better position to avoid charges of narrowness that have plagued more traditional notions.[5] Such accounts focus on a number of conditions that manifest the "self-government" of the person, while at the same time acknowledging the deeply embedded, interpersonally constructed, and historically situated nature of the self. The first set picks out those characteristics by which a person effectively makes competent decisions: rationality, self-control, freedom from psychosis and other pathologies, access to minimally accurate information, motivational effectiveness, and the like. The second set refers to requirements that the person's values and decisions are truly her own; these most often include the condition that person's reflect on their personal characteristics[6] and identify with (or at least not feel deeply alienated from) them. Whereas the first family of requirements ensures that the autonomous person effectively acts (rules), the second guarantees that the ruling is truly her own. Therefore the self-rule promised by the etymology of the word "autonomy" is established.

So on the view offered here, autonomy requires that the person be able to submit the factors of her personality to critical self-reflection.[7] This requires that factors relevant to identity, decision, and choice be such that, hypothetically, the person could reflect upon them without repudiation in light of how they came about. In this way, the autonomous person is competent (in the says described) as well as authentic in the sense of being moved by values that would withstand self-scrutiny.

Note also the reference to the *history* of the agent relative to the trait in question. I have argued in earlier work that the processes by which a person develops a trait are relevant to her autonomy vis-à-vis that trait.[8]

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

The way that attention to personal history should be captured is that a person cannot be labeled autonomous if some aspect of the manner in which a characteristic is developed would, if known, cause her to disavow that trait, to become deeply alienated from it. Let us say that a person discovered that the only reason she remains so devoted to her revolutionary activities is that she was kidnaped and tortured at an earlier time (a memory she had suppressed until now). Her autonomy is clearly in question if, were she to realize how these attitudes came about, she would disavow them. However, what matters is the person's relation to the attitude or characteristic *given* its etiology rather than her attitude *toward* that etiology (*simpliciter*): I might think that the way I was raised was too restrictive, but I accept the way I turned out nonetheless, because it wasn't *so* restrictive that I want to reject or disavow the character traits that developed from it.

The requirement of self-reflection demands that the person is autonomous (relative to some factor) if, were piecemeal reflection in light of the history of the factor's development to take place, she would not feel deeply *alienated* from the characteristic in question. To be alienated from some aspect of oneself is to experience negative affect relative to it, and to experience diluted or conflicted motivation stemming from it, and to feel constricted by it, as though by an external force. It is, moreover, to feel a need to *repudiate* that desire or trait, to reject it and alter it as much as possible, and to resist its effects. If I reflect on some addiction I have, for example – one that I did not bring upon myself voluntarily – I view it as distanced from me, as something about which I feel regret or dismay and that is less than fully motivating (relative to non-alienated desires).[9] Moreover, the reflection required of autonomous agents is considered to be piecemeal, requiring that agents reflect on *particular* aspects of their character without ever presupposing the ability to look at the whole of themselves from a completely disembodied perspective.

Further, a mere *capacity* to reflect is too weak: if a person has a capacity to reflect on herself but never does, and some of her first-order traits would be unacceptable to her if she did, we would not call her autonomous as she continues blithely to act on the basis of those traits.[10] It is not merely that the person can reflect but that, were she to do so, she would not feel alienated in the manner described. Moreover, the capacity to reflect alone, even if exercised, seems insufficient to pick out a meaningful conception of autonomy. An unwilling addict, who may be unable to resist the debilitating grip of his destructive cravings, but nevertheless retains a tragically robust ability to reflect on his life and take in all of

its deficiencies, is not autonomous despite this tragic self-knowledge. So an autonomous person must be able to alter those characteristics toward which she feels resistance, alienation, and repugnance.[11]

Non-alienation is also a different condition from the familiar requirement of identification, which one typically finds in discussions of autonomy. On the one hand, I can feel no alienation toward a characteristic but not fully identify with it, in the sense of wholehearted endorsement without regret.[12] We all contain some measure of internal conflict and complexity, and an attitude of ironic acceptance of the tensions of our own psyches is inevitable, and perhaps healthy, in a multi-dimensional and perplexing world. But to be alienated in the sense I mean here is to be actively derisive of some aspect of the self, to want to reject and resist it. An alienated person feels no affinity with such traits, wants to change or, if that is not feasible, distance herself from them; she is a divided and conflicted person, and is unable to present a minimally settled sense of herself to others in practical discourse. On the other hand, non-alienation is stronger than identification when the latter is considered as mere acknowledgment: I can admit that a trait is, alas, part of my identity (especially in my motivational structure), but still not want to repudiate and distance myself from it. Therefore, on the present view, a person is not autonomous relative to those aspects of herself that would produce such feelings of self-repudiation were she to reflect on them in light of how they came about. (Notice also how non-alienation adds an *affective* element to autonomy, in contrast to the picture of the disengaged cognizer described in our earlier discussion of reasons-responsiveness.)[13]

One final point: for a person to be autonomous on this model, the hypothetical reflection being considered cannot itself operate under the influence of factors that effectively prevent normal self-awareness. This prevents the possibility of a regress when considering the ways in which manipulative factors constrain both choice *and* reflection.[14] So self-reflection – even the hypothetical reflection being considered here – cannot be the result of distorting factors that guarantee that the self-appraisal in question has a particular result. Such factors include the influence of drugs or substances that prevent settled concentration, torture or intimidation that prevents the person from considering alternative ideas, educational backgrounds that severely limit opportunities to raise questions and come to minimally independent conclusions, and the like. As we will notice later, this condition will need to be refined in light of the ways that we all engage in "distorted" self-reflection in systematic ways.[15]

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

Interesting challenges, however, have been raised about the conceptualization (and related valorization) of autonomy, challenges that concern both the "competence" conditions and the "authenticity" conditions. For example, critics have claimed that autonomy problematically assumes herculean powers of self-knowledge, that the competence assumed in such accounts demands that agents have understandings of their motives and inner selves that few, if any, tend to realize. Moreover, such competency requirements have tended to emphasize the intellectual capacities over the emotional and affective.[16] This is shown in the characterization of competence as "rationality" and reflective self-endorsement in terms akin to the justification of belief. Concerning reflection, critics have charged that second-order appraisals of first-order motives and habits often reveal less authentic aspects of the self and, worse, cause dangerous disruption in people's deepest commitments, disrupting settled and authentic agency rather than securing it. Such emphasis on reflective re-evaluation and revision of the self both causes and reflects an unmerited valuation of change, instability and hyper-mobility.[17]

I want to investigate these charges in greater detail, and indeed I will emphasize and support versions of these claims. For the sake of brevity, we can examine these concerns as focused on the general requirement of "competent self-reflection" assumed in models of autonomy. We will further discuss the specific conceptual conditions of autonomy later; for now, we can assume that the conditions of autonomy at issue involve the competent self-reflection and inner endorsement just described. The idea is that autonomy requires that the agent in question be competent in the sense that she suffers from none of the disabilities that would systematically hamper reflective decision-making and that she exhibit minimal abilities to reflect, choose, and act. As a result of such reflection, the agent must not repudiate the characteristic in question to be autonomous. Let us survey, then, problems raised about such a model.

