

Positive Freedom and the General Will

Piper L. Bringham and Gerald Gaus

1 CONCEPTS AND CONTEXTS OF FREEDOM

Among the first essays one reads as a student of political philosophy is apt to be Isaiah Berlin's (1969) classic, "Two Concepts of Liberty." Despite its elegance and rhetorical power, the effects of this paper have been somewhat unfortunate: Berlin's thesis that negative and positive liberty are competing conceptions, and that a fundamental task of a political theory is to defend *a* conception of liberty against competitors, continues to influence theorists of freedom. Berlin famously defends negative liberty as a truer and more humane ideal than positive liberty, as it recognizes that "human goals are many," and no one can make a choice that is right for all people (Ibid., p. 171).¹ Not surprisingly, this claim elicited counter-criticisms that negative liberty was the truly flawed ideal (see, e.g., Pettit 1997, esp. chaps. 1 & 2; Taylor 1979). Behind this tendency to defend a specific concept of freedom as foundational, or generally correct, is an understandable philosophic impulse for a general theory of freedom, which would provide a unified and elegant understanding of free action and the various senses in which a person might be said to be free.² Even Gerald MacCallum (1972), who sought to transcend the debate between negative and positive liberty, endeavored to do so by providing a single

* Our thanks to the participants in the 2014 Tucson workshop on freedom for their questions and comments; our special thanks to Dave Schmitz, Carmen Pavel, and two anonymous readers.

coherent scheme that encompassed both.

An alternative understanding of a “theory of freedom” is one that, within a coherent normative framework, distinguishes different contexts, and tries to show why, say, the “negative conception” is especially important in contexts of interpersonal claims about freedom, while positive accounts make much more sense when talking about free agency, or what it means to lead a free life.³ Such an approach is not simply eclectic, drawing on different conceptions willy-nilly, but seeks to understand the complexity of our commitment to freedom, and how in some contexts a person can be free in one way and not in another; the aim is to show how a reasonable view of humans and their relations can make room for, and explain, these complexities and tensions. This, indeed, is in the spirit of T. H. Green, who explicitly introduced the idea of positive liberty into our political discourse. Green did not seek to supplant negative liberty as a false or “atomistic” doctrine; rather he acknowledged that “it must be of course admitted that every usage of the term [i.e. freedom] to express anything but a social and political relation of one man to other involves a metaphor It always implies . . . some exemption from compulsion by another” (1986, p. 229). While acknowledging this, Green explored an extension of this usage employing the positive conception — one that centers on the idea of free agency and decision-making, rather than the quintessential question of freedom of action in relation to interference by others.

The aim of this chapter, then, is not to defend the concept of positive liberty

“against” negative or republican liberty. Nor do we present a comprehensive account of freedom that relates different contexts, and the different understandings of freedom appropriate to each. The aim here is modest, examining but one context of freedom: relations of moral responsibility. We seek to show how one understanding of positive liberty — what we call *freedom as reasoned control* — is presupposed by our relations of moral responsibility. We argue that what may seem simply Rousseau’s quixotic goal — of insuring that all subjects of the moral law remain morally free — is necessary to the maintenance of responsibility relations within a moral community. Unless all are free in the sense of exercising reasoned control in accepting moral demands — i.e., their acceptance of these demands expresses their status as reasoning persons — they cannot be held responsible for their failure to comply. This then leads us to the second concern of the chapter: whether the concept of the general will can reconcile positive freedom and moral responsibility with regulation by a common moral law. In section 3 we briefly look at two classic accounts of the general will — those of Rousseau and Bosanquet — while section 4 turns to contemporary proposals that seek to understand how a general will might arise in a diverse society.

2 FREE REASONED MORAL AGENCY

2.1 *The Practice of Responsibility and Reasoned Control*

As Peter Strawson (1962) famously showed us, our moral practices are inescapably

about our reactions to what we perceive to be the good or ill-will of those with whom we interact. We make demands on them, and they on us; and we hold them (and ourselves) responsible for failure to meet these demands. The reactive attitudes are fundamental to these relations of responsibility; we experience resentment at those whose failure to meet our demands manifests an ill-will toward us, and indignation when, as a third party, we view others as the objects of such will. As Strawson stressed, we do not really have the option of deciding whether or not we should care about the attitudes of others toward us in these practices. "The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them. . . ." (Ibid., p. 197). Because we are so deeply committed to this interpersonal responsibility perspective, when we reflect on our practices we must be concerned about the conditions that are necessary to sustain this perspective — the conditions that make oneself and others fit to be held responsible. "The fact that fitness to be held responsible becomes salient to anyone involved in that practice, and that it represents what we think of as freedom in the agent," Philip Pettit observes, "means that the concept of free agency is intimately woven into the tapestry of inescapable human sentiments and responses" (2001, p. 20).

Following Strawson, our social morality can be understood as such a practice of

responsibility (1962, pp. 199ff; see also, Strawson 1961). Thus understood, our interpersonal morality is not simply about objective judgments of the rightness or wrongness of the actions of others (and ourselves): it is a system embedded in our attitudes toward others, and our judgments of their intentions and attitudes toward us. When I hold another responsible I do not simply judge his action against some standard; I react to his ill-will, his lack of respect or consideration. We deeply care about maintaining this web of personal relations. If so, we must care about the conditions that render individuals fit for moral responsibility and, as Strawson, Pettit and many others have stressed, a fundamental condition is freedom (see, e.g., Hayek 1960, chap. 5) — and not simply as regards only one “concept” of freedom. In any given responsibility context, negative freedom of a certain sort is certainly necessary for a person to be fit for responsibility. If one is threatened and coerced to perform a wrongful act such that, we say, one had no choice, it will generally not be an action for which one is responsible (Strawson 1962, pp. 189-90). The act is not a manifestation of one’s will, but the ill-will of the coercer (see further, Hayek 1960, chap. 9).

