


TOWARDS A THEORY OF THE GENERAL WILL*

1. Introduction

 ROUSSEAU, in his *Du contrat social*,¹ claims to resolve the conflict between “justice and utility,” reconciling individual moral autonomy with the state’s (justifiable) use of coercive force. His statement of the problem, and his resolution, are well known:²

[THE PROBLEM]: To find a form of association that defends and protects with the entire common force the person and the goods of each associate, and by means of which each unites himself to all, obeying only himself and still remaining as free as before.

[THE SOLUTION]: Each of us puts in common his person and his entire power under the supreme direction of the general will (*la volonté générale*), and we receive in a body each member as an indivisible part of the whole.

The general will is the mechanism by which the individual and the state, utility and justice, are reconciled; it is infallible (II.iii), indestructible (IV.i), and alone directs the forces of the state (II.i); it cannot be represented (III.xv), but its expression and execution is the source of the legislature’s undivided and inalienable sovereignty (II.i–ii); it may conflict with the particular will of any, and indeed every, citizen (I.vii, III.x, IV.i–ii); it may even, in a famous phrase, “force one to be free” (*on le forcera d’être libre*, I.vii 54).

But what *is* the general will, the possessor of these characteristics? Rousseau himself does not tell us, other than a few sketchy remarks that obscure

* I’d like to thank David Gauthier, Shelley Kagan, Arthur Ripstein, and Ken Westphal for comments on an earlier draft of this paper.

¹ Textual references are to the Garnier-Flammarion “Texte Intégrale” editions of *Du contrat social* (Paris 1966), by chapter, section, and page number; translations are my own.

² See I.vi 51–52: “Trouver une forme d’association qui défende et protège de toute la force commune la personne et les biens de chaque associé, et par laquelle chacun s’unissant à tous n’obéisse pourtant qu’à lui-même et reste aussi libre qu’auparavant. . . Chacun de nous met en commun sa personne et toute sa puissance sous la suprême direction de la volonté générale, et nous recevons en corps chaque membre comme partie indivisible du tout.”

rather than clarify.³ Nor is there consensus in the critical literature. The rest of this essay will offer a particular reconstruction of the general will. The term ‘reconstruction’ is worthy of note: the theory offered may not be the theory Rousseau had in mind, nor in its details could have had in mind. But there are good reasons to think it his, in outline if not in detail, and the reconstruction proposed has two virtues: it explains and justifies his claims about the general will, and it is continuous with modern problems in the philosophy of language and the philosophy of mind as well as political philosophy.

2. The Political Theory of Desire

The proper province of Rousseau’s political theory is the theory of desire, which Rousseau called *volunté*, ‘will’. Then as now, political discourse often centers on the satisfaction or frustration of desires. Hobbles began with a substantive hypothesis about human desires in the state of nature; taking rationality as the maximal satisfaction of desire, he offered his social contract as the result. Locke argued for limited government to allow for the pursuit of happiness, which is again the satisfaction of desires, as constrained by natural rights. Hume’s extended analysis of the passions is prerequisite to his ethical and political theories. Bentham based the hedonic calculus of utilitarianism on pleasure and pain, that is, the satisfaction or frustration of any given desire. Mill attempted to distinguish desires qualitatively, by their objects, but in the absence of a theory of desires met with little success. Then as now, little attention is paid to the analysis of desire itself, and political discourse is the poorer for it. Except for the tradition originating with Rousseau and culminating in Marx.

Rousseau’s work testifies to his interest in desire. His *Discours sur l’inégalité* carefully traces the evolution of desires, the expansion of needs, the mechanism of comparison, the emergence of esteem and vanity, the gradual transformation of *amour de soi* into *amour propre*. *Émile* and *La nouvelle Héloïse* treat at length the nature of education, taken as the process of molding belief and desire, attempting to transform *amour-propre* into something genuinely communal: the *amour de la justice* or *amour de la patrie* of the *Économie politique*. The so-called “Geneva Manuscript” contains a careful critique of civil religion (IV.viii) and means to inculcate the proper desires in the citizenry. Rousseau may have used the term ‘will’ (*volunté*) rather than the modern term ‘desire’, but that is mere terminol-

³ Such remarks, along with other political properties possessed by the general will, are explained in §5.

ogy; the general will is best understood through the analysis of desire.

The locutions discussed will be of the form “*S* wills that *p*,” where what is willed, ‘*p*’, is called the *object* of the desire—understood as any declarative sentence.⁴ I shall concentrate on desires that are *de re* or referential, in the form in which they are justifiably attributed to the agent.⁵ Problems familiar from other areas, such as substitutivity in opaque contexts, are relevant, but here they must be passed over. A complete theory of the general will would include such points.

3. Varieties of Desire

Any declarative sentence may count as the object of desire, but not all such will be relevant. Let us call the grammatical subject of ‘*p*’ what the desire is *about*: the item(s) purportedly referred to. Our attention will be restricted to *de re* desires about persons and their activities, that is, where ‘*p*’ has as grammatical subject a person, persons, or group of persons, attributing some activity to the subject.⁶ Clearly, such desires are either about oneself, about others, or about both.

Desires that are about oneself may be called *self-regarding desires*. They can be expressed in a variety of ways. For example:

(SRD) Schmidt wills that he, Schmidt, win the race.

⁴ My discussion will be in the contemporary linguistic mode for the sake of convenience: ‘*p*’ is the sentential complement of locutions expressing desire. (Of course, ‘*p*’ is not strictly a sentence, but a subjunctive clause closely related to a sentence; we can ignore this difference here.) A complete theory would incorporate much recent work in the philosophy of language in intensional contexts, supplemented by a full theory of intentional states from the philosophy of mind. Political philosophy would benefit immensely from such a synthesis.

