CHAPTER 10

Kant and the Social Contract tradition[1]

In memory of Pierre Laberge

The fundamental idea of the Social Contract tradition is that consent or agreement can justify basic social and political institutions: just societies are based on the consent of the governed, unjust societies are not. As is well known, the tradition has had many forms, notably in political philosophies of the early modern period and in the luxuriant variety of contemporary contractualisms. As is equally well known, a discouraging list of standard difficulties stands in the way of the thought that consent can justify fundamental political and social arrangements.

One long-standing dispute is about the sort of consent needed for justification: should it be the actual consent of those involved, or the hypothetical consent that would be given by beings with a distinctive (e.g. reasoned, informed, disinterested) view of the matter? If actual consent is needed, should it be explicit or is tacit consent enough? Neither view of actual consent seems satisfactory for purposes of justification: we consent explicitly to too little, but (as it seems) tacitly to far too much. If, on the other hand, consent is only hypothetical, then it is quite obscure why it justifies. Why should the consent of hypothetical idealised rational agents, or of hypothetical beings in an ideal speech situation, or of persons in an artfully tailored hypothetical Original Position, justify the principles by which we are to live, who have never been any of these supposedly ideal beings, and probably cannot and do not aspire to become any of them? Moreover, even if hypothetical consent could justify, how should we choose among the many versions of hypothetical consent that have been proposed? And if there is a choice to be made, does not this very fact suggest that considerations other than consent are basic to justification?

[1] 'Kant and the Social Contract tradition', first published in François Duchesneau, Guy Lafrance and Claude Piché (eds.), *Kant actuel: hommage à Pierre Laberge* (Montreal: Bellarmin, 2000), 185–200; slightly revised version in Elisabeth Ellis (ed.), *Kant's Political Theory: Interpretations and Applications* (University Park: Penn State University Press, 2012).

On the surface it seems that Kant too must encounter these difficulties. He is generally taken to be a Social Contract theorist, and a good one. In the most distinguished formulation of twentieth-century contractualism John Rawls writes that 'My aim is to present a conception of justice which generalises and carries to a higher level of abstraction the familiar theory of the Social Contract as found, say, in Locke, Rousseau and Kant.'[2] Many knowledgeable writers on Kant's political philosophy during the eighties and nineties spoke confidently of him as a Social Contract theorist.[3] Patrick Riley even titled an article 'On Kant as the most adequate of the Social Contract theorists'.[4] However, others do not see Kant as a Social Contract theorist.

In his great political writings of the 1790s – *Theory and Practice*, *Perpetual Peace* and *The Doctrine of Right* (which forms the first part of *The Metaphysic of Morals*) – Kant provides a fair amount of evidence for the view that he is a Social Contract theorist. Numerous passages in *Theory and Practice*, in *The Metaphysic of Morals*, and some in the other political writings, discuss and apparently endorse a conception of the Social Contract. Yet Kant's basic justification of political institutions appeals not to the Social Contract but to the 'Universal Principle of Justice (Right)'. This principle is stated at the beginning of section 2 of *Theory and Practice* and in the opening passages of *The Doctrine of Right*: 'Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law' (*MM* 6:230; cf. *TP* 8:289).

The Universal Principle of Justice makes no obvious reference to consent, and Kant does not identify it with the Social Contract. Indeed, many of his central writings on justice, and specifically on the justification of state power, barely refer to the notion of a Social Contract.

So it is not wholly surprising that some writers argue that Kant does not fall within the Social Contract tradition at all. For example, Reinhardt Brandt has stressed both premodern elements in Kant's political philosophy and the importance of the so-called *lex permissiva* rather than of consent in justifying coercive state power.[5] Leslie Mulholland argues in *Kant's System*

[2] John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 11.

[3] For example, Susan Meld Shell, *The Rights of Reason: A Study of Kant's Philosophy and Politics* (Toronto University Press, 1980), 152; Howard Williams, *Kant's Political Philosophy* (Oxford: Blackwell, 1983): 'Kant's theory of justice is, substantially, a theory of contract', 94.

[4] Patrick Riley, 'On Kant as the most adequate of the Social Contract theorists', *Political Theory* 1 (1973), 450–70.