A type of positive freedom is also necessary, for we must see the person to be held responsible as, in some sense, in reasoned control of her actions. A clear example of a person lacking such control is one in the grips of an obsessional neurosis. Consider the case of a nineteen-year-old girl with obsessional sleep ceremonies, which needed to be performed nightly. Clocks had to be stopped or

removed from the room, including her small wristwatch; flower pots and vases needed to be collected and put on a writing table; the door between her room and that of her parents must be half-open; “[t]he pillow at the top end of the bed must not touch the wooden back of the bedstead. . . . The eiderdown. . . had to be shaken before being laid on the bed so that its bottom became very thick; afterwards, however, she never failed to even out this accumulation of feathers by pressing them apart” (Freud 1973, p. 305).⁴ According to Freud, those in the grips of such obsessions have difficulty adjusting their actions to their settled aims and the intrusive thoughts make it very difficult for them to carry on with their lives. “All these things combine,” Freud concluded, “to bring about an ever-increasing indecisiveness, loss of energy, and curtailment of freedom” (1973, p. 299). This type of unfreedom is very much what Green had in mind. A person who is subject to some impulse that he cannot control is, Green said, “in the condition of a bondsman who is carrying out the will of another, not his own” (1986, p. 228). For both Freud and Green, in these cases the person has lost reasoned control of his or her activity, and is in this sense unfree.

Suppose one demands from our nineteen-year-old obsessive that she conform to the demand “Give your parents’ privacy at night; let them close their door!” And let us suppose that there is a good reason for her to see this as a *bona fide* moral requirement. Because she cannot exercise reasoned control over her actions, she cannot comply with the directive, which appeals to her reason — it gives her reasons

to act. Consequently, it would seem that her failure to comply cannot be taken as an indication that she bears ill-will toward her parents because she has set aside, ignored, or otherwise chosen not to comply with the directive; the impulse subverts her ability to control her behaviour via such directives. "When we see someone in such a light as this, all our reactive attitudes tend to be profoundly modified" (Strawson 1962, p. 194).

Now consider a second case, where you address a moral directive " ϕ !" to a twenty-year-old undergraduate, but the directive to ϕ is one that you know is beyond his ken to appreciate as based on reasons. Suppose, for example, that the reasons that ϕ -ing is morally required can only be understood given a complicated argument from an exceedingly complex original position set-up, an especially demanding transcendental deduction, or an intricate formal proof in first-order predicate calculus. Here by directing the undergraduate to ϕ you are directing him to *ignore* his reasoned control, for if he reasons as well as he can or as well as we expect of him, he just cannot see why he ought to ϕ . In the case of the compulsive she may well know that she has reason to give her parents privacy, and so perhaps an especially rigorous person might still say "she really did know better than to do what she did, though she just could not help herself." She lacks the ability to control her will through reasoned deliberation (or, as Green would say, she is not controlled by her rational will). It is not unintelligible for our rigorous critic to claim that the nineteen-year-old obsessive does manifest a sort of ill-will, albeit it one she cannot

control. Imagine that her compulsive behavior was a murderous rage; then we one might quite intelligibly say that she had an ill-will that was outside of her reasoned control. In contrast, in our second case the undergraduate's actions are subject to reasoned control, but he simply cannot see how complying with that directive is consistent with such control. Here no ill-will or lack of consideration can intelligibly be inferred: given the exercise of his rational control, you admit that he simply cannot see what is wrong with not ϕ -ing. This does not imply that it is false that he ought to ϕ ; it does mean that the practice of responsibility will, again, be profoundly modified in such cases.⁵

Quintessentially, to rationally hold a person morally responsible for failing at time t to obey a directive D to ϕ one must hold (i) she has reasoned control over her actions — more specifically, at time t such that if she concluded at t that she had sufficient reason to conform to D , she could have ϕ -ed at t and (ii) under some conditions C , the exercise of her reasoning to a certain level l , would lead to the conclusion that ϕ -ing was morally required at t . Of course both these conditions are subject to diverse specifications: longstanding disputes about moral responsibility have often centered on disagreements about these specifications. Regarding (i), for example, different accounts have sharp conflicts as to what “*could* have ϕ -ed at t ” means, as well as what sort of psychic pathologies render a person incapable of such control. And regarding (ii), in specifying C , varying theories of moral responsibility greatly differ in specifying from what impairments (drugs, emotional distress,

psychological pressure, and so on) the agent must be free.⁶ And they disagree about what level of reasoning (*l*) about *D* our practice of responsibility specifies such that, if the person did engage in *l* under *C*, she would conclude that ϕ -ing was morally required at *t*. Must she be ideally rational, or simply adequately reflective?⁷ These questions have been answered in widely different ways and many of the answers may be contextual, generating different and complex accounts of moral responsibility. Debates about these specifications are, of course, terribly important, but they do not obviate the claim that some versions of (i) and (ii) must be met to sustain our practice of responsibility. For you to be fit for moral responsibility for conformity to directive *D*, you must have been able to grasp the moral reasons for it, and you must have had the capacity to act on them. These are conditions of free reasoned moral agency.

2.2 Pettit's Objection to Rational Control

Pettit provides a detailed analysis of "freedom as rational control," but ultimately rejects it as an inadequate understanding of the sort of freedom presupposed by attributions of responsibility. Think again of our twenty-year-old undergraduate; we evaluate ϕ as the moral thing to do on the basis of a difficult derivation, and direct him to do it. So, we have seen that on our account:

...rational control is quite consistent with the agent's not having any beliefs to the effect, say, that this or that is what they ought to do and that they can be rightly held responsible for

whether or not they conform to that evaluation. Thus it is quite consistent with the agent being unfit to be held responsible for what they have done. There may be no standards acknowledged or embraced by the agent and no standards to which they may be expected to answer.

...

This line of thought shows that if an agent is fit to be held responsible then not only must their beliefs and desires constitute rational control over what they do; the agent must also have evaluative beliefs to the effect that this or that is what is required of them, whether in rationality or prudence or in morality, and they must have the desires to live up to such evaluations. Only agents who are capable of recognizing and responding to standards in this way can be held responsible for what they do and can count as free or unfree (Pettit 2001, p. 40).

Our understanding of “reasoned control” is not precisely equivalent to an obvious interpretation of “rational control.” Consider unreflective Ursula, whose actions always manifest rational control insofar as she has the capacity to, and actually does, base her actions on her current beliefs and desires, but she is cognitively unable to reflect on them. If she does not presently endorse *D*, she cannot reason herself to endorsing it, because she is unable to reason to level *l* (on any account where level *l* requires any significant reflection on her current beliefs). In this case Ursula would exercise “rational” but not “reasoned” control.