⁵ The “justifiable attribution” clause allows the discussion to sidestep, though not completely avoid, difficulties in knowing what intentional states an agent possesses. But aside from this epistemological difficulty, note the distinction between *having* a desire and *being aware* that one has a desire. The content-clauses appealed to in the philosophy of mind and the philosophy of language for describing contents of intensional attitudes try to bridge the distinction by imputing awareness of just those aspects captured in the sentential formulation of the content-clause. Yet (as *e.g.* Tyler Burge has pointed out) there are real problems about the “graspability” of contents so described, and, further, Rousseau does not share this faith about the transparency of desire-consciousness. For the purposes of this paper, though, I shall assume that agents are completely aware of their desires. A fuller and more realistic account needs to be buttressed by treating how people come to recognize what the objects of their will(s) are.

⁶ “Activity” is used here in its broadest sense: hating, building a house, reaching puberty, and being laughed at all count as activities.

is a useful paradigm.⁷ Selfish desires are self-regarding desires, but not conversely. For example, Schmidt may will that he, Schmidt, win the race for the sake of the pleasure it would give his wife, though Schmidt himself cares nothing for winning and indeed would *ceteris paribus* prefer that his friend Janofsky, who does care about winning, win the race. In this instance, Schmidt's desire is self-regarding but not selfish. More generally, what fulfills a self-regarding desire need not be desired for the sake of satisfying the desire, and may even frustrate other desires of the agent.

Desires that are about others may be called *other-regarding desires*. The object of the desire may be expressed by using 'you' (singular and plural), 'he', 'she', or 'they', as in the paradigm:⁸

(ORD) Schmidt wills that they, Janofsky and Kruger, pick up the piano. Altruistic desires are other-regarding desires, but not conversely. For example, Schmidt may will that they, Janofsky and Kruger, have a piano dropped on them. In this instance Schmidt's desire is other-regarding but hardly altruistic. More generally, what fulfills an other-regarding desire need not be desired for the sake of satisfying the desires of those the given desire is about, or may frustrate desires of those the given desire is about.

Desires about several people including oneself may be called *communal desires*, expressed in the object of desire by the pronoun 'we', as in the paradigm:⁹

(CD) Schmidt wills that we, the Red-Headed League, copy the *Encyclopædia Britannica*.

The pronoun 'we' includes the relevant identificatory belief that one is a member of the group in question. While self-regarding desires may be selfish and other-regarding desires may be altruistic, there is no counterpart in

⁷ The related formulation "Schmidt wills that he win the race" is ambiguous, since Schmidt could be thinking of Janofsky. Likewise, "Schmidt wills that Schmidt win the race" has obvious difficulties if Schmidt does not know or believe that he is Schmidt, a problem most evident in the case of descriptions—if Schmidt wills that the tallest man win the race, then the desire is self-regarding if Schmidt believes himself to be the tallest man in the race, whether his belief is true or false, and otherwise not.

⁸ As with self-regarding desires, the occurrence of the pronoun dispels ambiguity and avoids the problems surrounding associated identificatory beliefs.

⁹ The clause "*can* be expressed" is important: communal desires may be formulated as explicit conjunctive subjects, as in "Schmidt wills that he, Schmidt, along with Garfinkel and Haynes, throw the baby out with the bathwater." This may be reformulated as: "Schmidt wills that we, Schmidt and Garfinkel and Haynes, throw the baby out with the bathwater. Equally, Schmidt's desire that the common treasury be used for poor relief can be reparsed as a communal desire, provided that Schmidt is a member of the relevant community and knows himself to be so.

common parlance for communal desires. However, it is clear that what fulfills a communal desire may indeed frustrate the desires of any, or indeed every, member of the group or the group itself.¹⁰ Note that communal desires need not be shared: Schmidt may will that we, the Red-Headed League (including myself), copy the *Encyclopædia Britannica*, but no other member of the society need have the same desire; Haynes, also a member, may will that the Red-Headed League copy instead the collected works of Arthur Conan Doyle.

Desires may be related in several ways. We shall concentrate on the satisfaction or frustration of desires; the examples will be about pairs of desires, and the cases can easily be generalized to multiple sets of desires.

First, desires may be *cosatisfiable*: the fulfillment of one desire does not prevent the fulfillment of the other desire. Schmidt's desire that Janofsky lose the race and his desire that Haynes lose the race are cosatisfiable, since each will be fulfilled if Kruger wins the race. Yet the desires may have nothing to do with one another. Garfinkel may will that he, Garfinkel, learn to ride a unicycle, and that they, Loring and Haynes, be jointly awarded the Nobel Prize in biology. Other things being equal, these desires are cosatisfiable.¹¹ Furthermore, there is no restriction to a particular agent; different individuals may possess cosatisfiable desires. For example, Loring may want Schmidt to lose the race, and Janofsky may want Schmidt to play the saxophone. Other things being equal—*e. g.* barring cases in which Haynes refuses to play unless Schmidt wins or Schmidt threatens to lose if Haynes plays the saxophone—the desires of Loring and Janofsky may each be satisfied.

Second, desires may be *co-fulfillable*: one and the same event may fulfill different desires. Again, the desires need not have anything to do with one another (other than being fulfilled by the same event), and they may be possessed by the same person or by different people. Loring's desire that

¹⁰ The proviso 'or the group itself' is added because some groups may literally be thought of as possessing interests themselves, above and beyond the interests of its current members: the Zeppelin Preservation Society has an interest in preserving zeppelins, even if, at a given time, each of its employees is fed up with zeppelins and would gladly consign them to the flames. The job each employee holds, though, requires the employee to act for the sake of preserving zeppelins, regardless of any actual desires, occurrent or habitual, of the employee. Agency need not be restricted to humans, at least in the political realm; corporations are agents.

¹¹ Cosatisfiability is a matter of the actual facts of the case: if the judges awarding the Nobel Prize refuse to make a joint award if Garfinkel learns to ride a unicycle, then Garfinkel's desires are not in fact cosatisfiable. Such claims are either about the actual facts of the case or carry an understood *ceteris paribus* clause.