[5] See Reinhardt Brandt, 'Das Erlaubnisgesetz' in Brandt (ed.), *Rechtsphilosophie der Aufklärung* (Berlin: de Gruyter, 1982), 233–85, esp. 235ff.; see also Heinrich Bockerstette, *Aporien der Freiheit und ihre Aufklärung durch Kant* (Stuttgart: Frommann-Holzboog, 1982).

of Rights that his account of justice belongs not in the Social Contract tradition, but in the early modern Natural Law tradition.[6] Katrin Flikschuh argues that he belongs in neither.[7] Arthur Ripstein also emphasises the Universal Principle of Justice rather than contractualist elements in Kant's thought.[8]

Clearly, if Kant is a Social Contract theorist he is a peculiar one. Here I shall look at some of the peculiarities, and suggest that they may enable Kant to avoid some standard difficulties both of Social Contract theory and of contemporary contractualism. If I am right about the distinctiveness of Kant's relationship to the Social Contract tradition, it may not be feasible to see his position as fundamentally contractualist.

Justification by actual and hypothetical consent

Before turning to Kant, I shall consider some of the underlying claims of consent theories, including classical Social Contract theories and contemporary contractualist positions. Their basic thought is that consent can justify basic principles for social and political structures, hence indirectly those structures and even the coercion they inflict. This thought may seem intuitively plausible. When we have consented to something, we have no basis for objecting; we are not wronged when it happens; we have authorised what is done: *volenti non fit iniuria*.

On second thoughts, matters are far more complex. A start can be made by looking at the role of actual consent in justification. There are some cases where actual consent is not sufficient to justify and others where it is not necessary. Its importance is nevertheless not hard to discern.

Consider first how ordinary it is for us to think that actual consent is not sufficient to justify because it has been given to an action or policy of an unacceptable type. Suppose that A agrees that B may practise surgery on him; or to sell B one of his body organs (a kidney, say); or to let B push him from a dangerous height in return for payment, or to let B bully him, or to be B's slave. In these and countless similar cases we are likely to think that, all legalities apart, consent does not provide sufficient justification. We may be tempted to 'save' the claim that actual consent justifies by arguing that in such cases there was something defective about

[6] Leslie A. Mulholland, *Kant's System of Rights* (New York: Columbia University Press, 1990).
[7] Katrin Flikschuh, 'Freedom and constraint in Kant's metaphysical elements of justice', *History of Political Thought* 20 (1999), 250–71.
[8] Arthur Ripstein, *Force and Freedom: Kant's Legal and Political Philosophy* (Cambridge, MA: Harvard University Press, 2009).

the consent. We may argue, for example, that the danger or plain stupidity of consenting to dangerous treatment or profound violation of one's rights is evidence that such action cannot have been genuinely consented to by competent adults, that here consent is defective, just as in other cases it is defective because the consenting parties were not adequately rational, informed, free from duress and the like. However, this is a desperate line of argument: there is all too much evidence that people sometimes genuinely consent to action which may seem deeply unacceptable, even to action that profoundly injures, oppresses or degrades them.[9] Across the board insistence that any consent to such action must be flawed merely suggests an underlying refusal to consider the possibility that justification requires more than actual consent.

On the other hand there are plausible cases where actual consent does not seem necessary to justification. When laws are enforced, when parents control their children, when public authorities take emergency action, we have compliance rather than consent: yet the action taken may be wholly legitimate. No form of the Social Contract tradition even claims to show that all justifiable action receives actual consent. If we had to consent to every particular act of state for it to be legitimate, the central purpose of the Social Contract tradition – the justification of government – would be undermined. Clearly, actual consent is not always needed: it is no more necessary than it is sufficient for justification or just action.

Justification that appeals to consent in ethics and politics might, I suggest, more plausibly be thought of as having two stages, which are (generally) individually necessary and only jointly sufficient. First a certain type of action (relationship, institution, policy) must be shown acceptable; secondly, the parties involved in or affected by a particular instance of such an action (relationship, institution, policy) must (generally) consent to it. The only cases in which this second element of actual consent is not required is where the particular action (relationship, policy, institution) is of a type that has been shown exempt from such a requirement – for example because it is an instance of legitimate coercion, or legitimate exercise of authority.