In the context of fitness for moral responsibility and the reactive attitudes, reasoned control makes much more sense than mere rational control. Suppose one

more case: your neighbor currently possesses beliefs and desires such that, given them he sees no reason to refrain from dumping waste in your yard. However, if he reflected (under C to level l) on the reasons not to dump his waste in your yard he would accept the moral directive to stop, but he never bothers to think about the matter that deeply. If, because he failed to reason about his dumping and its effects, he goes ahead and dumps, this would by no means undermine your reactive attitudes toward him and his act. That he acted wrongly because he failed to take the time and effort to think things through hardly absolves him of ill-will. He possesses reasoned control, in contrast to unreflective Ursula, who simply cannot think things through. One such as the thoughtless neighbor, who cares that little about the impact of his actions on others, does not show them consideration, but rather a sort of contempt.

In the above passage Pettit stresses that rational control does not guarantee fitness for responsibility, and for him this is its core problem: he is searching for a concept of freedom F such that if a person possesses F , she is inherently fit for moral responsibility. Thus he quite rightly points out that freedom as reasoned control does not meet this condition — our undergraduate manifests freedom as reasoned control, but he is “unfit to be held responsible” for not ϕ -ing because, given the use of his reason (under C to level l), he does not “have evaluative beliefs to the effect that” ϕ -ing is required of him. However, on our account, freedom as reasoned control is a necessary, not a sufficient, condition for being fit for moral responsibility.

No account of positive liberty alone could provide sufficient conditions. Even if one possesses perfect freedom as reasoned control and even if one does hold the relevant evaluative beliefs, if one is a chattel slave subject to constant coercion and threats — one is almost entirely deprived of negative freedom — one will not be fit to be held responsible.

This leads to the fundamental difference in our analyses. Pettit seeks to develop a concept of freedom-as-control such that it is a conceptual truth that those who are free in this sense are also fit for attributions of moral responsibility. The status of being free entails the status of being a moral agent — a member of our moral community who is fit to be held responsible. Now certainly we would like to *show* that under an acceptable morality, the set of persons whose actions are controlled by their reasoning is very close to the set of moral agents fit to be held responsible, and so subject to the practice of responsibility. But this, as it were, is what we must seek to demonstrate, not by an analysis of the concept of freedom, but by an account of the nature and substance of morality. We ask: how can there be a morality such that, if not all, almost all, free (qua reason-controlled) agents are participants in the practice of responsibility?

2.3 Rousseau's "Quixotic" Quest

Recall Rousseau's fundamental problem: "*to find a form of association*" in which each, while uniting himself with all, "nevertheless obeys only himself, and remains as free

as before. Such is the fundamental problem to which the social contract furnishes the solution" (Rousseau 1988, p. 92 (Bk. I, chap. 6), emphasis supplied). Note that for Rousseau, the aim is not to find a *concept of liberty* that reconciles individual freedom with moral responsibility to others, but a form of *social life* in which these are reconciled. Many in the analytic tradition have found this quixotic or wooly-headed: how can one be subject to a common law and responsible to others for compliance and yet obey only oneself? This, it is often thought, is simply nonsense. So far from being nonsense, however, some approximate solution to Rousseau's problem is needed to achieve a community that sustains the relations that make moral responsibility possible. We might rephrase Rousseau's problem: "to find a form of moral community in which each free person is fit to be held responsible."⁸ As Rousseau was well aware, many moral communities fail to solve the problem: in many communities individuals possessing reasoned control are not fit to be held responsible, as the exercise of their reason does not lead them to endorse the oppressive or sectarian morality of their society. Like our undergraduate, they possess reasoned control (the capacity to control condition is met), but simply cannot reason themselves to the community's morality (they fail to satisfy the reasoning condition). Freedom and responsibility are thus driven apart. Rousseau appreciated that they cannot be brought together through conceptual analysis, but only through moral reform. Thus the fundamental link between positive freedom and the general will, for Rousseau believed that only a society whose rules were an expression of the

general will could unite positive freedom and responsibility.

3 THE GENERAL WILL: TWO CLASSIC ACCOUNTS

3.1 Rousseau's Limited Diversity Account

On Rousseau's view, because no person has natural authority over another, and because mere force could never legitimate authority, society must be founded upon agreement among its members (Ibid., p. 88 (Bk. I, chap. 4)). The form of society or moral community advocated by Rousseau is one of direct democracy, in which each citizen participates in government by voting for or against general laws that are presented as impartial and aimed at the common good. When citizens vote on such general laws with their common survival and well-being in mind, the outcome of the vote reveals the general will of the community.

There are well-known difficulties in interpreting Rousseau's theory of the general will (see Gaus 1997). However, on perhaps the most plausible reading, it points to a rather precarious combination of (i) a shared idea of the common good and (ii) a diversity of views about how individuals' other private interests relate to that common good. The general will is supposed to direct the community's forces in accordance with the common good, as it is "the agreement of [private interests that] has made [society] possible. It is what these different interests hold in common that forms the social bond" (Rousseau 1988, p. 98 (Bk. II, chap. 1)). Thus the general will arises out of the agreement or similarity between individual interests. Complete

unanimity or homogeneity among the interests of all members of society is not required;⁹ rather, the general will takes the *somewhat disparate* contributions of individuals and identifies through them the best collective decision about the common good. Without some measure of differences in the interests of society's members, according to Rousseau, a procedure aimed at discovering or revealing the common interest would be unnecessary: "everything would proceed on its own, and politics would cease to be an art" (Ibid., p. 100, n. 9 (Bk. II, chap. 3)). While citizens have moderately divergent private interests, these are not completely detached from the common interest, and in a well-ordered society, the benefits each individual gains from promoting the common interest outweigh any share of misfortunes that result from divergences between his private interest and what the common interest requires (Ibid., p. 149 (Bk. IV, chap. 1)). In this way, individuals reflectively endorse the laws selected by the general will as being consistent with their good. They thus freely conform to the general will because (i) they have reasoned control over their actions, and (ii) their reasoning, focusing on their own good, leads them to the conclusion that following the law is morally required.