Schmidt not receive the Nobel Prize and Janofsky's desire that Haynes not receive the Nobel Prize are co-fulfillable, since Garfinkel's being awarded the Nobel Prize satisfies each desire. Desires that are co-fulfillable are *a fortiori* cosatisfiable, but not conversely.

Third, desires may be *extensionally equivalent*: they each would be fulfilled by one and the same event, or set of events, though the event(s) be described in different ways. In the last example, Loring wanted Schmidt not to be awarded the Nobel Prize; these desires are co-fulfillable, but they are not extensionally equivalent, since their satisfaction-conditions diverge: Haynes's receiving the Nobel Prize would satisfy Loring's desire, and Schmidt's receiving the Nobel Prize would satisfy Janofsky's desire. However, "Garfinkel wills that he, Garfinkel, jump off the Brooklyn Bridge" and "Schmidt wills that he, Garfinkel, jump off the most famous bridge in New York City" are extensionally equivalent: since the Brooklyn Bridge is the most famous bridge in New York City, each desire would be satisfied by the same event or set of events, though described in different ways.¹² The way in which an event is described matters, since the strength of one's commitment to a desire may vary depending on the description of the desire. Again, a single person may have different but extensionally equivalent desires, or several persons may have extensionally equivalent desires. Desires that are extensionally equivalent are *a fortiori* co-fulfillable, and hence cosatisfiable, but not conversely.

Fourth, desires may be called *intensionally equivalent* when they are extensionally equivalent and the descriptions are the same, that is, when they describe the same events in the same way. For example, "Schmidt wills that he, Kruger, be run over by a truck" and "Loring wills that he, Kruger, be run over by a truck" are necessarily fulfilled by the same events, described in the same way. Intensionally equivalent desires are only individuated by the persons who have them; one person cannot have two distinct desires that are also intensionally equivalent.¹³ Obviously, intensionally equivalent desires are extensionally equivalent, hence co-fulfillable and cosatisfiable,

¹² These distinctions require a strong principle of event-individuation as well as a principle individuating descriptions. Such principles are notoriously difficult to spell out. Here I need assume only that they can be spelled out in some fashion; the exact principles do not matter. Note that desires are co-fulfillable if their satisfaction-conditions actually overlap, and they are extensionally equivalent if their satisfaction-conditions coincide. Each case permits deviations in other possible worlds.

¹³ One person may have distinct tokenings of a desire, which will then be intensionally equivalent, but then the *desire* does not differ from one time to the other, merely the tokenings.

but not conversely.

Each of these four ways desires may be related may be used as an equivalence-relation to generate equivalence-classes of desires. For example, if each member of a group has a desire that is mutually cosatisfiable with some desire each other member of the group possesses, then the group is universally cosatisfiable: each member can have some desire satisfied. (Of course, not every desire of every member may be cosatisfiable.) If each member of a given reference class possesses a desire in the equivalence-class generated by a particular relation, then we may say that a given relation is *universal*. The last relation, intensional equivalence, is the strongest, since then it is reasonable to say that each member has the same desire. A desire that is the same (with respect to intensional equivalence) and possessed by each member of a given reference class is said to be universally *universally held*.

4. Reconstructing the General Will

The general will is easy to define with these technical definitions. The following reconstruction captures its essential ingredients:

(GW-1) The *general will* is the set of all intensionally equivalent communal desires, each of which is universally held by the associates insofar as they are members of the community.

We may call a desire included in the general will, that is, a communal desire possessed (in its intensionally-equivalent form) by each associate crucial to membership in the community, a *social value*.

First, the clause “insofar as they are members of the community” marks out a subset of the intensionally equivalent communal desires held universally by each associate: only those explicitly taken to be criteria for membership in the community are included in the general will as social values. For example, it could turn out that each associate held an intensionally-equivalent form of the communal desire that we, the community of associates, build a cathedral, but unless this is a condition for membership in the general community it is not included in the general will.¹⁴ Universally-held desires that aren’t social values may be important to understanding the behavior of members of the community, but they are accidental to describing the character of the community as such.

Second, the clause that such desires be “universally held by the asso-

¹⁴ The same point can be stated as a condition of membership: if someone were not to hold an intensionally-equivalent form of the communal desire that is a social value, he or she would not be a member of the community.

ciates” reflects Rousseau’s insistence that the social pact be unanimous (I.vi); as he says, “unless there were some point on which all interests would be in accord, no society could exist.”¹⁵ Unanimity is essential for two reasons: (i) a minimal condition for an aggregate of persons to become a community is that their association produces a moral collectivity (I.vi); (ii) coercive force is only justifiable if each member freely subordinates himself to the general will (discussed in §6).

Third, they must be “intensionally equivalent communal desires,” since Rousseau insists on genuine community, the creation of a *moi commun*, in the social pact. The requirement that such desires are intensionally equivalent ensures that there is a fixed point among the desires of the associates, a particular set of desires possessed by each member of the community.¹⁶ The requirement that such desires be communal has several consequences. First, it reflects Rousseau’s insistence that an associate alienate himself and all his rights to the entire community (I.vi). The crucial desires in the social pact are those regulating how the community as a whole should be: “what generalizes the will is less the number of voices than the common interest that unites them.”¹⁷ Second, communal desires are Rousseau’s way of avoiding the Scylla of a state founded merely on self-interest and the Charybdis of utopian dreams of pure altruism. The former alternative, basing society on self-regarding, and indeed on selfish, desires was adopted by Hobbes, and cannot provide for any genuine community among the citizens.¹⁸ Other-regarding, and in particular altruistic, desires do not have the motivational force required either to bind a community together or to cause its formation.¹⁹ Communal desires, though, motivate the formation of a community

¹⁵ See II.i 63: “s’il n’y avait pas quelque point dans lequel tous les intérêts s’accordent, nulle société ne saurait exister.” Rousseau calls the formation of the community “le pacte social,” which, as he emphasized in IV.ii, must be unanimous.