[9] Phenomena such as adaptive and addictive preferences are deeply unsettling for attempts to base justification on actual consent. Some people adapt their preferences to the world and then go along with, accept, endorse, consent to being injured, exploited and oppressed; others are dominated by preferences they cannot adapt. See John Elster, *Sour Grapes* (Cambridge University Press, 1983) and *Alchemies of the Mind: Rationality and the Emotions* (Cambridge University Press, 1999); John Elster and O.-J. Skog (eds.), *Getting Hooked* (Cambridge University Press, 1999); Onora O'Neill, 'Justice, capabilities and vulnerabilities' in Martha Nussbaum and Jonathan Glover (eds.), *Women, Culture, and Dependency* (Oxford University Press, 1995), 140–52.

Daily life is full of examples of two-stage justifications. The myriad activities and transactions of domestic, commercial and professional life can be justified by showing that they are consensual, provided that such activities and transactions are of acceptable types, but not otherwise. Political life incorporates processes for allowing for actual consent and dissent: democratic consent given in elections, more rarely in referenda, most rarely in constitutional conventions, can justify governments, policies and states, provided that they are of acceptable types.[10]

Actual consent is not then unimportant either in political or in other forms of justification; but it needs to be supported by other arguments to show that the types of action (relationship, institution, policy) to which consent is given are acceptable. Moreover, even when actual consent is given to an action (relationship, institution, policy) that is of an acceptable sort, its justifying power reaches only a certain distance. Actual consent is a propositional attitude, so is given to a particular act or policy or institution as described by the agent. Since propositional attitudes are referentially opaque, actual consent will not automatically transfer from its initial object to the logical or causal implications of that initial object. This is why actual consent theories, although well designed to take note of consent given by those affected by action (relationships, policies, institutions) as perceived, are not well designed to show whether those affected also consent to related, even closely and necessarily related matters. Actual consent bears on aspects of the particular action (relationship, institution, policy) that an agent may take herself to encounter, and can offer a justification when these are of an acceptable type; it does so in a clear but necessarily narrow way. It cannot justify aspects of action (relationship, institution, policy) of acceptable types to which agents do not attend.

These considerations may suggest that any Social Contract theorist, and any contractualist, should rely as much as possible on appeals to hypothetical consent. Hypothetical consent theories are well designed to justify types of action (relationship, institution, policy); they do not demand demonstrations that anybody actually consents, so avoid the tiresome chore of demonstrating that consent has actually been given, and under which descriptions it has been given.

[10] These examples do not show that actual consent is a simple matter. On standard views, any actual consent that justifies must be informed consent. This small phrase covers a multitude of conditions: possession and use of adequate capacities to understand, to reason, and to exercise choice, absence of duress, ignorance and the like. Consent is a defeasible notion, which an indefinite number of conditions can render void, so without justifying power. See Neil C. Manson and Onora O'Neill, *Rethinking Informed Consent in Bioethics* (Cambridge University Press, 2007).

Although they are clearly designed to reach far, hypothetical consent theories do so at high cost. For two linked reasons a switch of focus to hypothetical consent fails to provide reasons for thinking that consent can justify. First, if hypothetical consent theories invoke conceptions of rationality (reasonableness, disinterestedness, etc.), their applicability to the human case will be in question unless they deploy empirically plausible conceptions of each notion. But many versions of hypothetical consent theory have invoked empirically false, highly idealised conceptions of rationality (reasonableness, disinterestedness, etc.). Second, if hypothetical consent theories do without idealising assumptions, and use only empirically accurate assumptions true of all choosers, they may not get very far. To repair the deficit of premises, hypothetical consent theories usually add more specific, empirical premises that refer to features and characteristics that are distinctively true of certain choosers, such as references to their citizenship or membership in certain states or communities. However, in doing this, hypothetical consent theories forfeit the possibility of justifying the features or characteristics they presuppose.

Hypothetical consent theories face an uncomfortable choice between idealised ('metaphysical') justifications and restricted ('political') justifications. 'Metaphysical' justification fails insofar as it relies on idealised claims that are false of many, even all, choosers; 'political' justification fails insofar as it relies on conceptions it needs to justify. For example, John Rawls's own later philosophy appealed to a 'political' justification that refers to peoples and citizens in its basic justificatory arguments, so presupposes rather than justifying boundaries, and (seemingly) at least minimal state powers.[11] Appeals to hypothetical consent, which assume and build on political notions such as these, cannot justify fundamental political arrangements.

Analogous dilemmas arise for contractualisms that appeal either to hypothetical preference orderings that meet idealised conceptions of coherence (connectedness, transitivity), or to the empirically established preference orderings true of specific choosers. If idealised accounts of preferences cannot be established, hypothetical consent theories may well slide towards actual consent theories and in so doing presuppose not only the structure of preferences orderings but the content of preferences – but as they do so, questions about the capacities of preferences to justify, and in particular about appeals to adaptive and addictive preferences, will become pressing.