Citizens support impartial laws aimed at promoting the common interest because considerations of their own private interest direct them to do so. However, private interest and common interest will only sufficiently align for all individuals when their situations are largely comparable. To see this, consider a society, perhaps much like our own, in which there is a vast diversity of cultures and traditions, a

variety of careers and ways of living, and significant stratification between social and economic classes. In such a society, because of the diversity of positions we occupy and circumstances in which we find ourselves, it is most dubious that laws will affect us equally, and thus, that we will be able to see them as arising from the essential identity of our individual interests. Such concerns demonstrate that, on Rousseau's account, although total homogeneity of interests among citizens is not required, *deep diversity* among citizens' interests and viewpoints jeopardizes the general will: "The more harmony reigns in the assemblies, that is, the closer opinions come to being unanimous, the more dominant, therefore, is the general will, *but long debates, dissensions, and tumult* proclaim the ascendancy of private interests and the decline of the state" (Ibid., p. 150 (Bk. IV, chap. 4), emphasis added).¹⁰ This gives us reason to be skeptical that Rousseau's account of the general will can provide a mechanism for our own diverse society by which we might unite freedom and moral responsibility. Because Rousseau's solution to our fundamental problem — to ensure that each free person is fit to be held responsible (or, to put it differently, to ensure that complying with moral demands is an exercise of positive freedom) — is a form of moral community that forbids, or simply cannot tolerate, deep diversity among citizens' viewpoints, it looks much less attractive as a solution for deeply diverse societies (see Chapman 1956, chaps. 6-7).

3.2 Bosanquet's Diversity Account

As the history of political thought is taught — at least in philosophy departments — even advanced students (and, indeed, scholars!) could be excused for thinking that Rousseau was the only theorist of the general will. In fact, other important political philosophers developed their own, very different, interpretations of the idea, including the British idealists who, it should be stressed, were especially interested in relating positive freedom to the general will.¹¹ In sharp contrast to Rousseau, Bernard Bosanquet emphasized the *need* for a deep diversity of views and ways of life as the foundation for the general will. For Bosanquet, the private wills of individuals — which are based on personal interests and moral beliefs, an individual's own knowledge and history, and his dispositions and traits — are diverse.¹² It is these diverse individual wills interacting with each other in complex ways that allow the general will to arise — here the general will is more of a correlated system of ideas than a core shared interest. Thus, for Bosanquet the division of labor in a free society expresses the general will; each person takes account of the will and abilities of others, and seeks to adjust his own will and activity to theirs. “Each unit of the social organism has to embody his relations with the whole in his own particular work and will” (Bosanquet 2011a, pp. 344-45): Private wills, on Bosanquet's view, presuppose the existence of communal life, and thus, must support and be supported by the individual wills of others in the moral community (see further, Gaus 2001). “The man's plans and principles all depend upon the support of other wills, and, apart from such agreement, there is no feature

of his life which he could possibly hope to realise" (Bosanquet 2011b, pp. 307-8). Nonetheless, each private will "stops at a certain point": no single individual grasps the entire system of moral relations — the general will (Ibid., p. 308). The complex interaction that takes place between individual members of the moral community, however, helps the individual to see possibilities for what is truly good and just, of which she could not conceive on her own. Thus on this view, a system of deep diversity gives rise to the general will, and the possibility for uniting freedom and responsibility within a diverse society appears anew.¹³

4 THE GENERAL WILL AND DIVERSITY: TWO CONTEMPORARY PROJECTS

4.1 Spontaneous Rules in a Publicly Justified Order

Although Bosanquet explores the possibility of a general will under conditions of deep diversity, his proposal relies on absolute idealism, which posited that the interrelated moral views of individuals are ultimately all part of a harmonious system of reality — the absolute (see further Gaus 1994). Thus in place of Rousseau's deliberative account, Bosanquet's general will proffers something of an invisible hand analysis. Because the general will is a system of moral relations that cannot be grasped by any single mind, it arises without design from diverse wills. A contemporary account along these lines, shorn of its idealist foundations, would require, (i) a plausible account of a system of morality as a spontaneous order, arising essentially unplanned out of the complex interactions of heterogeneous

agents and, (ii), a claim that the reasoned control of each agent endorses this emergent morality. Now the first claim certainly has been defended without resort to the absolute. Hayek, for one, has presented an evolutionary analysis of morality as a system of evolved rules to order complex and diverse societies (see Hayek 1973). That the interaction of diverse agents gives rise to moral rules to structure their interactions is by no means beyond credibility. Indeed, contemporary analyses of the emergence of social norms present detailed models of how this might occur (see, e.g., Bicchieri 2006, chap. 6; Brennan et al. 2013, Pt. II).

The pressing problem would seem to be claim (ii), that the reasoned control of each endorses this spontaneous order. Hayek himself was deeply skeptical that explicit justification of these rules is to be had (Hayek 1988, esp. chaps. 1 & 2). It is not clear, though, that such pessimism is warranted. As contemporary theorists of social morality and social norms stress, basic normative rules to structure the interactions of heterogeneous agents, especially those that help them escape social dilemmas with Prisoner's Dilemma-like structures, are critical to their successful agency; without such rules each does far worse than under universal conformity to them. As in Rousseau's account, agents, even very heterogeneous ones, surely share private interests in escaping from the traps of destructive social dilemmas. Moreover, recent evidence indicates that conformity to such rules is not only endorsed by the private interests of each individual, but by evolved commitments to fairness.¹⁴ If so, we have at least the outlines of how a contemporary Bosanquet-

inspired theory of the general will might proceed: if a system of rules and norms arises out of the interactions of diverse agents, which help them escape from social dilemmas, and if these rules correspond to their normative commitments concerning fairness, then given the interests and normative commitments of each, each may reason themselves to endorsing these shared rules of order.