II Difficulties With Self-Reflection

There are many initially compelling reasons to resist taking the reflective functions of the person as centrally indicative of her autonomy. Two families of reasons can be given on this score: one is that reflection itself is often costly, and carries with it effects on commitment and devotion that raise questions about its role in self-determination; a second is that the reflective voice in all of us often does not speak for our most settled and authentic personae in that such voices can cover over or mis-diagnose

the inner workings of our psyches. Let us look at these concerns more closely.

The first set of problems involve the way in which reflectively questioning our commitments and motivations can often disrupt and undercut those very commitments. This problem of first-order motivational distortion can best be brought out in a two-person case: consider longtime spouses or romantic partners. One day, one of them enters the breakfast room to announce that she has lately been reflecting on the value of the relationship for her and on her commitment to it. Now, even if the result of such re-thinking is to redouble the strength of her commitment, the partner hearing this may well be disappointed and shocked, and the ties between the two deeply shaken. Now if we collapse this dynamic into a single mental life, we have cases where self-evaluation leads to self-doubt and diminished motivation.[18] The paradox is that if a person reflects, she loses the autonomy she seemed to enjoy before the moment of re-appraisal.

Second, critics have charged that in many ways, our introspective judgments fail to reflect our settled, authentic selves. Such reflections merely give voice to a rationalizing super-ego attempting to quash the more central elements of our motivational system, elements that, if allowed to move us, would issue in action that is more truly our own. For an illustration of such a phenomenon, consider the character Jude in Thomas Hardy's *Jude the Obscure*. For a good part of the novel, Jude is clearly in love with his cousin Sue, though he is still married to his estranged wife Arabella, to whom he still feels a strong obligation of fidelity (backed by all the force of his North Wessex Christian upbringing). But Jude's most basic motivational drive is clearly his love for Sue, evidenced by the cold sweats he experiences at the thought of her leaving, and his fits of jealousy at the sight of her with another man. Reflecting on these emotions, driven by the thought that he is still officially married and that Sue is, after all, his cousin, Jude mis-characterizes these emotions as merely those of a platonic concern of a friend toward a family member. As the events in the novel soon bear out, Jude's true nature is not revealed by his reflective voices but by those first-order affective drives.

Now, in addition to revealing the important place that emotions have in the specification of our authentic selves, this case indicates how the voice of reflection may distort rather than clarify our self conceptions. Reflection, for Jude, produces profound alienation from his emotions and destroys whatever authentic motivation he might experience were he, as he eventually does, to allow his feelings of love to move him to

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

act. Only without the self-reflection that autonomy demands (under self-reflection views) can Jude, and those like him, act authentically.[19]

The other set of problems for requiring reflection of this sort concerns the inaccuracies (so to speak) of the judgments made from the higher order perspective of our reflective selves. For it is clear that only a marginal proportion of the self implicated in behavior and social interaction can ever be said to be available to conscious reflection, both generally and at any particular time. Factors connected with embodiment, demeanor, habit, and the emersion of the self in the ongoing flow of events operate outside of the purview of reflection, and often completely beyond its scope. Hence, a person's inner picture of her motivational matrix can be highly incomplete and, in many other ways, innacurate.

Psychoanalysis provides one of the starkest models of the self's misunderstanding of itself.[20] The fundamental theoretical commitment of psychoanalytic theory is the postulate that mental contents that are not integrated into the dominant – that is, consciously available – schema of self-organization exert influences on thought and behavior. The picture that emerges is, of course, of a conflicted and non-rational psychic mechanism whose operations are accessible to conscious reflections only in distorted form or through the mediation of therapeutic intervention or other complex self-interpretive techniques.

Of course, psychoanalysis is controversial, and many rightly raise questions about the reliability of (at least the details of) the postulates it produces concerning sub-conscious mechanisms. But evidence of systematic self-misunderstanding can be gleaned from several other traditions in individual and social psychology.[21] Cognitive dissonance theory, for example, trades on the postulate that a fundamental operation of mental reflection is to embrace propositions that accord with established self-conceptions and resist those that destabilize them, independent of the epistemological ground of such information. Internal coherence trumps probable truth. In addition, any number of (often self-serving) biases shape judgements about internal states, capabilities, and traits. A general tendency has been observed toward attributing responsibility for positive outcomes to the self while explaining failures in terms of environmental factors (the "self-serving attribution bias"). Also, people will find flaws in evidence that portrays them in an unflattering light, and they are selective in sorting through memory when considering evidence for desirable traits.[22]

More generally, what psychologists call the "fundamental attribution error" refers to the systematic tendency to mis-estimate the role of either

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

personal characteristics or environmental factors in explaining one's own or others behaviors.[23] People routinely and predictably attribute behavior to character traits or dispositions when overwhelming evidence (available to them) indicates otherwise. And agents' perception of themselves and their own motives follows the same pattern as observation of others: the tendency to misidentify motives and the causes of our behavior is psychologically ubiquitous.[24] And our appraisals of our own emotions and attitudes tends to be equally prone to "error," indicating little, if any, advantage we have in having direct (introspective) access to such feelings.[25]

Material of this sort, admittedly presented here only selectively, does much to bolster skepticism concerning the possibilities of self-transparency. What emerges from these several angles is a picture of systematic self-delusion or, at best, a fundamental disconnect between introspective understanding and actual structures of motivation, thought, and behavior. In these ways, to the extent that autonomy demands that we reflect accurately on our motivations, desires, and reasons, most of us are systematically heteronomous in identifiable ways.

Indeed, this is a more empirically-minded way of expressing what commentators writing in a post-modern mode have been saying about the liberal conception of the self for some time – namely, that such a conception wrongly assumes a transparent, unified, fully rationalized self-conception of a sort no one realistically can realize.[26] Even when avoiding the psychoanalytic models mentioned earlier, such critics decry the fiction of a fully self-transparent consciousness as a basic presupposition of the model of the (autonomous) person at the heart of liberal theories of justice.

With all these reasons for questioning the reliability of our reflective functions in capturing and representing ourselves, why should we continue to require reflective endorsement of any kind for autonomy? Answering such a question involves two complicated steps. The first is to examine the role the concept of autonomy plays in various theoretical and practical contexts, here the context of liberal political theory, thereby locating the manner in which self-reflection figures in that dynamic. The second is returning to the concept of autonomy and refining the conditions of self-knowledge that (1) capture what is required by the concept's role in those political/theoretical settings, and (2) squares with the information just outlined concerning the systematic limits of the typical person's self-understanding. What I will suggest is that autonomy, when viewed a certain way, plays a role in the legitimation of political principles in such a manner that reflective self-appraisal will be a crucial requirement, despite (and in come cases because of) the complex effects

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

that such reflection has on motivation, self-understanding, and social interaction.