[11] The terms 'metaphysical' and 'political' are used to distinguish different types of justification in Rawls's later work, first appearing in his 1985 'Justice as fairness: political not metaphysical', *Philosophy and Public Affairs* 14 (1985), 223–52; revised version in John Rawls, *Collected Papers*, ed. Samuel Freeman (Cambridge, MA: Harvard University Press, 1999), 388–414.

However, these limitations of hypothetical consent theories do not show that theories of justice are impossible. There is no particular reason for thinking that all reasoning that aims to show which *types* of action (relationship, institution, policy) are acceptable (so setting the context for justifying their particular instances by reference to actual consent) needs to build on a conception of hypothetical consent. The claim that actual consent is a second step, needed to complete a justificatory process (in all but a restricted set of cases, where coercion is shown legitimate), rather than a complete justification by itself, could be linked with a variety of conceptions of justification and of justice. However, if the basic moves of a theory of justice do not appeal to consent, we may no longer be considering any recognisable version of the Social Contract tradition. With these thoughts, back to Kant.

Kant's Social Contract: hypothetical consent

Kant states his views on the Social Contract most clearly in Part II of *Theory and Practice*. The section begins with the assertion that the Social Contract is completely distinctive and defines the 'unconditional and first duty in any external relation of people in general, who cannot help mutually affecting one another' (*TP* 8:289). He then drops all discussion of the Social Contract for several pages in order to set out the Universal Principle of Justice and to explain why it entails the three components of republican justice, namely freedom, dependence on law[12] and equality under law (*TP* 8: 290).

Only towards the end of the section does Kant resume discussion of the idea of the Social Contract. When he does so, he makes a negative and a positive claim: the Social Contract should not be thought of as an historical event; it should be thought of as an idea [*Idee*] of reason offering practical guidance.

Kant expresses the negative point as follows:

> But it is by no means necessary that this contract (called *contractus originarius* or *pactum sociale*), as a coalition of every particular and private will within a people into a common and public will (for the sake of merely rightful legislation) be presupposed as a *fact* (as a fact it is indeed not possible) – as

[12] Kant means quite specifically dependence *on law*, which he sees as protecting the independence. Common dependence on law is one of the guarantees of adequate independence, an independence Kant in fact extends only to some (active citizens). In *PP* he states that the Social Contract mandates common dependence or subordination to law. In *MM* he points out that 'civil independence . . . allows him to owe his existence and sustenance not to the arbitrary will of anyone else among the people, but purely to his own rights and powers as a member of the commonwealth', but once again distinguishes the case of the active and passive citizen (*MM* 6:139–40).

> if it would first have to be proved from history that a people, into whose rights and obligations we have entered as descendants, once actually carried out such an act, and that it must have left some sure record or instrument of it, orally or in writing, if one is to hold oneself bound to an already existing civil constitution. (*TP* 8: 297)

Here Kant is on uncontroversial ground. His position amounts to selective rejection of the claim that actual (past) consent (by others) can justify state coercive powers.[13]

Kant's second, positive claim can be, and often has been, read as showing his adherence to some version of hypothetical consent theory. For he asserts that the idea of the Social Contract is an Idea of Reason. In *Theory and Practice* the point is summarised as follows: 'the idea [*Idee*] of a social contract would remain . . . not however as a fact . . . but only as a rational principle for appraising any public rightful constitution'.[14]

However, I believe it would be a misreading to take this passage as showing that, having rejected the view of the Social Contract as an historical event, Kant concludes that hypothetical consent is the source of political justification. Consider the most explicit passage on the Social Contract in *Theory and Practice*, and in particular its modalities, which Kant repeatedly italicises:

> The [Social Contract] is instead *only an idea* of reason (*eine bloße Idee der Vernunft*), which, however, has its undoubted practical reality, namely to bind every legislator to give his laws in such a way that they *could* have arisen from the united will of a whole people and to regard each subject, insofar as he wants to be a citizen, as if he has joined in voting for such a will. This is the touchstone of any public law's conformity with right. In other words, if a public law is so constituted that a whole people *could not possibly* give its consent to it (as, e.g., that a certain class of subjects should have the hereditary privilege of *ruling rank*), it is unjust; but if it is *only possible* that a people could agree to it, it is a duty to consider the law just, even if the people is at present in such a situation or frame of mind, that if consulted it would probably refuse its consent. (*TP* 8:297)

Kant does not say here that the *hypothetical* consent of (idealised) rational beings constitutes the touchstone for justifying constitutions and laws,

[13] It is selective because while he rejects the view that the Social Contract was an historical event, he does not consider whether actual consent to basic institutions might be currently but tacitly expressed.