¹⁶ Extensionally equivalent may not be sufficient, since the difference in descriptions among extensionally-equivalent desires might license different inferences, some of them practical, which may be drawn from the content of the general will. (Whether it does license such inferences will depend on the principle for individuating descriptions and on the nature of inferences in intensional contexts.) In addition, intensional equivalence is obviously a tighter ‘fit’ among desires, welding the associates together more firmly.

¹⁷ See II.iv 69: “ce qui généralise la volonté est moins le nombre des voix que l’intérêt commun qui les unit.”

¹⁸ Indeed, Hobbes has a notoriously difficult time in the *Leviathan* explaining how it is rational for members of his society to keep the contracts into which they enter.

¹⁹ It might be thought that exchange-relations are paradigmatic other-regarding desires that do have such motivational force: Schmidt wills that he, Janofsky, give Schmidt

and weld its members together in holding the same social values. For example, if Schmidt, Kruger, and Garfinkel are associates, then a particular social value may be represented as follows:

Schmidt wills that we, the associates, support free speech.

Kruger wills that we, the associates, support free speech.

Garfinkel wills that we, the associates, support free speech.

Each member of the community of associates, then, agrees on the appropriate activities for the community.²⁰ These common social values produce genuine collectivity.

The reconstruction given in (GW-1) places no restrictions on the content of the general will, on what may count as a social value. Rousseau, however, suggests a further clause: “the general will... changes its nature having a particular end, and cannot *qua* general pronounce on a person or a fact.”²¹ We might capture this by saying that the relevant desires must be concerned with institutional structures: to specify roles rather than persons, to be restricted to social arrangements that are themselves general.²² We can thus introduce the following characteristic:

(GW-2) A general will is *impersonal* if its social values are concerned only with institutional structures.

Communal desires are “impersonal” in that they contain no proper names, demonstratives, or definite descriptions. The range of permissible social values is thus (somewhat) limited with impersonal general wills. Hence Rousseau adopts the following principle:

[P1] Every general will is impersonal.

This is a strong claim. Is it defensible?

ten widgets in exchange for five gimcracks, while Janofsky wills that he, Schmidt, give Janofsky five gimcracks in exchange for ten widgets. On this line, mere co-fulfillable desires will be sufficient. But this is mistaken. While such other-regarding desires can motivate particular acts of exchange, they cannot motivate the formation of a community—consider the notorious difficulties involved in motivating the exchangers to abide by the terms of their bargain.

²⁰ Recall that the relevant identificatory beliefs are built into the use of ‘we’, and that ‘activities’ is taken in the broadest sense.

²¹ See II.iv 69: “la volonté générale à son tour change de nature ayant un objet particulier, et ne peut comme générale prononcer ni sur un homme ni sur un fait.” See also the next note.

²² See especially II.vi 74–75: “La loi considère les sujets en corps et les actions comme abstraites, jamais un homme comme individu ni une action particulière. Ainsi la loi peut bien statuer qu’il y yu aura des privilèges, mais elle n’en peut donner nommément à personne; la loi peut faire plusieurs classes de citoyens, assigner même les qualités qui donneront droit à ces classes, mais elle ne peut nommer tels et tels pour y être admis.”

There may be good reasons for adopting [P1], but it does not seem to be *required* by Rousseau's insistence on collectivity alone. A community of associates might indeed hold as a social value that the Reverend Jim Jones be their leader. Such a community may alter its social values, or perhaps simply disband, on the death of Jones, but the mere fact that it names Jones as leader seems no reason to deny that the community legitimately has a general will. Thus [P1] is too strong; some general wills may not be impersonal. Rousseau needs further argument.²³

Within the prescribed limits, then, anything at all may count as a social value. Were a community of associates to hold universally the communal desire that they play baseball every Sunday, baseball-playing would be a social value of their community. A Nazi community may hold racial purity as a social value; a Catholic community might establish an Inquisition. Rousseau does not set any *a priori* limits on the scope of the political. Baseball, genetics, religious belief, or anything at all may be in the political realm. Nor does Rousseau recommend particular social values. However, in consequence of its definition, the general will does have several properties that are political in nature.

5. Political Properties of the General Will

Sovereignty—the ability to determine the content of the general will—is clearly vested in the community as a whole, since the general will is determined by the communal desires universally held by the associates *qua* associates. Each associate contributes his share in forming the moral collectivity. Sovereignty thus conferred is inalienable (II.i): the general will is by definition tied to the community itself.²⁴ Moreover, sovereignty is indivisible, since it rests on the unanimous consent of each associate to the social values; “the will is either general or it is not; it is either that of the body of

²³ In the rest of the paper, the examples and arguments offered will conform to Rousseau's practice, involving only institutional structures, but nothing hangs on this fact; [P1] is not presupposed.

²⁴ If we reject [P1], allowing some general wills to not be impersonal, then the community might appoint a leader to dictate the content of the general will. But the source of the dictator's legitimacy is itself founded in the communal exercise of the general will, and the dictator may be deposed by another act of the whole community, for, as Rousseau says, a will cannot bind itself as regards the future (II.i 64). Thus sovereignty is inalienable, though it is possible to will in accord with the will of another. In such a case one is guided by the will of another. Yet even here the liberty to will as one pleases is retained; it is impossible to alienate one's own will (as Rousseau points out in his discussion of slavery, I.iv 49).

the people, or only of a part.”²⁵ The general will is also unrepresentable, “for the same reason it is inalienable: it is the same or different; there is no medium” (II.xvii). A genuine ‘representative’ would be able to bind the wills of the constituency represented. But this is impossible; the general will, by definition, must consist of desires each associate *actually* has. Yet the result sounds needlessly paradoxical, since Rousseau admits several forms of government with “representatives” who act as *deputies*: they do not and could not replace the general will, but they serve it.²⁶ Since popular sovereignty is inalienable and indivisible, Rousseau concludes that government is merely a trust wherein the legislature expresses the general will and the executive branch puts it into practice (III.i).