III Autonomy and Varieties of Liberalism

The context in which the conception of autonomy at issue here will be tested is that of modernist liberal theories of justice, ones where political authority is generated by way of citizen endorsement of collective social values. Liberal views reject a metaphysically ordered hierarchy of values, and thereby embrace a degree of value pluralism. No single overriding value and no fixed ordering of values can be determined to be objectively valid for all agents, on this view.[27] Liberalism rests on the idea, then, that political power is legitimate only if it is endorsed or accepted by citizens living under it "in light of their common human reason" (as Rawls puts it).[28] This implies that the principles expressive of this power rest on respect for citizens' abilities to rationally endorse the content of those principles. Therefore, liberalism rests on respect for individual autonomy as conceived generally as the "moral power" of judging both principles of justice and conceptions of value.[29] This respect is afforded equally to all and is reflected in the manner in which both basic principles and more specific social policies are derived (that is, democratically).

Justice, then, is formulated in a way that expresses this respect, where people are considered ultimately able to reflect upon and embrace (or reject or revise) conceptions of value for themselves.[30] Liberalism can be seen to rest on the fundamental valuation of persons as having a basic interest in pursuing their own conceptions of what is valuable, and doing so "from the inside." This conception of justice as the set of principles claimed as legitimate by those living under them utilizes what some have labeled the "endorsement constraint" on value assumed in liberal theory.[31]

It is important to note how liberalism, in this general sketch, is fundamentally opposed to certain kinds of perfectionism (in both moral and political theory). Although there are varieties of perfectionist liberals, most of those views take the fundamental (perfectionist) value to be respected by just institutions as autonomy (or its equivalent) itself.[32] What liberalism of all these sorts opposes is the view that there are values or moral imperatives that are valid (for a person) independent of that person's subjective appraisal, and hence first-person endorsement, of that value. Not only the (European) medieval worldview concerning a metaphysically structured value scheme in which humans played only

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

a part, but contemporary views of the objectivity of value must be put to one side here, at least as a means to provide foundations for political principles.[33] The tradition of political thought in which autonomy plays a crucial role and is the subject of examination here is one that contrasts deeply with that perfectionist standpoint.

But we must recognize a sharp distinction in liberal views of the authority of the state. In one, which we can call the Hobbesean variant, collective choice (via either the original social contract or ongoing democratic mechanisms) is seen as an aggregation of individual rational desires. The purpose of political institutions, on this view, is to provide stability and peace in order that citizens may pursue their own rational life plans, separately and for their own reasons. The ground of political authority, in this tradition, is self-interested rationality manifested in strategic interaction with others.[34] In the second tradition, emanating from Locke, Rousseau, and Kant, citizens are understood to have a *moral* connection to the authority of the state, insofar as such authority is a collective manifestation of their own autonomy. Collective choice, on this model, is simply the social version of the independent self-government that grounds all morality and obligation. Political authority, then, is grounded in a moral obligation (rather than simply rational bargaining).[35]

Indeed, we can generalize this distinction to apply to any social interaction whose purpose is to generate norms that will, in turn, constrict, guide, or constitute the resultant activities of the participants. In the Hobbesean case, agents view each other in a purely strategic manner, where knowledge and empathic understanding of the other's experiences or perspective are, at best, of instrumental importance to the interacting parties. There is no constitutive relation between recognition of the thoughts, preferences, and experiences of others and the binding nature of the outcome of such an interaction. Whatever one's social compatriots think or feel, on this model, one relates to them as instrumental to the achievement of the outcome of the exchange. Call this the purely strategic relation.

In the other case, the interpersonal exchange involves at least a respectful understanding of the other's perspective (at some level of abstraction or description), an understanding that is a crucial component of the reciprocity involved in this kind of social dynamic. And this attempt at understanding forms an ineliminable part of the normative grounding of the outcome. That is, participants view both the process of collective deliberation and confrontation, as well as the result, as normatively significant in part because of their shared understanding and projected

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

moral judgment. Such an interaction must involve mutual respect and sense of reciprocity in the familiar Kantian sense, where one attributes basic moral weight to the capacities of one's co-citizens to deliberate and decide. But it also includes an attempt at empathic grasping of the subjectivity and motivations of others. This need not involve a flawless or even accurate understanding of another's deliberative processes, but it does require an attempt to see the point of view (together with affective and subjective elements of it when relevant) of those with whom one shares a common relation to a collectively formulated outcome, incomplete though this process will inevitably be. Again, this can take place at virtually any level of abstraction, rising, say to the point of merely saying, "I think I can understand what it is like being a motivated human who is passionate about a cause such as that." We will call this empathic respect.[36]

The normative hold that the outcomes of this type of interaction has on participants will be constitutively related to this emphatic respect. Such a theory will be grounded much more firmly in one's own perceived value-commitments. In the case of strategic interaction – the Hobbesean model – one's commitment to outcomes of interchange extends only as deep as one's occurrent self-interest, and that outcome and commitment remain stable only as a function of the initial power relations that made the compromise with the objectified other possible. We will return to this point later.

Liberal political theory, then, presupposes a conception of the (autonomous) person that is both the object of respect (upon which those principles are built) and the model for basic interests that those principles protect. (Rawls's use of the index of social primary goods as a measure of just distributive shares is an example of this, based as it is on the projection of persons as capable of forming and embracing conceptions of both justice and the good.) Parallel to this commitment, though, is the liberal presupposition of value pluralism noted earlier. Liberal theory developed (historically) by rejecting various medieval and Scholastic metaphysical conceptions that postulated a teleologically structured order of the universe. These rejected pictures of the world served to specify completely the virtues and values for both individuals and societies. Liberalism, in both its Hobbesean and Kantian varieties, replaced this metaphysical framework with (what would later galvanize into) a conception of moral commitment with the human *will* at its center. Political principles, then, and the sense of obligation binding citizens to them, are seen as grounded in the individual and

collective judgments of the people involved, expressed by their rational wills.[37]

Value pluralism is the understanding that various individuals will embrace irreducibly divergent, but equally valid, moral conceptions. And political principles must take into account what Rawls calls the "fact of reasonable pluralism" – citizens pursue divergent comprehensive moral conceptions but recognize this divergence itself and accept it as a permanent fact of modern life.[38] Social values and the political principles reflective of them are generated (in part, at least) by way of collective choice and deliberation, and not given fully formed from above.