[14] Kant repeats this criticism of actual consent approaches quite often: for example, 'these writers have assumed that the idea of an original contract (a basic postulate of reason) is something which must have taken place in reality, even where there is no document to show that any contract was actually submitted to the commonwealth' (*TP* 8: 302; see also *MM* 6:143 and 6:162; *WE* 8:39).

and thereby fixes the terms of the Social Contract. He suggests that it is *possible* agreement that is decisive. He appeals to *possible* universal consent, not to hypothetical consent. The criterion is apparently modal and not hypothetical.

On first consideration there are unappealing aspects to a modal conception of justification. One problem of any modal view is that it looks too weak either to identify determinate constitutional requirements, or to justify them. Surely all sorts of constitutions and legislation, including much that seems palpably unjust, *could* be universally consented to. A second problem is that it is far from obvious why an appeal to *possible* consent should count as an Idea of Reason.

Kant's Social Contract: republican institutions

However, on Kant's view the criterion of possible consent has crucial substantive implications, and can itself be justified. I shall turn to the implications first. Kant argues that the criterion of possible consent can vindicate a republican constitution which guarantees freedom within the law: constitutions that lack this structure *cannot*, he claims, be universally consented to; constitutions with this structure *can* be universally consented to.

A succinct formulation of the components of a republican constitution is given in *Perpetual Peace*:

> A constitution established, first on principles of the *freedom* of the member of a society . . . second on principles of the *dependence* of all upon a single common legislation . . . third, on the law of their equality – the sole constitution which issues from the idea of the original contract, on which all rightful legislation of a people must be based – is a *republican* constitution. (*PP* 8:350; cf. *MM* 6:340; *TP* 8:290)

Kant's arguments from the idea of the Social Contract to the three basic components of a legitimate republican constitution are no doubt overly compressed. Yet I do not think it is hard to discern how his thought runs. If the idea of a Social Contract is that of a constitution which *could* secure universal consent, then any constitution that exemplifies it requires the freedom of individuals, without which the possibility of genuine consent or dissent is undermined, at least for some, making universal consent impossible. Secondly, it requires *their common dependence on or subordination to law*: if anyone were above or outside the law, freedom could be systematically or gratuitously undercut, and once again the possibility of

genuine consent or dissent would be undermined, at least for some, and universal consent consequently impossible.[15] Thirdly, it must endorse the *legal equality* of citizens, since the subordination of some individuals to others (rather than to the law) would once again undercut freedom, and with it the possibility of genuine consent or dissent, at least for some, and universal consent becomes impossible. In saying that these principles must be exemplified in any constitution that can be derived from the idea of the original contract Kant does not insist on constitutional uniformity: but he does claim that all just constitutions must meet these three quite demanding 'republican' conditions.[16]

There has, of course, been much criticism of Kant's limited conception of a republican constitution, and in particular of his lack of concern with democracy, his exclusion of those whom the world he knew made socially dependent (women, day labourers) from full citizenship and his tough views on the duty of obedience even to unjust rulers. I shall leave these interesting issues largely aside and comment on the fundamental reasoning which lies behind Kant's idea of the Social Contract and so behind the three elements of any republican constitution.

Kant's Social Contract: the Universal Principle of Justice

Kant's conception of the Social Contract undoubtedly has appealing features. These lie not so much in the particular constitutional demands for which he argues, as in his distinctive conception of the reasoning that lies behind his accounts of justice and of justification. These features of Kant's argument may enable him to avoid some of the difficulties of Social Contract theory and of contemporary contractualism that were discussed above (pp. 172–6).

[15] The term 'Idee' (Idea: capitalized in many translations) is a technical one for Kant, see below. In *Theory and Practice* he speaks of 'die Idee von einem ursprünglichen Vertrag, die immer in der Vernunft zum Grunde liegt' (*TP* 8:302); 'The social contract is *only an idea* of reason, which, however, has its undoubted practical reality, namely to bind every legislator to give his laws in such a way that they *could* have arisen from the united will of a whole people' (*TP* 8:297).

