modern contractarianism. Contractarianism is the political philosophy that government is or ought to be created or empowered in some sense by the consent of the citizens, who agree to cooperate in certain ways to enable the government. There are many variant visions of just what level of agreement is required. It can range from actual to tacit to hypothetical to, most recently, so-called reasonable agreement. David Hume (1748) famously ridiculed the doctrine of contractarianism by pointing out that no then-extant state had been created by agreement and that yet many of them seemed, in some plausible sense, morally acceptable.

No leading contractarian today holds that there must be actual agreement. The focus of contemporary contractarians is on what citizens, under certain, sometimes quite abstract, circumstances, would supposedly consent to or ought, either rationally or morally, to consent to. Interpretation of these abstract circumstances to bring them down to ground enough to allow criticism or judgment of actual states is commonly beyond the enterprise of contractarian philosophy. Thomas Hobbes (1651) argued that citizens ought rationally comply, in their own self interest, with an extant state, even a relatively harsh state. Succeeding generations of contractarians have generally abandoned this rational criterion and have grounded their arguments in moral claims or in metaphysical, as opposed to self-interest, claims of what is rational.

The urgency of the focus on consent seems, except in Hobbes, to depend on finding a ground for making citizens obligated to obey government within certain limits. For Hobbes it was sufficient to oblige citizens with the threat of coercion. However, he often used the word obligation when his modern sense was to be obliged by force. Indeed, he said there are ‘two species of Natural obligation’: being obligated to submit to laws of physics and to fear of the stronger (Hobbes [1642] 1983: ch. 15, f7). Such slippage in the meaning of obligation has been a frequent problem in political analysis, especially in contractarian theory. In particular, scholars often moralize Hobbes’s contractarian arguments.

The move to grounding obedience in normative obligation that is derived from consent is the familiar moral move of grounding an obligation to fulfil a promise in the fact of one’s having consensually promised. What is at issue in promise keeping is usually an exchange (see generally, Hardin 1988). I promise to do something for you in return for your first doing something for me. There is less agreement, and a ridiculously large philosophical literature, on the strength of the obligation, if any, to fulfil a gratuitous promise, such as the notorious deathbed promise, especially if the person who has been promised something has incurred no costs in anticipation of fulfillment. In the law of some nations, gratuitous promises are not legally enforceable. In political contractarianism, however, the structure of the consent or the promise is potentially confused. Do I promise my government to obey? Or do I promise my fellow citizens? If government is not a moral agent analogous to a person, then the bite of the promise must depend on its being a promise to my fellow citizens. What then do my fellow citizens exchange with me? Their own forbearance in obedience to our government.

For Hobbes, our coordination on our government gives it the power to maintain order in our interest. Hence, he has no need of consent. John Locke (1690) argued that anyone who accepts benefits from the state has tacitly agreed to its rule. This view seems grossly simplistic. For example, it seems to entail that the trapped slave who accepts the benefit of police protection therefore agrees to the regime that supports slavery. Many contemporary writers evidently hold that hypothetical consent, but not tacit consent, is morally compelling: If I would have agreed, I can be treated as though I actually did agree. However, the emphasis in contemporary political philosophy is not on the obligation of citizens to obey but rather on the justification of a political order. For such justification, the philosopher Thomas Scanlon (1982) argues that agreement need merely be reasonable, that one must accept a government that one could not reasonably reject. Unfortunately, this move seems merely to shift the burden of understanding from ‘agreement’ to ‘reasonable’ and plausibly to undercut the original point of contractarianism. Instead of grounding obligation to obey government in consent we ground it in what the philosopher asserts it would have been reasonable to agree to.

The normative doctrine of contractarianism apparently has a very strange pedigree. It came into vogue with the death of theological grounds for claiming citizens must behave in certain ways. One might call theological politics a matter of obligation to god or an obligation to follow god’s will. But, given the power of god to demolish those who disobey, this sounds rather more like being obliged by force than like being obligated by the rightness of god’s demands. Hence, the duty of obedience to government under theological accounts was a matter of the citizen-believer’s self interest. With god no longer in the theory of the state, the problem of invoking a duty of obedience became acute. One solution to that problem was simply to suppose citizens actually have a moral duty that they
themselves voluntarily take on through agreement, essentially through contracting. This was a very odd move. It translated the self-interest notion of political duty under theological theories to a purely normative notion of moral obligation.

This move left a major difficulty in political theory: How do we explain actual behaviour of citizens if their only incentive for obedience is the normative incentive of having promised to do what, in the event, often turns out to be against their interests? Under god’s indirect rule, at least in many theories, interest and obedience were perfectly coordinated. Under normative consent theories, they often are contrary. This grievous difficulty has never been overcome in any contractarian theory. There are essentially three moves that have commonly been made to address the apparent difficulty. One of these is simply to assert that obedience to the state is a normative matter, that it is simply right. This move runs against the grim fact that obedience to many states cannot plausibly have been good and cannot plausibly therefore be right. Hence, we require specific argument for why this particular state merits obedience.

The second move is the quasi-logical move of Kantians and others to assert that, among other things, failing to keep a promise is inconsistent and incoherent. This transcendental move simply makes no sense to many theorists but seems compelling to others.