Such public endorsement of dominant political values must also occur against the backdrop of the inevitabilities of social existence. That is, contrary to the traditional assumption of a state of nature by which to measure the benefits of a specific political arrangement (a pre-social arena to which disgruntled citizens can retreat), political principles are judged by citizens who take the ongoing, historically embedded dynamics of social existence as an unavoidable fact (along with the pluralism of value-conceptions this brings with it).[39]

Therefore, interaction and collective deliberation among divergent viewpoints is fundamental to the process of legitimation and justification of social power. This view has been the dominant theme in the recent work of both Rawls and Habermas. The latter has developed the most complex picture of the centrality of discursive communicative action in the justification of both moral and political principles (indeed all of the claims to validity that underlie the use of language itself).[40] Indeed, on Habermas's view of the development of individualized *identity* (individuation), the person (child) internalizes the social meanings and normative structures of the surrounding, usually parental, voices.[41] The dialogic interaction with a "generalized other" takes the place of the assumption of a disembodied and objective viewpoint of Enlightenment (that is, purely Kantian) thinking. Intersubjective validity replaces depersonalized objectivity, and such intersubjectivity is established by ongoing, linguistically mediated social interaction with surrounding others. Normative (moral) validity is fixed in reference to a principle whereby all affected by a decision could freely accept the consequences of its general observance given their needs and interests.[42] Individuation occurs with the development of capacities of questioning, reflection, and critique as a component of the participation in dialogue and internalization of social meanings that such a test of validity requires.[43] Hence, as a view of *personal* development, this model mirrors the requirements

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

of legitimacy that liberal theories require of *social* principles. That is, the ability to reflectively negotiate collectively generated norms and to present critical points of view in the dynamic of such deliberations is central to the legitimacy of the political principles expressive of those norms. The relevance of this comparison will arise presently when we discuss the requirements of self-understanding in the social negotiations so described.

While the social contract tradition has expressed the establishment of the collective endorsement of political principles as a *hypothetical* agreement among rational parties, recent developments in liberal theory have underscored the need for *actual* social interaction and ongoing negotiation to be seen as constitutive of political legitimacy.[44] The traditional view of a hypothetical and philosophically determined ground for agreement has been rightly challenged by those who insist that these abstract conceptions of social and individual life utilized in such hypothetical models very like betray actual biases and exclusionary tendencies inherent in the contemporary social milieu out of which they arise (valorizing certain middle-class, white, male value-conceptions to the relative denigration of other, marginalized groups).[45] Even standard liberal theorists have claimed the centrality of democratic deliberation in the determination of the principles of justice, at least in their final form.[46]

Therefore, insofar as actual public deliberation and communication must occur for the principles of liberal justice to be settled upon and political legitimacy to be established, *self*-expression will be crucial in the functions of the citizen acting in this process. The ability to settle upon and give, publically, *reasons* for claims will function as an ineluctable element in the determination of just principles. Final determination for the order of values that will be represented in the principles of a just society must be given to citizens themselves, and such values must be defined by way of ongoing, open, discussion among autonomous citizens (and/or their representatives) in a diachronic process of refining justice and maintaining legitimacy. Thus, person's themselves must be in a position to reflect upon, and report in public settings, the value commitments that they wish to receive weight in such political deliberations.

But there are importantly different positions one can occupy with regard to the expression to others of one's own experiences, ideas, and preferences. In cases where the person speaks for herself in expressing her beliefs, desires, values and experiences, there are two ways that this first-person representational "authority" can be understood. In the first way, we view the person as the *epistemic* authority on what she is representing:

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

she speaks for herself because she knows herself (perhaps more than others or perhaps absolutely). In the second way, she has *personal* authority where she is designated as the expressive voice of those value commitments, independent of her actual ongoing hold on the content of those expressions. The point of this distinction is that one can enjoy the second type of self-representational authority without claiming the first. I might be assigned a role of expressing some material for reasons independent of my epistemic position in regards to it.

Epistemic authority over self-expressions is grounded in the assumption of a "truth of the matter" regarding the content of what is to be represented. That is, granting someone expressive authority on epistemic grounds makes sense only if (1) there is a settled truth concerning which representation is appointed or accepted, and (2) the representing agent is in the best epistemic position to know that content.[47] However, in cases where these two provisions fail to hold, the case for granting representative authority on epistemic grounds becomes weaker. And as we will see later, the assumptions about the kind of social deliberation involved in processes of liberal legitimacy cannot presuppose that there is a "fact of the matter" concerning value-statements independent of the person's own internal grasp and endorsement of that value; being in the best position to know what is "true" is less important than being in a position to adopt *for oneself* that to which one is committed.

Now let us return to the conception of the *autonomous* person to which liberal theory relies so that we can connect our endorsement of self-reflection as a component of autonomy while acknowledging the difficulties in its operation outlined earlier.

IV Autonomy and Self-Reflection Revisited

In contexts where interaction with others brings about collective decisions, the normative anchor that such outcomes provide for the participants depends heavily on the acknowledgment of the autonomy of one's co-deliberators. But autonomy in what sense? The model I have suggested here requires that the autonomous person exhibit minimal cognitive competence and hypothetical self-endorsement (interpreted as non-alienation) via self-reflection. That is, authenticity obtains when, were one to turn a reflective eye toward the motives, values, and concepts that structure one's judgments (and do so in a piecemeal manner), one would not feel deep self-alienation, self-repudiation, and unresolvable conflict.

An important point to note here is that the hypothetical self-reflection involved in this test for authenticity does not imply accurate self-knowledge or self-transparency. The test is purely subjective in that it takes as its perspectival orientation the agent's own point of view, independent of any external account of the motives, values, and beliefs to which she might turn her attention. Moreover, the non-alienation characteristic of the autonomous person has both phenomenological and affective elements: the agent would not feel a sense of self-repudiation in the internal grasping of her sense of the motives and impulses that move her to action.[48]

But in what way can this model of the liberal (autonomous) person square with the accepted levels of self-*mis*understanding that we outlined above? To see if it does, we need first to recall the distinction between two kinds of self-expressive authority – epistemic and normative authority: the lack of self-understanding we accepted only touches the assumption of (some kind of) epistemic authority on the part of self-representing agents. Insofar as the reasons for granting self-representational authority in collective decisions are normative rather than epistemic, then failures of epistemic access to the content of one's expressions – one's self-knowledge in this case – will be less serious. (Though they will by no means be irrelevant: see later discussion.)

Second, we need to focus on the distinction between the two kinds of liberalism noted earlier: In the Hobbesean variant, the point of granting individual rights of self-expression and participation in the process of legitimating state power is that such expression functions as a conduit for the promotion of the rational interests of the parties. State power, remember, is justified as a coordination device for the maximal satisfaction of such interests. Therefore, the authority granted to citizens to express their own judgments is clearly *epistemic* authority: it is the authority to judge and express their own interests, interests that are well defined independently of the process of subjective grasping and endorsement. That is, according to Hobbesean contract theory, state power is designed to protect the *idealized* desires of the participants. Their own judgments of what those desires are may well, for the reasons outlined in our examination of self-knowledge, be systematically distorted.