Furthermore, Rousseau asserts that the general will is infallible (II.iii). Yet it does not address matters of fact (II.iv), and so “infallibility” cannot mean that the facts are as the general will says them to be. Instead, “infallibility” has to do with value: the community alone establishes its social values in the general will; what it *takes* to be its social values *are* its values.²⁷ That is, the content of the communal desires included in the general will is solely up to the associates themselves, since they must agree in order to establish a community. Rousseau’s arithmetical way of putting this point must have seemed obscure to his contemporaries:²⁸

Take away fromn these same [private] wills the pluses and minuses which cancel each other, and the general will is left as the sum of the differences.

Suppose Schmidt and Janofsky wish to form a community. Now each has his own communal desires: Schmidt wants the proposed community to support free speech and racial equality, whereas Janofsky wants it to support free speech but not to support racial equality (to racially discriminate).

²⁵ See II.ii 64: “Car la volonté est générale, ou elle ne l’est pas; elle est celle du corps du peuple, ou seulement d’une partie.”

²⁶ Technically such deputies *try* to express or enact the general will, though the expression or enactment is in fact reserved to the members of the community themselves—which is why the community can replace the deputies, or alter the deputy-structure of the society. See the remarks about being ‘guided’ by the will of another in n 24.

²⁷ That is, the general will is “infallible” the same way the Supreme Court is infallible: the US Constitution simply is what the Supreme Court says it is, and the social values simply are what the community says they are. There is no external check on value, no independent test of what ought to be desired.

²⁸ See II.iii 66: “Ôtez des ces mêmes volontés les plus et les moins qui s’entre-détruisent, reste pour somme des différences la volonté générale.” The same point is suggested in II.i 63: “C’est ce qu’il y a de commun dans ses différents intérêts qui forme le lien social.”

Schmidt's positive desire that the community support racial equality (a 'plus') is cancelled Janofsky's negative desire that the community not support racial equality (a 'minus'). Hence racial equality, and racial discrimination, cannot be part of the content of the general will of any community Schmidt and Janofsky are associated with. Cases in which their communal desires do not differ, as in supporting free speech, will be left "as the sum of the differences," and so be included in the content of the general will of the community they form. Infallibility is nothing more mysterious than the communal constitution of social values.

The associates unite by the formation of a community, but they also unite in two further ways: they institute a government, and the particular form of government instituted specifies collective political interaction. Forming a community is only the first stage.

The institution of a government is not contractarian (III.xvi). If the particular form of government to be established is not specified in the content of the general will, then instituting a government is a distinct second stage. And for this stage unanimity is *not* required: majoritarian vote will suffice (IV.ii).²⁹ Further, different forms of government will be appropriate in different circumstances, as Rousseau sensibly points out (III.viii). Once a particular form of government is established, the citizens interact in prescribed ways: in representative democracy, for example, they select by vote deputies to express and execute the general will (III.iv); in direct democracy, laws are enacted by the community as a whole, again by vote. Legislative action is therefore a third stage, distinct from the preceding two.³⁰

Yet there is a difficulty. At the second stage, why should the individual subordinate his will to the decision-procedure of voting? At the third stage, why should the individual subordinate his will to whatever decision-procedure is involved in the particular form of government instituted? Rousseau rejects this question as ill-formed: each associate should vote as though on behalf of all (II.iv), where a vote is an opinion that something would benefit the state (IV.i). Thus 'vote' does not have its customary

²⁹ If unanimity were required, the second stage would not be distinct from the first. But this is too strong. A community whose social values do not determine the particular form of government is certainly possible: the associates may be united by a common interest, say the love of baseball, without this interest determining the particular form of government.

³⁰ Legislative action is a distinct stage, although if the government to be instituted is a direct democracy the difference between the democratic institution of the government (second stage) and the instituted government's legislative action (third stage) is slight—which is why Rousseau says that direct democracy is established "by a simple act of the general will" (III.xvii).

meaning:³¹

What is asked if [the associates] is not precisely whether they approve the proposition or reject it, but if it is consonant with the general will, which is their own; each, in casting his vote, expresses his opinion on the matter. . .

That is, each associate proposes a new social value. Voting is not simply a record of individual preferences about a given proposition, but rather an expression of preference by each individual who is already committed to the general will. Clearly, the more associates who propose intensionally equivalent forms of the same communal desire, the closer such a desire is to being incorporated in the general will.³² Yet this is not sufficient: a desire 'close' to being incorporated in the general will is not yet part of the general will; why should the dissenting voter be bound by the vote?

While [GW-1] is the foundation of a theory of community, it is static; no provision is made for altering the general will short of dissolving and reconstituting the community anew each time. But social values alter and unforeseen situations arise; communities of living individuals are dynamic. Hence a mechanism for reconstituting the general will short of the complete dissolution of the community might be adopted by a given community. Thus we may define the following characteristic:

(GW-3) A general will is *generable* if includes as a social value a reconstitutive mechanism.

Reconstitutive mechanisms can be of a wide variety of sorts: two-thirds majority vote, decree of the dictator, consensus of the barons. Rousseau seems to think that any community *must* have a reconstitutive mechanism. That is, he endorses the truth of the following strong claim:

[P2] Every general will is generable.

It could perhaps be maintained that in forming a community there must

³¹ See IV.ii 149: "Ce qu'on leur demande n'est pas précisément s'ils approuvent la proposition ou s'ils la rejettent, mais si elle est conforme ou non à la volonté générale qui est la leur; chacun en donnant son suffrage dit son avis là dessus. . ." Hence we may define a 'vote' as a single associate's expression of a communal desire that associate would like to be held universally (in intensionally-equivalent forms) by each associate insofar as each is a member of the community. The process of voting is the expression of such desires by each associate.