To see this better, suppose we have a diverse group of agents G , who are caught in social dilemmas in which, if each acts as she thinks best, each will be worse off than if all cooperated — but if all the others cooperate, each does best by defecting. Respect for bodily integrity and property, and sustainable resource use, are a few of the critical social dilemmas all societies face. In G rule R may have evolved such that, given the diverse interests of the members of G and their diverse judgments about the fair way to resolve social dilemmas, each does better by her own lights when conforming to R than she would do if there was no rule, or widespread violation of R . Because the character of R is such that compliance furthers their interests and is consistent with their values and commitments regarding fairness, when they exercise their reason to level l members of G would conclude that they have sufficient reason to conform to R 's directives. And if this is so, then diverse agents living under R , exercising their freedom as reasoned control, will be fit for responsibility. As with Bosanquet's account, it is not required that they all, collectively, perceive this; what is critical is that all are fit to be held responsible for violations of these rules of cooperative order, which advance the interests of all.

And, should the presently evolved system not be so endorsable, we could inquire whether explicit revisions could make it so. This, of course, is not to say the project can be carried through, but it certainly does not seem a will-o'-the-wisp.

4.2 Diverse Agents Solving Moral Problems

The rules and order project sketched above is something of a compromise between Rousseau's and Bosanquet's visions of the general will. Like Rousseau, it sees the general will as focusing on rules for advancing the interests of each; like Bosanquet, it sees these rules as arising out of the diverse interactions of heterogeneous agents rather than a "shared point of view."¹⁵ Some contemporary lines of analysis indicate that we might even move closer to Bosanquet's vision of a general will in which, as heterogeneous moral agents, when one agent's insights into the general will "stop," another picks up. This idea nicely captures an interpretation of the philosophical import of recent work by Lu Hong and Scott E. Page about the necessity of diversity for solving our problems (Page 2007a) and, we propose, this includes our moral problems.

The starting point of Hong and Page's analysis is a group of cognitively diverse agents: people who look at the same problem in different ways.¹⁶ In their terms, our heterogeneous group has diverse *perspectives* — different ways of representing a given problem and solutions. Suppose we take a problem that might confront a moral community, say choosing the most fair taxation system.¹⁷ Perspectives are

diverse if the set of what is feasible is mentally represented in different ways. So, for example, suppose a group is thinking about the problem, and in the end (though not necessarily at the outset) all can recognize that they can choose from five possible tax reforms ($a-e$); their question is how well each of these proposals does in solving their problem of achieving the most fair system. Each perspective represents the underlying structure of the options in different ways. Alf's might represent the set of possible tax reforms based on the number of votes each proposal is likely to receive in the Senate, Betty's might arrange them based on their simplicity, while Charlie's perspective might organize them based on how much they are likely to cut taxes for the middle class. The way in which a person organizes the set of possible solutions determines which options are similar to others and which are very different; since a perspective determines how one structures the options, it also structures how one thinks of alternatives and locates new solutions.¹⁸ For simplicity's sake, let us suppose that each of our three individuals can, after reflection, decide how well each member of the set solves their fairness problem, though at the outset each simply considers a few options that their perspective makes salient. Alf, for example, might start off with the option most likely to pass the Senate, Betty the simplest alteration of the tax code, and Charlie looks at the proposal that is best for the middle class. Suppose, then, that if all the options were presented to them they would be ranked as $\{a > b > c > d > e\}$; call this the objective fairness ordering. Because they have different perspectives on the problem, they do not immediately see all the

possibilities. Suppose we have the three perspectives:¹⁹

Alf: $e-c-d-b-a$

Betty: $d-c-b-e-a$

Charlie: $e-d-c-b-a$

Suppose that Alf deliberates about the fairest policy, using the simple method (what Hong and Page call a *heuristic*: a strategy for locating solutions within a perspective (Page 2007a, chap. 2))²⁰ of taking the first option in his perspective, and then searching for a better answer until he gets to the point where the next thing he thinks of is worse; at that point he stops his search. In more formal terms, this would be a “local optimum” for Alf; given the way he understands the problem, he has arrived at the best answer he can see — the next step is worse. Employing this idea, Alf might start with e (the most likely proposal to pass the Senate), but he then sees whether he can find a better policy by searching his perspective. So he next considers c , which is indeed an improvement. Having found this improvement Alf is excited and keeps on searching; however, the next proposal that his perspective suggests is d , which is worse than c . Disappointed, he stops. But Betty can, as Bosanquet indicated, pick up from there; she locates c on her perspective and finds that, using the same heuristic as Alf, she can arrive at a better solution, b . She, however, gets stuck at b , for it is surrounded on both sides by worse options (c and

e). Charlie then thinks about his middle-class tax perspective and where b is on it; he sees that, again, using the same heuristic, he can improve on b , and can move to a , the “global optimum.” They thus arrive at the most fair option.

This toy account shows why the reasoned control of each would support a — on each of their views it is the best and, interestingly, none would have arrived at it on their own. Thus the general will, as Bosanquet argues, augments the reasoning of each; by each reasoning to, say, level l (correctly employing our simple heuristic given their perspective, for example), they can collectively arrive at a result that is better than anyone could have arrived at by alone employing l . Although none of our agents can arrive at a , *the objectively best answer*, by exercising their reasoned control alone, if they reason together, picking up in the process where others left off, by the end of their collective problem-solving enterprise, they can each see how the exercise of their reason leads them to endorse a . The idea, as we saw above with Bosanquet, is that people can build on the solutions of others to produce moral improvement.²¹ Of course there is a critical difference: in the end all of our perspectives do fully grasp the general will (the global optimum); unlike in the rules and order version (§4.1), the general will arises from an explicit search.

To better appreciate the power the Hong-Page analysis, let us imagine a community containing members of unequal individual ability — from the very good to the mediocre. Now divide the community into two groups: one homogenous group composed only of the best reasoners, in which the individual members each

possess approximately the same perspective, and another group that contains a diverse group of competent reasoners, but not the best. Under certain conditions,²² the diverse group will outperform the better, but homogeneous, group (this is the Hong-Page *Diversity Trumps Ability Theorem*) (Page 2007a, pp. 158-65). The key to the result is that the best problem solvers tend to have similar tools and solve problems in similar ways, so that a group of the best problem solvers tends not to perform much better than any single member of the group does. On the other hand, a group of relatively good, but diverse problem solvers, can perform collectively better than the group of homogeneous experts (Ibid., p. 137). Indeed, when the relevant conditions are satisfied, a group of diverse problem solvers must outperform the best individual problem solver. This does not show that ability does not matter, but rather, once individuals meet a certain ability threshold, diversity matters even more than does ability for solving difficult problems (see Page 2007b, p. 11).²³