So for Hobbesean liberalism, full self-knowledge (as a condition of the autonomy assumed in citizens) is a necessary condition for the validity of outcomes of collective choice. Only when actual interests are expressed in deliberation will the process of aggregating such interests – which, on the Hobbesean model, is the fundamental role of the state – operate

correctly. And the interests in question are determinable independent (in principle) of the person's judgment about them: a person can be mistaken about her interests (as well as her motives, psychological states, and the like, as we saw earlier).

However, one need not accept the Hobbesean account of political authority. While I cannot argue for this here, there is good reason to avoid seeing the legitimacy of political institutions as fundamentally the coordination of individually determined interests of the citizens living under it.[49] The assumption of individualized self-interest, to take one example of a core assumption of the Hobbesean view, is highly problematic; moreover, it is not at all clear that even if citizens were self-interested and motivated by individualize desires, social stability of the sort promised by Hobbesean political theory would materialized or be maintained.[50]

On the other hand, the Kantian variant of liberalism grants individual's powers of self-government for a different reason – only when political principles are embraced authentically by those governed by them are they valid for those people. Collective deliberation in order to legitimate state power and to generate new legislation functions by way of mutual respect and what I called "empathic respect" for others' differences. For the later Rawls, for example (who departs from the literal Kantianism of his earlier view but remains in the category I am here labeling "Kantian"), the overlapping consensus that legitimates principles of justice must be "affirmed" from within each citizen's comprehensive moral view, and hence must involve a moral commitment to cooperative interaction with others whose views differ. For Habermas, valid (political) claims presuppose sincere and free interchange among participants, all of whom implicitly accept the normative presuppositions of discourse itself. Sincerity involves not simply reporting what is in fact true but expressing what one deeply believes. One can be sincere but incorrect, and it is sincerity that is presupposed in discursive interchange. Therefore, *personal* (self-representational) authority (in my sense) is what is granted in communicative action.

Thus, no presumption of epistemic authority over a person's motives and desires must be granted in this matrix; representational authority is all that is needed to ground the mutual respect (and empathic understanding) that, I have argued, functions to legitimate state power. It is as if we say to each other: you may be often mistaken about what truly moves you and what is in your best interest, but nevertheless you always get to speak for yourself on such matters. The reason for this position is moral/political, not epistemic: in order to ensure the personal

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

endorsement necessary for the validity of value commitments one must embrace and express for oneself such commitments; externally determined validity (a "fact of the matter" fixed independent of such endorsement) of values is not recognized.

Democratic institutions that arise as part of (and, some would argue, a *constitutive* part of) principles of justice require that citizens (perhaps through their representatives) be in a position to advance *reasons* for the interests they wish to see promoted collectively in their society. Democratic deliberation, then, also requires participants' abilities to reflectively endorse, indeed publically defend, the points of view, values, interests, and opinions that are the inputs to such deliberative processes (the "outputs" of which are social principles and policies). This provides further reason for the presupposition that the autonomous person is able to reflectively grasp and present her values and perspective. This accords her the kind of representational authority over those points of view but also necessitates their capacity to reflect on their values as part of the dynamic of social interchange that produces collectively justified principles. So autonomy as competent, self-reflective endorsement (non-alienation) is central to this understanding of justice and politics.

Therefore, for reasons of social legitimation, interpreted as the liberal principle of legitimacy for political institutions and principles, self-reflection is a crucial mark of the autonomous citizen whose status is respected and whose interests are protected in just political arrangements. Only if a person is put in a position to *speak for herself*, can the collectively generated principles of justice claim the legitimacy required by liberal theory. Advancing her interests in a way that thoroughly bypasses reflective endorsement of them threatens to violate the requirement that values promoted in a society obtain validity only by being subject to the citizens' endorsement of them. So liberal legitimacy presupposes a model of the (autonomous) person able to reflectively endorse her interests, respect for which is reflected in the structure of the principles themselves.[51]

What, then, should be the standards of self-understanding and cognitive competence that autonomy, used in this context, requires? To answer this question, we must say a bit more about the epistemic standards of public reason, within which autonomous self-expression plays such a crucial role. This is a complex subject, so, in addition to what has already been said, we will be brief.[52] First, in order for public justification to proceed in a way consistent with the endorsement constraint, we must assume at least a modest internalism as our epistemic standard of justification at the individual level. That is to say, no value claim can be said to be valid

for a person (or no belief about such a claim or its components) unless there is an inferential relation between such a claim and other elements of that person's belief/value corpus. Pure externalism would deny this and claim that some beliefs are justified for a person wholly independent (in principle) of that person's belief set. But the endorsement constraint implies that, ideally at least, a person could come to embrace (or at least not be deeply alienated from) the value in question. This is not possible unless there is a hermeneutic or otherwise inferential relation between that value and things the person already holds.[53]

Second, a person must have a level of understanding of her own psyche so that she is a relatively *consistent* representative of a viewpoint. If manifest inconsistencies arise from or are involved already in her corpus of desires and values, then the process of deliberation and negotiation cannot fruitfully proceed. So absence of manifest inconsistencies – where fully contradictory beliefs or values are held in ways that could bring them easily to mind – is a necessary part of autonomy competence. But this is compatible, it must be stressed, with sincere ambivalence and measured changes of mind. I meet this requirement even if I am torn in two directions on an issue or if I alter my view in light of new information and deliberation itself. But a person who is notably pulled by inconsistent desires in ways she does not admit – acting on or expressing one at one moment and doing the opposite the next – is not a competent deliberator and hence not autonomous in the requisite sense.

Third, mis-identification of motives as specific as those described in the various attribution errors described earlier need not disturb the self-expressive authority assumed in autonomy-based liberalism. One need not correctly identify the motivating reasons for action or decisions, as long as one takes responsibility for such decisions once they are made. As for mis-labeling of either the character or the source of our emotions and attitudes, public deliberation need not be seen as a process of *discovery* of stable and independently existing attitudes that such deliberation serves merely to coordinate or (as appropriate) aggregate. At least under what I am calling the Kantian rubric, public discourse is itself a process of moral importance not reducible to its revelatory role in uncovering nascent internal states of the agent. We are not merely counting votes. So when people's interaction in public debate functions in ways that "distort" their reporting of their own attitudes, their public stance in that debate thereby becomes the position they are committed to, independent of its representational accuracy concerning the internal states of the person.

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

A matter of some importance here is the manner in which commitments to beliefs or value claims can be made valid upon the decision to commit oneself to them. This is akin to the existentialist point that our existence (and our choices) precedes our essence, and our commitments follow in part from our choices themselves, thereby constituting our being. However, we should add (hence deviating from Sartrean doctrine) that the validity of a norm for a person need not be understood as wholly subjective. As Charles Taylor has argued, a fundamental aspect of human agency is a commitment to "strong valuations" – value judgments whose validity lies, in part at least, beyond the merely subjective choice to accept them. Moreover, the process of public reason itself demands the giving of reasons to others that are (1) sincere (so held as valid by the person making the claim), and (2) grounded in considerations that could appeal to those others, hence not wholly grounded in subjective choice.