[16] Kant makes well-known and severe comments on the duty to obey and the wrongness of all rebellion even against a harsh ruler yet insists that his position is entirely different from Hobbes's (*TP* 8:303–5; 84–5). Hobbes had claimed that the sovereign has no obligations to the people, so can by definition do no injustice to a citizen. Kant insists that 'Whatever a people cannot impose upon itself cannot be imposed upon it by the legislator either' (*TP* 8:304; 85). He characterizes Hobbes's position as 'quite appalling' [*erschrecklich* – also translated as terrifying]. However, although Kant insists that the people have inalienable rights, these rights do not include a right to rebel against rulers who violate their rights, 'a coercive right against the one who has done him injustice'. In many (arguably not in all) his discussions of obedience Kant argues that the only legitimate action citizens may undertake in the face of injustice is to make representations: 'the freedom of the pen is the only safeguard of the rights of the people' (*TP* 8:304; cf. *WE*).

Kant reaches the view that his conception of the Social Contract is *'only an Idea* of Reason' (*TP* 8:297). The notion of an 'Idea of Reason' is a technical term which Kant explicates as early as the beginning of the *Transcendental Dialectic* of the *Critique of Pure Reason*, where he writes:

> A concept is either an **empirical** or a **pure** concept, and the pure concept, insofar as it has its origins solely in the understanding . . . is called *notio*. A concept made up of notions, which goes beyond the possibility of experience, is an **idea** [***Idee***] or concept of reason. (*CPR* A 320 / B 377)[17]

Already in this section of the *Critique of Pure Reason* Kant offered his version of the Social Contract as an example of an Idea of Reason:

> A constitution providing for the **greatest human freedom** according to laws that permit **the freedom of each to exist together with that of others . . .** is at least a necessary idea, which one must make the ground not merely of the primary plan of a state's constitution but of all laws too. (*CPR* A 316 / B 373)[18]

What lies behind the idea that this constitutional principle, unlike others, is an Idea of Reason? Put schematically Kant's justification of the Social Contract as an Idea of Reason relies on three linked claims. Proceeding from the most to the least abstract, these claims are:

A. The Categorical Imperative is the supreme principle of practical reason, so an Idea of Reason.
B. The Universal Principle of Justice states a version of the Categorical Imperative, restricted to the public domain.
C. The republican conception of the Social Contract is a special case of the Universal Principle of Justice, adjusted to certain historical conditions.

These three stages of justification may be traced as follows:

[17] See also a related passage that also articulates the modal basis of the Universal Principle of Justice: 'For, provided that it is not self contradictory that an entire people should agree to such a law, however bitter they might find it, the law is in conformity with right' (*TP* 8:299). The German is even sharper: 'Denn wenn es sich nur nicht wiederspricht, daß ein ganzes Volk zu einem solchen Gesetze zusammenstimme, es mag ihm auch so sauer ankommen, wie es wolle: so ist es dem Rechte gemäß.'

[18] Kant means quite specifically dependence on law. He sometimes speaks of dependence and at others of independence in this context. However, he has closely similar ideas in mind. Common dependence on law is one of the guarantees of adequate independence; independence Kant in fact extends only to some (active citizens). In *Perpetual Peace* he states that the Social Contract mandates common dependence or subordination to law. In the *Metaphysic of Morals* he makes the same point by referring to independence, 'civil independence . . . allows him to owe his existence and sustenance not to the arbitrary will of any one else among the people, but purely to his own rights and powers as a member of the commonwealth', but once again distinguishes the case of the active and passive citizen (*MM* 6:139–40).