On this account, diversity generates a type of “superadditivity” (Page 2007a, pp. 339-40): “When a collection of people work together to solve a problem, and one person makes an improvement, the others can often improve on this new solution even further. Problem solving is not the realization of a state but a process of innovation in which improvements build on improvements” (Page 2007b, pp. 13-14). Bosanquet believed that the general will arises through a complex interaction of individual viewpoints, and this, it turns out, is the case with problem solving through diversity. To realize the benefits of diversity, however, the members of our

diverse community must interact so that they can build upon the work of each other to spot solutions where others have gotten stuck or “stopped.”²⁴ It will also be necessary to describe with some specificity the particular moral problems our diverse agents are working to solve.²⁵ It is implausible to think of our agents as trying to divine the general will in terms of the “entire system of moral relations.” Rather, we can think of them as trying to solve the discrete moral problems with which they are confronted. This might include concrete questions surrounding the distribution of burdens and benefits, the morality of physician-assisted suicide, the death penalty, certain interrogation techniques, and abortion, difficulties involving the ethical treatment of nonhuman animals, and so forth. By solving these discrete problems (and so many other distinct moral problems to which we have already engineered more settled solutions), the general will in terms of the “entire system of moral relations” can emerge, even though no one is attempting to discern the whole scheme.

We do not wish to suggest that the Hong-Page theorem and its approach to diversity are unproblematic. As we saw in our toy example, the approach supposes that while each perspective disagrees about how each option is related to the others, there is consensus about the value of each option as a solution to the collective problem. We certainly might query the stability of this combination of diversity in how the options are understood with consensus about the value of each. The theorem also supposes that each perspective, while viewing the options very

differently, nevertheless successfully communicates its findings to individuals with other perspectives. These are real problems.²⁶ However, the point is not that the Hong-Page approach to diversity is without difficulties, but that it demonstrates, in a sophisticated and thoughtful way, how the very diversity that seems to militate against the formation of a general will might, instead, be the engine of its discovery.

5 POSITIVE FREEDOM AND MORAL RESPONSIBILITY IN A DIVERSE SOCIETY

As we have argued, to sensibly hold another morally responsible we must suppose that the person has reasoned control over his actions and that, exercising that control, he would be capable of acting on the relevant moral directive. From this thesis one might proceed in two ways. One might suppose that moral responsibility is unproblematic — *of course* we can be held morally responsible. Given this, the preferred conception of freedom simply will be one that shows us to be fit for moral responsibility. We might say this this approach reasons *from* our acknowledged responsibility *to* a conception of freedom that supports it (this, we suspect, is along the lines of Pettit's analysis). The other tack, which we have followed here, is to first analyze the concept of freedom as reasoned control, and then *inquire*: are agents with such freedom fit for responsibility, given the character of their morality? This allows that in certain social orders agents with positive freedom may not be fit for responsibility, because they could not use their reasoned control to comply with it. On this second view we need to ask under what conditions free agents will be fit for

responsibility.

Taking this second route, we have argued, ties positive freedom to the general will: how could a morality be such that each person's positive liberty is congruent with the demands of a shared morality, and so each is fit to be held responsible? Rousseau's own solution, we have suggested, does not seem appropriate to a diverse society — it cannot show how members of a diverse society are all fit to be held responsible for violating the moral rules that structure their lives. However, Rousseau's account does not exhaust theorizing about the general will. Picking up on some clues from Bosanquet, we have very briefly sketched two analyses according to which a highly diverse society can share a general will. Our aim was not to establish these accounts, but to show that Rousseau's project — of showing how each can be free yet obey the moral laws — is alive and important today.

References

- Benn, SI 1988. *A Theory of Freedom*. Cambridge: Cambridge University Press.
- Benn, SI & Weinstein, WL 1971. "Being Free to Act, and Being a Free Man." *Mind*, New Series, vol. 80 (Apr.), pp. 194-211.
- Berlin, I 1969. "Two Concepts of Liberty," in I. Berlin, *Four Essays on Liberty*. Oxford: Oxford University Press, pp. 118-172.
- Bicchieri, C 2006. *The Grammar of Society*. Cambridge: Cambridge University Press.
- Bosanquet, B 2011a. "The Antithesis Between Individualism and Socialism Philosophically Considered" in *The Philosophical Theory of the State and Related Essays*, eds. GF Gaus & W Sweet. South Bend, IN: St. Augustine's Press, pp. 324-44.
- Bosanquet, B 2011b. "The Notion of a General Will" in *The Philosophical Theory of the State and Related Essays*, eds. GF Gaus & W Sweet. South Bend, IN: St. Augustine's Press, pp. 305-311.
- Bosanquet, B 2011c. "Psychological Illustration of the Idea of a Real or General Will" (chap. VII) in *The Philosophical Theory of the State and Related Essays*, eds. GF Gaus & W Sweet. South Bend, IN: St. Augustine's Press, pp. 160-76.
- Bowles, S & Gintis, H 2011. *A Cooperative Species*. Princeton: Princeton University Press.
- Brennan, G, Eriksson, L, Goodin, RE & Southwood, N 2013. *Explaining Norms*. Oxford: Oxford University Press.
- Chapman, JW 1956. *Rousseau — Totalitarian or Liberal?* New York: Columbia University Press.
- Feinberg, J 1986. *Harm to Self*. Oxford: Oxford University Press.
- Freud, S 1973. *Introductory Lectures on Psychoanalysis*, trans. J Strachey Harmondsworth, UK: Penguin Books.
- Gaus, G 1994. "Green, Bosanquet and the Philosophy of Coherence," in *The Routledge History of Philosophy*, gen. eds. SG Shanker & GHR Parkinson, vol. 7 *The*