But the endorsement constraint continues to operate here. For it implies that subjective embrace of a value is a necessary component of its validity (for a person). So a person's *act* of embracing a view, or embracing a view as part of a process of publically expressing it in the dynamics of public deliberation, makes it her *own* in this crucial sense. Even if I am somewhat out of touch with my motives, or systematically mistaken about the psychological sources of my opinions and values, I commit myself to them as I advance them to others in public discourse. I therefore, *construct* myself (in part) by committing myself to this or that belief. At least I construct and commit myself provisionally in that I am open to reasons from others and, as a sincere and non-strategic communicator, I listen to others in ways that may lead me to reconsider my own views. But as a participant in this process, I commit myself to views I judge to be right by expressing them, not (or not always) by simply discovering them as a settled aspect of my nexus of other beliefs, desires, and values.

In this way, the fact that reflective self-appraisal tends to undercut the person's own commitments (or merely serve as a rationalization of some of them) becomes less troubling: the public stand one takes in discourse and deliberation becomes the position to which one is held responsible in the process of generating valid social norms and the legitimacy they enjoy. It is hoped (not entirely without reason) that the *process* of public interchange itself can induce dynamic reconsideration of one's own position on various matters that will reduce whatever disconnect there might exist between a public report and a private drive.

Hence, psychoanalytic, and indeed post-modern, pictures of the fragmented and decentered self do not conflict with this picture of liberal

autonomy insofar as the requirements of self-understanding in the model of autonomy at work make no demands of strict internal unity, stable emotional or attitudinal matrices, opaque psychological mechanisms, and the like. What it demands is that the person's characteristics (values, desires, and the like) be subject to her own reflective appraisal and, if not found to be deeply repugnant, presented publicly as a position for which she is held responsible. If the very act of discursive interchange in effect constructs the value position that is the focus of this responsibility, that is consistent with the anti-perfectionism implicit in the version of liberalism alluded to here: under this view, there is no pre-determined value scheme that lies outside of human embrace and construction waiting to be found.

V Summing Up

Several objectives were pursued in this chapter. One was to claim that conceptions of autonomy should not rest on a single conception of the "self," since conceptualization of selves are (validly) understood to be multiple and variable. Second, a model of autonomy was put forward and (in part) defended, though problems with a central element of that model (the requirement of self-reflection) were aired and expanded. But we came back to the view that autonomy requires self-reflection because of the role that the concept of autonomy plays in certain political principles prominent in current theoretical constructions.

These constructions found no independent defense here, of course, and those who reject them in whole or in part will not be particularly satisfied with the chapter's conclusions. But these constructions were sketched in broad enough form (breadth that carried with it that degree of vagueness and imprecision) that they should seem compelling to many, if only because they are intended to represent a large current in modern(ist) approaches to political legitimacy and justice. To show that autonomy in something like the form defended here is necessary for the acceptability of those broadly construed theoretical constructions is no mean accomplishment, fragile though it is.

Notes

1. See, for example, Michael Sandel *Liberalism and the Limits of Justice*, 2nd ed. (Cambridge: Cambridge University Press, 1998); Iris Young, *Justice and the Politics of Difference* (Princeton, NJ: Princeton University Press, 1990); and Daniel Bell, *Communitarianism and its Critics*. Oxford: Clarendon, 1993.

P1: JtR
052183951 3c14.xml CY500B/Christman 0 521 83951 3 August 24, 2004 13:45

2. For a survey of literature on autonomy, see John Christman, "Constructing the Inner Citadel: Recent Work on the Concept of Autonomy" in *Ethics* vol. 99 no. 1 (Fall, 1988), 109–24. See also Catriona Mackenzie and Natalie Stoljar, "Introduction: Autonomy Reconfigured" in Mackenzie and Stoljar, eds., *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self* (New York: Oxford University Press, 2000), pp. 3–31; and the essays in *The Inner Citadel: Essays on Individual Autonomy* (New York: Oxford University Press, 1989). Other discussions of note of the concept include Lawrence Haworth, *Autonomy: An Essay in Philosophical Psychology and Ethics* (New Haven: Yale University Press, 1986); Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1990); Alfred Mele, *Autonomous Agents* (New York: Oxford University Press, 1995); Diana T. Meyers, *Self, Society and Personal Choice* (New York: Columbia University Press, 1989); and Bernard Berofsky, *Liberation from Self* (Cambridge: Cambridge University Press).
3. Isaiah Berlin, "Two Concepts of Liberty" in *Four Essays On Liberty* (Oxford: Oxford University Press, 1969), 118–172.
4. See references in note 1 as well as Charles Taylor, *The Ethics of Authenticity* (Cambridge, MA: Harvard University Press, 1991), 33ff.
5. Some, however, continue to insist on a close connection between autonomy and the self: see Marina Oshana (Chapter 4 in the present volume). For a focused argument for the separation between self and autonomy, see Bernard Berofsky, *Liberation from Self*.
6. Typically, the focus of models of autonomy are specifically the agent's desires. However, there is good reason to broaden this to include any aspects of the person relevant to identity, action, and choice. One can lack autonomy relative to emotions, skills, physical factors, knowledge, and general states of being as well as to desires *per se*. For discussion, see my "Liberalism, Autonomy, and Self-Transformation," *Social Theory and Practice* 27, 2 (2001). See also Richard Double, "Two Types of Autonomy Accounts," *Canadian Journal of Philosophy* 22, no. 1 (March, 1992), p. 66.
7. I defend this view also in "Liberalism, Autonomy, and Self-Transformation."
8. In my original formulation of this idea, I claimed that the person must reflect upon and accept the processes of self development himself. I now see the limitations of this formulation, and have amended this requirement as indicated in the text. See "Autonomy and Personal History," *Canadian Journal of Philosophy* 21 no. 1 (March, 1991), 1–24. For criticism of this version of the view, see Alfred Mele, "History and Personal Autonomy," *Canadian Journal of Philosophy* 23 (1993), 271–80; and for a reply, see "Defending Personal Autonomy: A Reply to Professor Mele," *Canadian Journal of Philosophy* 23 (1993), 281–90. For a discussion of historical views of autonomy, see Alfred Mele, *Autonomous Agents: From Self-Control to Autonomy* (New York: Oxford University Press, 1995). For a historical account of moral responsibility, see John Martin Fisher and Mark Ravizza, *Moral Responsibility and Control* (Cambridge: Cambridge University Press, 1998).
9. The concept of self-alienation is analyzed in different form in certain areas of psychoanalytic theory: see, for example, Karen Horney, *Our Inner Conflicts: A Constructive Theory of Neurosis* (New York: Norton, 1945). For a discussion of

autonomy and personal integrity, see Diana T. Meyers, *Self, Society, and Personal Choice*, 59–75. A parallel idea plays a role in Ronald Dworkin's distinction between those aspects of a person's personality for which she should be held responsible (for the purposes of distributive justice) and those that are part of her "circumstances" and hence subject to egalitarian redistribution. See *Sovereign Virtue: The Theory and Practice of Equality* (Cambridge, MA: Harvard University Press, 2000), 286–91, 322–23.