A1 From the Supreme Principle of Practical Reason to the Categorical Imperative

Despite the fact that critique of reason is evidently Kant's central task, the way in which he proceeds is often overlooked or slurred, and the frequently asserted status of the Categorical Imperative as a principle of reason – indeed the supreme principle of practical reason – is left quite obscure. I think this obscurity is remediable. Kant presents extensive and coherent accounts of his views on the vindication of reason in a number of works, and in particular in the Doctrine of Method of the *Critique of Pure Reason*. A bald summary is all that I offer here.[19] I think that it would run along the following lines: we should take seriously the thought that we may lack any adequate account of reason, in that we may be unable to vindicate any generally authoritative ways of disciplining and organising either thinking or doing. What we provisionally take to be the powers of human reason may, as Kant suggests vividly in the prefaces of the *Critique of Pure Reason*, betray us and leave us in darkness and confusion. Moreover, we cannot expect to find any transcendent vindication of reason, such as rationalists have sought. In the Doctrine of Method of the *Critique of Pure Reason* Kant argues that if any ways of proceeding, any theoretical or practical principles, are to have a general or unrestricted authority for organising our thinking and doing, they must be principles that can be followed by all for whom they are relevant: principles could hardly have authority for those unable to follow them. The fundamental requirement of reason is modal: any reasoned way of thinking or acting must be one that is (viewed as) possible for its intended audiences to follow.

This line of thought can be seen as vindicating at least a very abstract and minimal, doubly modal principle for anything to count as theoretical or practical reasoning. Anything that is to count as theoretical reasoning must at least be based on principles which are followable in thought by all whom it is to reach: intelligibility for the intended audience is an indispensable condition for any theoretical reasoning. Anything that is to count as practical reasoning must at least be based on principles that are followable in choosing or willing, that is adoptable by all whom the reasoning is to reach. Universalisability, alias the Categorical Imperative, is a doubly modal requirement, which insists that reasons for action *must* be ones that *can* be adopted by all for whom they are to provide reasons.

[19] See the chapters in Part I for more detail.

B1 From the Categorical Imperative to the Universal Principle of Justice
The Categorical Imperative is broader in scope than the Universal Principle of Justice. Its most familiar version, and the one that makes its doubly modal character most explicit, is the Formula of Universal Law: *act only on that maxim through which you can at the same time will that it be a universal law* (*G* 4:421). It covers maxims for all sorts of action, inward and outward, personal and public. By contrast the Universal Principle of Justice is restricted in two ways. First, it is concerned only with maxims for outward action, that is with the aspects of action which could be enforced (hence not, for example, with maxims for virtue or with moral worth). Second, it is concerned only with maxims for *structuring* individuals' external freedom, that is with maxims for shaping the public domain, hence not with maxims for other outward aspects of individual conduct, such as the outward aspects of duties to self or personal relations.

The Universal Principle of Justice formulates a restricted version of the Categorical Imperative for the domain of 'the external use of freedom' – that is to say the domain of justice: it requires the rejection of any basic maxims for structuring the domain of *the external use of freedom* which *could not* be consented to or adopted by all. So the Universal Principle of Justice is a restricted version of the Categorical Imperative. This is wholly explicit in the formulation from the *Rechtslehre*:

> Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law. (*MM* 6:230)

Here there is no appeal to actual or to hypothetical consent: there is only a doubly modal claim that justice *requires* rejection of principles that *cannot* be adopted by all.

C1 Universal Principle of Justice to Social Contract
The Universal Principle of Justice remains very abstract; it might be satisfied in many different ways. It does not even claim that state structures and coercion are needed to secure a world in which people do not infringe one another's freedom. Indeed, since state coercion obstructs external freedom it may seem incompatible with the Universal Principle of Justice.

Kant argues that, contrary to initial assumptions, in *our world* only coercive structures can secure the Universal Principle of Justice. We are neither reliably altruistic, nor sufficiently scattered across the surface of the earth to avoid interaction. Consequently without a system of enforced

laws we will inevitably limit and invade one another's external freedom. In our world justice requires coercion: we need a power that will 'hinder hindrances to freedom' and cannot achieve any more complete realisation of freedom for all.[20] The Universal Principle of Justice conjoined with coercive enforcement is the basis for Kant's version of the Social Contract, which is a less abstract idea than his Universal Principle of Justice, in that it takes account of specific historical circumstances of human life in which unsociable beings 'cannot help mutually affecting one another' (*TP* 8:209), so demands a constitution which deploys state power to enforce laws to which all are subordinate and which guarantee legal equality. *Theory and Practice* contains Kant's most significant discussion of the idea of the Social Contract, because these specific historical conditions are borne in mind throughout the essay. By contrast, in *The Doctrine of Right* Kant's initial focus is broader, and the specific historical conditions that make coercion essential are addressed only in a second move, after consideration of the more abstract Universal Principle of Justice (*MM* 6:230ff.).