- Nineteenth Century*, ed. CL Ten, London: Routledge, pp. 408-36.
- Gaus, G 1997. "Does Democracy Reveal the Will of the People? Four Takes on Rousseau." *Australasian Journal of Philosophy*, vol. 75 (June), pp. 141-162.
- Gaus, G 2001. "Bernard Bosanquet's Communitarian Defense of Economic Individualism" in *The New Liberalism: Reconciling Liberty and Community*, eds. D Weinstein & A Simhony. Cambridge: Cambridge University Press, pp. 136-158.
- Gaus, G 2011. *The Order of Public Reason*. Cambridge: Cambridge University Press.
- Gaus, G & Hankins, K (forthcoming). "Searching for the Ideal: the Fundamental Diversity Dilemma." In *Political Utopias*, eds. M Weber & K Vallier. Oxford: Oxford University Press.
- Gauthier, D 1986. *Morals by Agreement*. Oxford: Oxford University Press.
- Green, TH 1986. "On the Different Senses of 'Freedom' as applied to the Will and the Moral Progress of Man," in TH Green, *Lectures on the Principles of Political Obligation and Other Writings*, eds. P Harris & J Morrow. Cambridge: Cambridge University Press, pp. 228-249.
- Hayek, FA 1960. *The Constitution of Liberty*. London: Routledge & Kegan Paul.
- Hayek, FA 1973. *Rules and Order*. London: Routledge.
- Hayek, FA 1988. *The Fatal Conceit*, ed. WW Bartley III. Chicago: University of Chicago Press.
- Hong, L & Page, SE 2004. "Groups of Diverse Problem Solvers can Outperform Groups of High-ability Problem Solvers." *Proceedings of the National Academy of Sciences*, vol. 101, no. 46 (Nov. 16, 2004), pp. 16385-16389.
- Kramer, MH 2003. *The Quality of Freedom*. Oxford: Oxford University Press.
- Landemore, H 2013. *Democratic Reason*. Princeton: Princeton University Press.
- MacCallum, G 1972. "Negative and Positive Freedom" in *Philosophy, Politics and Society*, 4th ser., eds. P Laslett, WG Runciman & Q Skinner. Oxford: Basil Blackwell, pp. 174-93.

- Page, SE 2007a. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies*. Princeton: Princeton University Press.
- Page, SE 2007b. "Making the Difference: Applying a Logic of Diversity." *Academy of Management Perspectives*, vol. 21, no. 4 (Nov.). pp. 6-20.
- Page, SE 2008. "The Power of Diversity." *The School Administrator*, vol. 65, no. 9 (Oct.), pp. 35-41.
- Pettit, P 1997. *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.
- Pettit, P 2001. *A Theory of Freedom: From the Psychology to the Politics of Agency*. Oxford: Polity.
- Rawls, J (2007). *Lectures on the History of Political Philosophy*, ed. S Freeman. Cambridge, MA: Harvard University Press.
- Rousseau, JJ 1988. *The Social Contract*, in *Rousseau's Political Writings*, eds. A Ritter & JC Bondanella, trans. by JC Bondanella. New York: Norton.
- Sen, A 2002. "Arrow Lectures" in A. Sen *Rationality and Freedom*. Cambridge, MA: Harvard University Press, pp. 581-712.
- Smith, M 1994. *The Moral Problem*. Oxford: Blackwell.
- Strawson, P 1961. "Social Morality and Individual Ideal." *Philosophy*, vol. 36 (January), pp. 1-17.
- Strawson, P 1962. "Freedom and Resentment." *Proceedings of the British Academy*, vol. 48, pp. 188-211.
- Swanton C 1992. *Freedom: A Coherence View*. Indianapolis: Hackett.
- Taylor, C 1979. "What's Wrong with Negative Liberty" in *The Idea of Freedom*, ed. A Ryan. Oxford: Clarendon Press, pp. 175-194.

¹For a sophisticated defense of a negative liberty view, see Kramer (2003).

² Philip Pettit (2001) explicitly argues this; for another such example, see Swanton (1992).

³ This is the sense in which S. I. Benn (1988) offers a “theory of freedom” (see also Benn & Weinstein, 1971). From a very different perspective, this is Amartya Sen’s (2002) approach.

⁴ “She found out the central meaning of her ceremonial one day when she suddenly understood the meaning of the rule that the pillow must not touch the back of the bedstead. The pillow, she said, had always been a woman to her and the upright wooden back a man. Thus she wanted — by magic, we must interpolate — to keep man and woman apart — that is, to separate her parents from each other, and not allow them to have sexual intercourse. . . . If a pillow was a women, then the shaking of the eiderdown till all the feathers were at the bottom and caused a swelling there had a sense as well. It meant making the woman pregnant; but she never failed to smooth away the pregnancy again, for she had been for years afraid that her parents’ intercourse would result in another child. . . .” (Freud 1973, pp. 307-8, paragraph break deleted).

⁵ But perhaps the teacher should be considered as a moral expert. Suppose the undergraduate was directed by a doctor to administer some medicine to his ailing mother who is in his care. Although he may be unable to understand the biochemical molecular interactions that are responsible for making this particular medicine an effective treatment for his mother’s condition, we would nonetheless hold him morally responsible for disregarding the doctor’s expert counsel. In other words, although he may not be able to appreciate the reason why the medicine will cure his mother’s condition, he nonetheless has a reason (i.e., the direction of an expert) to administer the medication. Perhaps we should think that the undergraduate has similar reason with respect to the demanding transcendental

deduction or intricate formal proof. This matter brings to light difficult problems of the epistemic warrant of expert testimony. Even if the undergraduate does not understand the science behind the expert medical advice, presumably he does have understandable evidence that the doctor is an expert — if there is no such sufficient evidence then he is not responsible for ignoring the medical advice. And here lies the rub: does the undergraduate have evidence that this instructor is an expert on *morality* (as opposed to proofs, or the literature in moral philosophy)? There are many reasons to answer in the negative. On simply epistemic grounds, if the body of purported experts disagree among themselves, then there is apt to be insufficient evidence as to whom is the expert. More generally, an ideal of moral autonomy may entail that free moral agents cannot morally follow expert authorities without full understanding — such action would have no moral worth. Thus it is reasonable to suppose that the undergraduate need not defer to the doctor's conclusion that *morality requires* the undergraduate to administer the medicine to his mother, only to the doctor's judgment that *this medicine* is the one he should administer. In any case, if our undergraduate is truly unable to appreciate the reasons why he should administer some kind of aid to his ailing mother, while his lack of appreciation does not make that demand false, it does suggest that our practice of holding him responsible will be profoundly modified (much as it is with young children whose understanding of moral requirements often fails, but whom we might expect to defer to a parent during the learning process). Our thanks to a reader for Oxford University Press for bringing this matter to our attention.