10. Gerald Dworkin's account of autonomy suffers from this (relatively minor) weakness, I think, in that all his model requires is the "capacity to raise the question of whether I will identify or reject the reasons for which I now act" (*The Theory and Practice of Autonomy*, p. 15).
11. This does not imply, as some liberal theorists have been (rightly in some cases) accused as claiming, that a person must be able to alter all aspects of her values and convictions upon reflection. The requirement is that she must be able to shed only those traits or commitments from which she feels deeply alienated. For discussion, see my "Liberalism, Autonomy, and Self-Transformation."
12. See, for example, Harry Frankfurt "Identity and Wholeheartedness" in *The Importance of What we Care About* (Cambridge: Cambridge University Press), 159–76.
13. Cf. Meyers, *Self, Society, and Personal Choice*, 72.
14. For further discussion, see my "Autonomy and Personal History".
15. For further discussion and elaboration of this condition, see my "Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves," *Philosophical Studies* 117 (2004): 143–64.
16. Discussion of this issue can be found in Diana T. Meyers, *Subjection and Subjectivity: Psychoanalytic Feminism and Moral Philosophy* (New York: Routledge, 1994) as well as *Self, Society, and Personal Choice*, pp. 28ff.
17. See, for example, Bernard Williams, "Persons, Character, and Morality" in *Moral Luck* (Cambridge: Cambridge University Press, 1983), 1–19, and Robert Bellah, et. al., *Habits of the Heart* (New York: Harper & Row, 1985), See also "Liberalism, Autonomy, and Self-Transformation."
18. A similar argument is made by Bernard Williams concerning moral principles, see "Styles of Ethical Theory" in *Ethics and the Limits of Philosophy* (Cambridge: Cambridge University Press, 1985), 71–92.
19. These kinds of distortions are merely more specific instances of the kind of disconnect that critics have noted about the requirement of second-order reflection on first-order aspects of the self for autonomy. See Marilyn Friedman, "Autonomy and the Split-Level Self," *Southern Journal of Philosophy*. vol. 24 no. 1 (1986): 19–35, and Irving Thalberg, "Hierarchical Analyses of Unfree Action," reprinted in *The Inner Citadel*, 123–136.
20. For an overview, see Morris Eagle, "Psychoanalytic Conceptions of the Self" in Jane Strauss and George Goethals, eds., *The Self: Interdisciplinary Approaches* (New York: Springer-Verlag, 1991), 49–65.
21. For discussion of recent social psychological work on the self that reflects this tradition, see Roy Baumeister, "The Self," in *Handbook of Social Psychology*, Daniel T. Gilbert, Susan T. Fiske and Gardner Lindzey, eds., vol. I (Boston,

MA: McGraw-Hill, 1998), 680–740. For discussion of the historical development of theories of the self, see Susan Harter, "Historical Roots of Contemporary Issues Involving Self-Concept," in Bruce A. Bracken, ed., *Handbook of Self-Concept: Developmental, Social, and Clinical Considerations* (New York: John Wiley & Sons, Inc., 1996), 1–38, and Kenneth J. Gergen *The Concept of Self* (New York: Holt, Rinehart, and Winston, 1971), 1–12.

22. See Baumeister, 690f for overview and discussion of these observations.
23. See Richard Nisbett and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, NJ: Prentice Hall, 1980), 120ff. For a discussion of the relation between such errors and moral philosophy, see Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 1998–99 (1999) 315–331.
24. See Daryl J. Bem, "Self-Perception Theory" in L. Berkowitz, ed., *Advances in Experimental Social Psychology*, vol. 6 (New York: Academic Press, 1972). For discussion, see Ross and Nisbett, *Human Inference*, 195–227.
25. This is shown in experiments in which subjects are given artificial stimuli inducing certain emotions but will mid-identify both the source and the nature of that emotion (ignoring, for example, the readily apparent artificial source): See Ross and Nisbett, *Human Inference*, 199–210, for an overview.
26. See, for example, Judith Butler, *The Psychic Life of Power* (Stanford, CA: Stanford University Press, 1997. For discussion, see Diana Meyers, *Subjects and Subjectivity*.
27. This is not to say that liberalism, by definition, is anti-perfectionist. There are plenty of perfectionist liberal views around: see, for example, Joseph Raz, *The Morality of Freedom* (Oxford: Oxford University Press, 1985); Will Kymlicka, *Liberalism, Community, and Culture*, (Oxford: Clarendon, 1989); and William Galston *Liberal Purposes* (New York: Cambridge University Press, 1991). For further discussion of the contours of liberalism, see my *Social and Political Philosophy: A Contemporary Introduction* (New York: Routledge, 2003), chapter 4.
28. This is meant to express the principle of liberal legitimacy: see Rawls, *Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001), 41.
29. See Rawls, *Political Liberalism*, 25–35, for discussion.
30. This formulation is meant to be neutral about the fundamental *grounds* for this respect, leaving open the possibility that such ground is ultimately "political" rather than metaphysical. For discussion, see Rawls, *Justice as Fairness: A Restatement*; Charles Larmore, *Patterns of Moral Complexity*. (Cambridge: Cambridge University Press, 1987); and John Gray, *Post-Liberalism: Studies in Political Thought*. New York: Routledge, 1993).
31. Kymlicka, *Liberalism, Community and Culture*, 10–12. Kymlicka sees this constraint on the view of *value* that is assumed in liberal theory, since he claims that liberalism rests on this unique conception of value rather than the assumption of the "priority of the right over the good" claimed here. For discussion of this difference, see Christman, *Social and Political Philosophy*, 97. Also, Gerald Gaus (Chapter 12 in the present volume) claims that liberalism

should be defined as the tradition of political philosophy that puts ultimate value on individual liberty (conceived as a presumptive right to noninterference). I will only mention in passing here the reason that moves me in another direction – that "liberty" cannot function in this way as a basic value since it is an essentially contested and, more importantly, derivative, political value (derivative from the conception of the "right," or justice, operative in the society). For discussion of the concept of liberty, see my *The Myth of Property*, chapter 4, and Ronald Dworkin, *Sovereign Virtue* (Cambridge, MA: Harvard University Press, 2000), chapter 3.