³² As Rousseau says in IV.ii 147: "Plus les avis approchent de l'unanimité, plus aussi la volonté générale est dominante." Different degrees of closeness may be appropriate for different desires; there is no need to be bound by majority rule. Rousseau suggests several practical alternatives, such as a two-thirds rule, in IV.ii. Further, it is clear that no associate may be excluded, since this would destroy the generality of the vote, even though unanimity is not required (II.ii).

be beliefs not only about how the community is to be but also how it is to continue to be, and the latter beliefs lead directly to reconstitutive mechanisms.³³ Yet while this may be true, it does not lead to the conclusion that a community must have a reconstitutive method *other* than dissolution and the attempted re-formation of the community. Hence [P2] is too strong; it does not follow from Rousseau's insistence on community. However, it seems as though most communities of interest to political philosophy will include a reconstitutive mechanism. This the restriction to communities with generable general wills, while not strictly required by the argument, is reasonable given Rousseau's interests and intentions.

Rousseau focuses on communities with generable general wills whose reconstitutive mechanism is the decision-procedure of voting. The passage defining 'vote' cited above continues as follows:³⁴

The counting of the votes provides the declaration of the general will. Thus when the opinion contrary to my own prevails, that shows nothing else than that I was in error, and that what I had taken to be the general will was not such.

A decision-procedure can declare the general will, as Rousseau indicates in this passage, only if it is a reconstitutive mechanism. When the general will is generable, "a citizen consents to all the laws, including those passed against him."³⁵ Yet this just points up the basic problem: why should the dissenting voter be bound by his previous consent to the reconstitutive mechanism? To explain this, we need to examine Rousseau's theory of freedom.

6. Freedom and Force

The general will reconciles the individual and the state; it may even "force one to be free." The proposed analysis of the general will illuminates the latter claim, and hence the former as well.

³³ This seems to be the import of I.v 50: "La loi de la pluralité des suffrages est elle-même un établissement de convention, et suppose au moins une fois l'unanimité."

³⁴ See IV.ii 149: "du calcul des voix se tire la déclaration de la volonté générale. Quand donc l'avis contraire au mien l'emporte, cela ne prouve autre chose sinon que je m'étais trompé, et que ce que j'estimais être la volonté générale ne l'était pas."

³⁵ See IV.ii. 149: "Le citoyen consent à toutes les lois, même à celles qu'on passe malgré lui, et même à celles qui le punissent quand il ose en violer quelqu'une." The rationality of such consent is interesting, but Rousseau does not pause to consider it—taking 'rationality' itself to be socially relative—and neither shall we. Rousseau is less interested in analyzing the rationality of social forms than their content since, whether rational or irrational, we do associate in such communities.

The classical tradition of political philosophy takes “freedom” to be the absence of *constraint*: the individual’s pursuit of the fulfillment of his desires is not to be interfered with or constrained in any way, subject to certain limitations (typically given by the proviso that no harm is brought to others). Now such limitations need not be taken as part of the very notion of freedom: they are rather a restriction of freedom, imposed by society.³⁶ Further, ‘constraint’ is often understood as *external* force, or the threat of same, but this too is not strictly necessary. Aside from the notorious difficulty of spelling out the internal/external distinction, neither force nor external agency in fact need be present. Rousseau says that the impulse of mere appetite is slavery (I.vi)—the heroin addict is at the mercy of his desires; his freedom is constrained, though not by force or external agency. Thus the central intuition in the classical tradition is that the pursuit of desires should not be constrained. As Rousseau says:³⁷

That man is truly free who only wills what he is capable of, and does what pleases him.

Freedom, then, is not having one’s desires frustrated, no matter the cause. Put positively, it is the ability to pursue (with hope of attainment) the fulfillment of one’s desires. Rousseau calls this “natural liberty” (I.viii).

Desires may be frustrated in many ways. Circumstances may prevent their fulfillment; other people may interfere; one’s abilities may alter. Yet one case is central for our purposes: two (or more—henceforth I drop this reminder) desires may not be cosatisfiable, so that at least one desire is frustrated. By definition, to the extent that one desire is frustrated, there is a lack of freedom; to the extent that another may be fulfilled, there is freedom. Schmidt and Janofsky each desire to win the race; at least one will be disappointed. Further, a single individual may have two desires that are not cosatisfiable. The Brooklyn Philharmonic and the Bronx String Quartet perform at the same time, so it is impossible to attend both; Kruger, who would like to see each of them, must have at least one of his desires frustrated.³⁸ Now Rousseau’s strategy is subtle. He takes cases in which the

³⁶ This is *not* to say that such restrictions are unwarranted. Far from it. But they need not be included in the analysis of freedom. One who can do as he pleases to satisfy his desires, including harming others, has relatively more freedom than one who is bound by the Harm Principle. This is compatible with the claim that restrictions on freedom are justifiable and warranted.

³⁷ See *Émile* (99): “L’homme vraiment libre ne veut que ce qu’il peut, et fait ce qu’il lui plaît. Voilà ma maxime fondamentale.” I am indebted to David Gauthier for this reference, and also for the relevance of the added below, ‘with hope of attainment’, since freedom is properly the *sufficiency* of power to desire.

³⁸ Note that Kruger’s case as described is compatible with his desiring to see the Brook-

autonomy of the individual seems opposed to the (justifiable) use of coercive force by the state and reduces them to the familiar case of an individual whose desires are not cosatisfiable.

Consider a community holding racial equality as a social value. By definition, Schmidt, an associate of the community, actually has an intensionally-equivalent form of the communal desire that the community not discriminate among races. Yet Schmidt may not want members of other races to move into the house next door. This latter desire conflicts with the general will. But since Schmidt actually has the desire included in the general will, Schmidt has desires that aren't cosatisfiable.³⁹ Rousseau's first move is to reject the (apparently complete) opposition between the individual and the state, locating the conflict instead between non-cosatisfiable desires actually held by individuals.