⁶ Benn (1988, pp. 155-64) argues that the person must be free from three classes of defects — epistemic, of practical rationality, and of psychic continuity. See also Joel Feinberg's analysis of voluntary action in *Harm to Self* (1986, p. 115).

⁷ Compare, for example, Michael Smith's view (1994, pp. 151ff) with that of Gerald

Gaus (2011, §13).

⁸ There is evidence that Rousseau, at least on occasion, saw the problem in the way we have rephrased it. In arguing against unlimited forms of government with “absolute authority” and requiring “unlimited obedience,” Rousseau writes that “Such a renunciation [of one’s liberty] is incompatible with man’s nature, and *to strip him of all freedom of will is to strip his actions of all morality*” (1988, p. 89 (Bk. I, chap. 4), emphasis supplied). For Rousseau, then, any form of moral community that strips people of their positive freedom, will also rob them of moral responsibility because such freedom is a necessary condition for responsibility.

⁹ This holds except with respect to the initial formation of the state, which must be accomplished via the unanimous vote of those who will become citizens (see Rousseau 1988, p. 151 (Bk. IV, chap. 2)).

¹⁰ See also, e.g., Rousseau 1988, p. 112 (Bk. II, chap. 9): “The same laws cannot be appropriate for so many diverse provinces, which have different moral habits as well as contrasting climates and cannot all tolerate the same form of government.”

¹¹ For an excellent analysis of different accounts of the general will, see Chapman (1956, chap. 10).

¹² Unlike Rousseau, who thought that vast differences in circumstances would make it difficult or impossible for individuals to be governed under a single state (see note 10, above), Bosanquet thought such vast diversity of circumstance could be accommodated within a general will (see Bosanquet 2011b, p. 309: “It is even possible, and obviously usual, to support by our private will different arrangements in different localities, adapted to different conditions; and, in fact, this principle runs throughout our whole social and political life.”).

¹³ Bosanquet also stressed that for these diverse minds to form a coherent system

they must be organized by common psychological structures (see his fascinating 2011c).

¹⁴ Bowles and Gintis (2011), Bicchieri (2006, chaps. 1 & 3), and Gaus (2011, chap. 3) each offer examples of how normative rules that prevent us from succumbing to social dilemmas are endorsed by our evolved fairness commitments. This distinguishes these accounts from the traditional contractarian project, as exemplified by David Gauthier's *Morals by Agreement* (1986).

¹⁵ According to Rawls, Rousseau's idea of the general will requires a shared "point of view" (2007, pp. 229ff.).

¹⁶ Note that what is required here is *cognitive* diversity — different cognitive tools that individuals can bring to bear on the problem. This is distinct from identity diversity, involving gender, race, ethnicity, culture, age, etc. Still, identity diversity will often correlate with or drive cognitive diversity because people's cognitive tools are shaped by their life experiences, as well as by learning routines, rules and scripts from other contexts. And these things, in turn, are often shaped by our cultures, values and identities. Nonetheless, the benefits of identity diversity and its correlation with cognitive diversity vary by context (see Page 2007b, p. 19: "As life experiences often frame how people see social issues, for public policy problems identity differences can translate directly into diverse perspectives. On more scientific and technical problems, the linkages are less direct.").

¹⁷ There are many such communal problems, of different levels, from responses to terrorist acts to places for dumping trash (see Page 2007a, chap. 1, esp. pp. 30-33).

¹⁸ Not all perspectives will be useful ones. If we're trying to determine the best places to dump trash, a perspective that considers only nearness to the individual's current location will probably not be as useful as one that organizes solutions based

on long-term viability (see Page 2007a, p. 35).

¹⁹ For a somewhat similar example, see Landemore (2013, p. 102).

²⁰ In this example, we consider only one simple heuristic. However, agents may employ more than one heuristic, and different agents may employ diverse heuristics, which, when combined with each other or paired with different perspectives, allow agents to locate solutions that might not be found otherwise.

²¹ Two other important cognitive tools also contribute to diversity: interpretations and predictive models. *Interpretations* classify or sort parts of the world; they provide mappings of features of the world into mental categories (Page 2007a, chap. 3, esp. pp. 79-81). For example, we might sort that place called *Rocco's Little Chicago* into the category of pizza restaurants, or classify that cold, brown, creamy ball as a kind of chocolate dessert. *Predictive models* then use these interpretations, which have sorted features of the world into categories, to tell us what might happen (Ibid., chap. 4, esp. pp. 92-94). For example, I might predict that I'll like that dessert because I like creamy, chocolate things. These interpretations and models make regular appearances in our everyday thinking, from predicting whether one will like a dessert, to predicting which policy will effectively deter violent crime, or which area of specialty in philosophy will be most sought-after on the job market.

²² They are: (i) the problem is difficult enough that no single individual always locates the best solution; (ii) the problem solvers are smart, and have relevant cognitive tools; that is, they are each capable of composing a list of their local optima; (iii) the problem solvers are diverse — for any proposed solution other than the global optimum, there is at least one individual problem solver who can locate an improvement; (iv) the problem-solving group must be at least reasonably sized and drawn from a relatively large population so that the collection of problem

solvers is sufficiently diverse (Page 2007a, pp. 159-162). For a formal proof of the theorem, see Hong and Page (2004).

²³ We have only considered here the “Diversity Trumps Ability” theorem, which has strong conclusions, but also requires a set of conditions to be met. Page also develops the “Diversity Beats Homogeneity Theorem,” which is less demanding, and shows that more diverse groups of problem solvers will on average beat their more homogenous equals (2007a, pp. 153-157).

²⁴ Strictly speaking, diverse members of the community cannot realize the benefits of diversity just by interacting with one another; they must do so with a positive outlook toward the advantages that diversity can generate: “If people do not believe in the value of diversity, then when part of a diverse team they’re not as likely to produce good outcomes” (Page 2007b, p. 7) (see also Page 2008; Page 2007a, pp. xiv-xviii).

²⁵ Our thanks again to a reader for Oxford University Press for encouraging us to clarify this point.

²⁶ These worries are developed in Gaus and Hankins (forthcoming).