There is good reason to distinguish the two principles. The Universal Principle of Justice, as stated in *The Doctrine of Right* and elsewhere, might or might not require coercive powers. For example, if certain communities of rational beings were never inclined to do one another injustice, or if human beings lived the isolated lives of Rousseau's earliest state of nature and could not interact or do one another injustice, they would not need coercive structures. But Kant thinks that this is not the human condition. We are evidently not perfect altruists, and unsociability runs through our sociability. And we unavoidably interact since we are now numerous yet still inhabit the closed space of the earth's surface. In the conditions in which we live, the Universal Principle of Justice can only be realised by way of the more specific idea of a Social Contract, which makes explicit that for us only constitutional structures of a republican type are objects of possible universal consent.

Is Kant a Social Contract theorist?

The strands of argument can now, I think, be pulled together. Kant's conception of the Social Contract is a special case of his Universal Principle

[20] 'if a certain use of freedom is itself a hindrance to freedom in accordance with universal laws (i.e. wrong), coercion that is opposed to this (as a *hindering of a hindrance of freedom*) is consistent with freedom in accordance with universal laws, that is, it is right. Hence there is connected with right by the principle of contradiction an authorization to coerce someone who infringes upon it' (*MM* 6:231).

of Justice: it spells out what this principle requires in actual human conditions. The Universal Principle of Justice is in turn a special case of the Categorical Imperative: it spells out the most basic maxims to which our lives must conform if the structures of the public domain are to meet the requirements of the Categorical Imperative. The Categorical Imperative is the supreme principle of practical reason, indeed an Idea of Reason, because its double modal requirement formalises the recognition that we are not in the business of offering reasons to others unless we offer considerations that (we take it) can be followed by all and taken as reasons. Kant's version of the Social Contract can count as an Idea of Reason because he derives it from his account of practical reason, rather than from any appeal to actual or to hypothetical consent.

Can the fact that Kant labels a derivative, but still very abstract Idea of reason 'the Social Contract' make him into a Social Contract theorist? Labels perhaps do not matter much, and it is probably not useful to make essentialist claims about what makes a theorist a Social Contract theorist. But I think that Kant has in fact moved away from the basic insights of other versions of the Social Contract tradition to a different type of justification, to which neither actual nor hypothetical consent is fundamental. His basic thought is that when we think that others cannot adopt, *a fortiori* cannot consent to, some principle we cannot offer them reasons for doing so.

This distinctive if minimalist approach to justice and to justification has the advantage that it avoids some of the central difficulties of justification that have arisen in the Social Contract tradition. Kant's thought is not that coercion is justified because its exercise, or the institutions that exercise it, or the principles behind these institutions, have been, are or would be consented to, but because certain types of republican coercive institutions *could* be consented to by all who live under them, whereas their rejection *could not* be consented to by all. Non-republican institutions are non-starters for universal consent. The rejection, removal and replacement of non-republican institutions is the foundation of justice, the basis for identifying acceptable types of action (policies, institutions) and for establishing the proper contexts for actual consent by those who will be actually affected by particular acts (policies, institutions).

Constitutions that cannot be objects of possible universal consent are all of them unjust on Kant's view. Societies that do not secure the rule of law (anarchic or despotic societies) undermine or jeopardise external freedom for some or for all, so do not secure even the possibility of universal consent. Societies that leave some persons above or outside the law (monarchs, dictators, states within states) undermine or jeopardise external freedom

for some or for all, so do not secure even the possibility of universal consent. Societies that do not secure equality of status under law (feudalism, caste societies) undermine or jeopardise external freedom for some or for all, so do not secure even the possibility of universal consent. We may think that Kant's position would be improved by adding that further forms of social subordination which he accepted, but which also obstruct the possibility of consent, are unjust, and so that the rejection of patriarchy and the institution of full democracy are also required components of a just constitution. However, justifying these further claims may not be a simple matter.

If Kant is 'the most adequate of the Social Contract theorists' this may ironically be because he abandons the idea of the Social Contract as some sort of agreement or contract, actual or hypothetical, and thinks of it simply as formulating the necessary conditions for the possibility of universal consent to a political order for unsociable yet interacting rational beings.

However, although Kant insists that the people have inalienable rights, these rights do not include a right to rebel against rulers who violate their rights, 'a coercive right against the one who has done him injustice'. In many (arguably not in all) his discussions of obedience Kant argues that the only legitimate action of citizens in the face of injustice is to make representations: 'the freedom of the pen is the only safeguard of the rights of the people' (*TP* 8:304; cf. *WE*).