The second move should be obvious. Rousseau has to argue that in cases of conflict, the general will should predominate. Indeed, the general will must predominate even when the individual in question strongly prefers the desire included in the general will to be frustrated so that his other desire may be satisfied (a situation Rousseau explicitly recognizes in I.vii 54). Schmidt may be more committed to not allowing members of other races to move in next door than to supporting non-discrimination. Indeed, this is likely.⁴⁰ Why should the general will predominate?

lyn Philharmonic much more than the Bronx String Quartet. His decision is not problematic, but his desires are opposed nonetheless—a fact presupposed by the very notion of “decision.” Such cases of conflict are ubiquitous. Buridan's Ass is a special case, where the desires are equal in strength; weakness of will is a special case, where one desire is for what is believed better in some fashion.

³⁹ It is not important to the example that Schmidt's desire—“Schmidt wills that they, members of other races, not move in next door”—is other-regarding rather than communal; it's possible for a member of the community to hold incompatible communal desires. For example, Janofsky could have twelve reasons on balance supporting his communal desire that the community be Protestant, but also six reasons supporting the communal desire that the community be Catholic. At most one of these could be included in the general will, but that does not of itself simply eradicate the other desire. Indeed, Janofsky may revise his reasons and come to believe that the communal desire included in the general will is on balance the weaker of the two, and attempt to bring it about that the opposite desire be adopted in its place. The means he uses depend on the reconstitutive mechanism, if there is one: persuasion of all in a direct democracy; fiat if he is a dictator; bribing the barons in an aristocracy. If the general will is not generable, he may seek to dissolve the community and reconstitute it anew with the social value altered. The is that an associate may hold incompatible communal desires, and such cases are fully in line with the analysis offered here.

⁴⁰ Rousseau often says that private interests conflict with the general will (*e. g.* I.vii, II.ii–

Rousseau's answer is not evident, but we may construct a reply based on his remark (emphasis added):⁴¹

Hence in order that the social pact not be a mere empty formula, *it must tacitly include this clause*, that whoever should refuse to obey the general will shall be constrained [to do so] by the whole body: which signifies nothing other than that he shall be forced to be free.

We may recast the clause as a definition:

(GW-4) A general will is *effective* when it includes as a social value the (higher-order) desire that (i) the general will itself predominate in cases of conflict, and (ii) no other desires defeat this very desire.

The social value in [GW-4] is self-referential in virtue of (ii), which must be added to block the possibility of infinite regress.⁴² Hence Rousseau is committed to the strong principle:

[P3] Every general will is effective.

Is this defensible?

Just as there may be communities whose general will is not generable, it seems as though there may be communities whose general will is not generable.⁴³ There are three possible cases. First, the relevant social value

iii, III.ii, IV.i-ii); they may be in accord with it, but rarely at length, and typically strive against it. The general will, by its communal nature, tends to equality (in the sense of equal treatment: II.iii), whereas the private will tends to individual preferences (II.i).

⁴¹ See I.vii 54: "Afin donc que le pacte social ne soit pas un vain formulaire, il renferme tacitement cet engagement qui seul peut donner de la force aux autres, que quiconque refusera d'obéir à la volonté générale y sera contraint par tout le corps: ce qui ne signifie autre chose sinon qu'on le forcera d'être libre." See also IV.ii 149: "La volonté constante de tous les membres de l'État est la volonté générale; c'est par elle qu'ils sont citoyens et libres."

⁴² The threatened regress is as follows. If only clause (i) were satisfied, then there may be cases in which an individual not only prefers some particular desire to the social value included in the general will, but also prefers that the social value in (i) be frustrated. This puts us back in the initial condition: why should the social value in (i) predominate over the individual's preference? But the addition of (ii) provides an escape, for even in such cases the individual prefers that his preference that the social value of (i) be frustrated is itself frustrated. Technically, then, cases of conflict are reducible to cases of three or more non-cosatisfiable desires: the initial desire, the social value it conflicts with, and the social value of [GW-4] (and perhaps desires relative to the latter social value).

⁴³ Effectiveness is not the same as generability, though closely related. A generable general will need not be effective. There may be a reconstitutive mechanism, say majority vote, but after the vote is taken I dissent from the result; generability does not ensure bindingness. But there is a real distinction here. A reconstitutive mechanism is useful in cases in which there might be unanimity but no real call to have all the

might not be included for extraneous factors: simply inadvertence due to inexperience or inattention; the fact that the associates are united by common interests far removed from the political realm (*e. g.* the belief that bees are divine); and the like. Second, the content of the general will could be such that the associates believe, whether correctly or not, that such conflicts would never arise. In this case, there is no need for a social value ensuring effectiveness; the general will need not legislate possibilities that are believed not to occur. Third, the relevant social value as described might not be included due to other social values included in the general will. In a community in which the general will is not impersonal, the fiat of the dictator might provide effectiveness; equally, there might be other procedural rules present, or the relevant social value as described might be altered to permit side-constraints (every fourth offender may have his own desires predominate), or modified by other higher-order desires included in the general will.

Yet the core of truth in Rousseau's thesis is that any community that is more than a mere paper fantasy *must* agree on higher-order matters of procedure. Whatever form such procedures take, they must themselves be incorporated in the general will. The first kind of case is not interesting: such communities will readily feel the lack of such procedures, if they endure at all. The second kind of case is more interesting. If the associates are incorrect, then that community will also feel the lack of effective procedures. But if they are correct, there is no need to include rules of procedure. It seems unlikely that there could be such a community, but it is a genuine possibility. However, such ideally-running communities are in a way similar to communities that do have effective rules of procedure: there are no cases of conflict in which the action of the general will is unjustified, as a corollary of the fact that there are no cases of conflict. Thus the puzzles raised by conflict are avoided. That does not obviate Rousseau's attempt to resolve such puzzles when they *do* occur.