32. See Joseph Raz, *The Morality of Freedom*; Will Kymlicka, *Liberalism, Community, and Culture*; and Ronald Dworkin, *Sovereign Virtue*.
33. See, for example, Steven Wall, *Liberalism, Perfectionism and Restraint* (New York: Cambridge University Press, 1998); Thomas Hurka, *Perfectionism* (New York: Oxford University Press, 1993); and George Sher, *Beyond Neutrality: Perfectionism and Politics* (Cambridge: Cambridge University Press, 1997).
34. A paradigm case of this approach can be found in David Gauthier, *Morals By Agreement* (Oxford: Oxford University Press, 1986), but see also Jan Narveson *The Libertarian Idea* (Philadelphia, PA: Temple University Press, 1988), chapter 14.
35. This variant is seen most clearly in Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), but it survives in the view developed in *Political Liberalism* – that is, according to "political" liberalism – where principles of justice are established via an overlapping consensus among reasonable comprehensive moral views – citizens are able to "affirm" the principles from "within their own comprehensive views" (*Political Liberalism*, Lecture IV) – that is *morally*. To do otherwise is to adopt the view that justice is a mere *modus vivendi*. Also making much use of the kind of distinction described in the text (or at least one parallel to it) is Habermas, especially in the distinction be makes between "strategic" and "communicative" social interaction. See *Moral Consciousness and Communicative Action* (Cambridge, MA: MIT Press, 1991), 58. For a similar distinction in approaches to political justification, see Gerald Gaus, "Liberalism," *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu), p. 5. See also Michael Sandel, *Liberalism and the Limits of Justice*, 1–7.
36. This is an explication of a Kantian conception of the grounds of justice, utilizing the views of several writers in this tradition, most notably Rawls and Habermas (about whom more will be said later). But the call for including a sense of *empathic* respect is motivated by the arguments of Susan Moller Okin (see *Justice, Gender and the Family* (New York: Basic Books, 1989), 187).
37. This is compatible, it should be repeated, with perfectionist brands of liberal thought, as long as such perfectionism retains this "endorsement constraint" and admits of a pluralism of (allegedly objective) values.
38. *Political Liberalism*, 35ff.
39. This point is stressed in Kant (influenced, no doubt, by Rousseau): see "On the Common Saying that It May Be True in Theory but Not in Practice," in *Practical Philosophy*, Mary Gregor, trans. (Cambridge: Cambridge University

Press, 1996), 273–310. See also Jeremy Waldron, *The Dignity of Legislation* (Cambridge: Cambridge University Press, 1997), 47–52 (see especially p. 52, n. 43). This point, and its importance, is overlooked in much recent liberal theory: see, for example, Gerald Gaus, "Liberalism" *Stanford Encyclopedia of Philosophy*.

40. *The Theory of Communicative Action*, vols. I and II, Thomas McCarthy trans. (Boston, MA: Beacon Press, 1984, 1987). See also *Moral Consciousness and Communicative Action*, pp. 43–194.
41. Habermas, "Moral Development and Ego Identity," in *Communication and the Evolution of Society*, Thomas McCarthy, trans. (Boston, MA: Beacon Press, 1979), 69–94.
42. See "Moral Consciousness and Communicative Action," 120–22, and *Between Facts and Norms*, 107. In the truncated version here, I combine what Habermas calls a rule of "argumentation" (principle "D") with the fundamental principle of morality he labels "U."
43. This view of normativity and personal individuation is controversial, and certainly much more complex than this. For critical discussion, see for example, Seyla Benhabib, "The General and Concrete Other," in Eva Feder Kittay and Diana T. Meyers, eds., *Women and Moral Theory* (Totowah, NJ: Rowman and Littlefield, 1987) 154–77; and Allison Weir, "Toward A Model of Self-Identity: Habermas and Kristeva," in *Feminists Read Habermas* (New York: Routledge, 1995), 263–82.
44. To say that political principles will be "fleshed out" is to align oneself with Rawlsian political liberalism, understood a certain way, where the justification of principles is hypothetical (even via the use of public reason): these principles are justified if an overlapping consensus involving them could be established. For other theorists, actual social deliberation and democratic communication *constitutes* the justification of principles. See, for example, Habermas, *Between Facts and Norms*.
45. For arguments along these lines, see Iris Marion Young, *Justice and the Politics of Difference* (Princeton, NJ: Princeton University Press, 1991); Nancy Fraser, *Justice Interruptus* (New York: Routledge, 1997); and Jürgen Habermas, *The Inclusion of the Other* (Cambridge, MA: MIT Press, 1998). Even Rawls eventually claimed that the dynamics of public reason – real world, ongoing, interaction among persons and groups provides the ultimate anchor for the overlapping consensus on which justice is grounded: see "The Idea of Public Reason Revisited," in *The Law of Peoples* (Cambridge, MA: Harvard University Press, 1999), 129–180.
46. See, for example, Jeremy Waldron, *The Dignity of Legislation*.
47. For discussion of epistemic authority over one's own desires and motives, see Gerald Gaus, *Justificatory Liberalism* (New York: Oxford University Press, 1996), Part I.
48. It is in this way that autonomy can be seen to involve a level of self-*trust*, as has been pointed out by several writers. See, for example, Paul Benson, "Free Agency and Self Worth," *Journal of Philosophy* 91 (1994), 650–68; Trudy Grovier, "Self-Trust, Autonomy, and Self Esteem," *Hypatia* 8 (Winter, 1993), 99–120; and Carolyn McLeod and Susan Sherwin, "Relational Autonomy,

Self-Trust, and Health Care for Patients Who are Oppressed," in Mackenzie and Stoljar, eds., *Relational Autonomy*, 259–79.

49. For a defense of the Hobbesean approach, as I am using that label, see Gauthier, *Morals By Agreement*. For an argument that purely instrumental rationality (on which the Hobbesean model is predicated) cannot adequately account for social stability and political authority, see Jon Elster, *The Cement of Society* (New York: Cambridge University Press, 1989), chapter 3, and *Solomonic Judgments* (New York: Cambridge University Press, 1989), chapters 1 and 4. For criticism of a different type, which strikes at the heart of the Hobbesean framework, see Donald Green and Ian Shapiro, *The Pathologies of Rational Choice Theory* (New Haven, CT: Yale University Press, 1994). For general discussion of this issue, see Habermas, *Theory of Communicative Action, Vol. II*, 119–52.
50. For a specific argument of this sort, see Thomas Christiano, "The Incoherence of Hobbesian Justifications of the State," *American Philosophical Quarterly* 31 (1994), 23–38. For a general discussion, see my *Social and Political Philosophy: A Contemporary Introduction*, chapter 2.
51. Though arrived at from a different direction, the claims being defended here involving the relation between autonomy and a persons being designated as speaking for herself resemble closely Paul Benson's views: see his Chapter 5 in the prevent volume.
52. And I rely greatly on the detailed and powerful analysis of "public justification" and its role in political legitimacy developed by Gerald Gaus in *Justificatory Liberalism*.
53. Assuming some qualified internalism for the purposes of political philosophy is not the same as claiming this as the best epistemic account, period. However, for an argument against strict externalism as an epistemic standard, see John Pollock, *Contemporary Theories of Knowledge* (London: Hutchinson, 1987), 133–49. Also, what is meant by "hermeneutic" here is that a coherent interpretation could be applied to the belief (or value set) that includes the contested element.