The third kind of case poses no real problems for Rousseau's approach. The fiat of the dictator itself rests on the general will, and hence requires further effective justification. Equally, side-constraints or other higher-order

associates meet to redefine the general will. Thus generability might be useful for cases in which, say, the content of the general will is extended. But some cases where reconstitution is called for might be cases of conflict, and in those cases the general will needs to be effective as well as generable. Note also that effectiveness does not entail generability: there need be no reconstitutive mechanism or other means for altering the content of the general will. In most interesting cases, though, generability and effectiveness go together.

desires regulating the application of the general will in cases of conflict are social values included in the general will itself, self-regulating and self-limiting mechanisms. For these rules of procedure to apply there must *at some point* be a social value of effectiveness of the sort suggested in [GW-4], for otherwise these rules of procedure would not avoid the initial difficulty themselves: why should the individual be bound to rules of procedure contained in the general will? Hence the general will must contain, in a sense, its own justification. Rules of procedures and limitations must themselves be licensed by an effective social value.

Thus *some* version of [P3] seems justifiable. General wills need not be effective in short-lived communities, but if they are to be long-lived, the lack of effectiveness will be readily apparent and effectiveness introduced. General wills need not be effective in very lucky communities either, but in such cases the puzzles plaguing conflicts do not arise, since conflicts themselves do not arise. Most communities, on the other hand, will have to include some version of the social value of effectiveness—the communities interesting to social and political philosophy—and here Rousseau’s analysis can apply. Moreover, it seems clear that social decision-procedures for adjudicating disputes are a *sine qua non* for any society, and indeed that our commitment to such is, or ought to be, greater than attachment to any particular or occurrent desire. The effectiveness of a general will is something any member of the society ought rationally to insist on. Local losses would then be offset by global gains, and the exchange of some frustration for the greater good of realizing most desires makes it an acceptable bargain.⁴⁴

How an associate can be “forced to be free,” at least in communities with an effective general will, should now be clear. In cases of conflict, the general will (1) frustrates some desire that the individual has, a process that may involve physical force, though it need not do so.⁴⁵ But (2) the general

⁴⁴ Rousseau suggests the rationality of subscribing to effective general wills in several places. At I.vi 50 he points out that forming societies is the obvious means to ensure individual freedom; at I.viii 55–56 he treats of membership in society by reckoning the advantages and disadvantages, arguing that on balance it is rational to adopt effectiveness; at II.iv 71 he describes the choice to join a community rather than to persist in the state of nature as “un échange avantageux.” Furthermore, we can see why Rousseau insists on the homogeneity of communities with regard to the general will, for an agent will join a community whose social values he adopts, and if the general will is effective then it will help the agent attain satisfaction of most of his desires most of the time. Hence it is rational to adopt an effective general will.

⁴⁵ More detail is called for here. It might be possible that certain kinds of violations of freedom, or certain means involved in the frustration of a desire, would *not* be justifiable on the part of the general will, even if for the sake of fulfilling the other

will only thwarts some desire for the sake of fulfilling two other desires the individual actually possesses, namely (a) the desire included as a social value in the general will, and (b) the social value [GW-4], the desire that the general will be effective (or the appropriate version thereof). Not preventing the satisfaction of desires—indeed, contributing to the realization of (2a) and realizing (2b)—is freedom. Schmidt's desire that members of other races not move in next door is frustrated, but only for the sake of fulfilling other desires he actually has, namely his desire that the community support racial equality along with his desire that the general will be effective.

Conclusion

The reconstruction of the general will described above allows more analytic rigor to be applied to Rousseau's arguments. It explains his claims on behalf of the general will. Yet there is a final worry. The theory of the general will seems best construed as a theory of community; its political properties are secondary. But if Rousseau himself was attempting to sketch a political theory, then his apparent negligence with regard to impersonality, generability, and effectiveness are even more striking, for these are precisely the characteristics possessed by the communities of greatest interest to the political philosopher. Is Rousseau's lack of precision, his blurring of distinctions, his non-theoretical discussion a sad comment on his abilities?

There is a better way of viewing Rousseau's project. In the tradition of the Enlightenment, social and political structures were taken to be justifiable to the extent that they were products of rational choice (at least hypothetically). Nothing intrinsic to man involved sociality. Aristotle's conception of man as a political animal was simply discarded. But Rousseau, the great critic of the Enlightenment, inverted its premisses: rationality itself is a function of the nature of social interaction, and the core of social interaction is man's very sociality itself, an intrinsic feature of (modern) man. In articulating the theory of the general will, Rousseau was not proceeding in the fashion prescribed by the Enlightenment, from rational Robinson Crusoes to the permissible forms of society; he rather characterized the central notion of community itself as embodied in existing political forms. In concrete cases, most communities possess impersonal, generable, and effective general wills. Notwithstanding the views of the Enlightenment, human so-

desire. Such constraints, however, would themselves have to be included in the general will as higher-order social values regulating the mechanism of the general will itself. If not, such communities are faced with dissolution in these cases. If the cases are sufficiently rare, a community might get by without these rules. It is a difficult problem. Rousseau simply leaves it unsolved, and indeed unaddressed.

cieties are not mere aggregates; they genuinely strive towards community. The theory of the general will codifies and expresses the social practices of human communities. Recognition of that fact can only lead towards an articulate self-understanding and improved social organization, which itself would define the very notion of rationality. The political theory of desire is not merely ancillary to politics, it is the central political fact to be understood. Rousseau never finished his project, a complete theory of the general will. But he needs to be rehabilitated, to have his theory taken seriously, as Locke and Hobbes are taken seriously. For:⁴⁶

The unchanging will of all the members of the state is the general will; it is that by which they are citizens and free.

⁴⁶ See IV.ii 149: “La volonté constant de tous les membres de l’État est la volonté générale; c’est par elle qu’ils sont citoyens et libres.”