Finally, it is supposed by John Rawls and others that humans in a just society will become just, so that such a society will actually be supported by moral motivations. There might be difficulties in making the transition to such a just society but, if Rawls’s psychological argument is roughly right, there is no incoherence in supposing there could be one at least in principle.

THE BACKGROUND: HOBBES, LOCKE, ROUSSEAU, KANT. Although there were forerunners, the most important early contractarian was Hobbes. In Leviathan (1651) he proposed that there are two ways that government could be created to resolve the problem of social order in a particular community. That problem is acute because, in the so-called state of nature, Hobbes supposed life without government must be nasty, brutish, and short. Each person would independently attempt to defend against attacks from others by, often, attacking pre-emptively. This would be true even if most people were not vicious and were not inclined to steal from others. Hobbes supposed that most people are essentially self-seeking and centrally concerned with their own welfare. Hence, he posed solutions to the problem of disorder that depend on the rational, self-interested actions of people.

Order among such people could come about in two ways. First, members of an anarchic community could agree with each other to establish government as if by contract. Hobbes was writing before the heyday of contract law, but modern readers commonly interpret his argument as an analogue of the self-interested signing of a contract of exchange. And second, an extant government of some other power or a usurper could impose order. Hobbes called these two resolutions government by institution and by imposition, respectively. The actual problem that motivated him was the turmoil and disorder of the mid-seventeenth-century England in which he lived. For that problem his actual complaint was that deliberately attacking an extant government was self-destructive because it could lead to the collapse of that government and hence to disorder, under which virtually all would suffer great losses.

It is clear that the central problem for Hobbes was that order be established and maintained. He did not expect to achieve that with moral commitments but only with rational incentives for obedient behaviour. What was needed, he therefore thought, was merely coordination on a powerful sovereign or government that could be relied upon to suppress disorderly activities. In the absence of the threat of disorder and theft, individuals could then go on to prosper in their own ways. Oddly, although he is commonly regarded as a, or even the, major contract theorist, contract was at most only a device for reaching coordination on a sovereign and it was not the only device. Indeed, Hobbes himself thought it might never actually have been used. Yet he nevertheless thought that the extant, non-contractarian governments of the world, which at some point in their histories had been imposed rather than contractually established, were governments that should generally be obeyed.

Locke (1690) had a more sanguine vision of the state of nature as a time in which people could come together and agree on a government. Moreover, he was more inclined than Hobbes to argue as though there really had been such an agreement. (He was presumably the main object of Hume’s [1748] criticism of contract theorists.) But he saw clearly that there is still a problem of the consent of later generations who were not parties to any supposed past contract. He resolved this problem with his doctrine of tacit consent from accepting advantages from a government. His vision seems clearly to be one of moral, and not only rationally self-interested, commitment to obey government. That moral vision has arguably dominated contract thinking since his time. That is a somewhat odd twist on later visions of contract in the law, which is often seen as rationally rather than morally grounded. Worse, it raises the problem that it does not likely motivate many people in actual states. It seems therefore to ground state policies to coerce those who do disobey. For a theory whose proponents often claim grounding in a vision of autonomous free agents, this is a disconcerting twist.

If Locke saw clearly the problem of the obligations of future generations to a contract they did not sign, as though you and I were obligated for the contractual debts of our distant forebears, Jean-Jacques Rousseau (1762) saw, with Hume, the implausibility that any real society of any complexity could have unanimously agreed to a government. Because he nevertheless wanted to motivate obedience to the state with the moral obligation to live up to one’s consent, he invented the notion of the General Will. Even many of those who are ardent exponents of Rousseau’s thought have difficulty putting this notion into simple, comprehensible words. It is about the abstract idea that we do want order. The order Rousseau wanted clearly goes beyond Hobbes’s relatively minimal demand for order to include strong positive commitments to the community.
On such a notion, Rousseau supposed, we could unanimously agree. Thereafter, we could make specific decisions with less than unanimity. Locke similarly saw the need for unanimity in his moral vision. But he restricted the demand to the stage of agreeing to reach a decision on the form of government. That is, we have to agree to enter a room together in order then to design a government. But once in the room, we can vote by majority rule on actual elements of the design. Rousseau seems to have demanded much more than this, but it is very hard to say just what he wanted.

Immanuel Kant (1785) resolved this problem in Rousseau’s vision with his so-called transcendental argument about what any rational agent could will. Kant supposed that the moral law could only be grounded in principles that any rational agent could will to be universal laws in the sense that they would apply to all rational agents equally. Hence, I could not will a law that would make an exception for me. He supposed further that all rational agents would reach the same conclusions about whether any proposed principle could be a moral law. If this were true, then we could plausibly come together and will at least the elements of the principles that would underlie a morally acceptable government. Hence, we could contract on the uniquely correct form of government. That government would be republican in the sense that it would represent us all. It would not, however, be democratic because a merely democratic government could allow a majority to legislate against a minority. It could therefore be a government of interests and not a government of right.

Both Rousseau’s and Kant’s contractarian arguments might seem to ignore a problem that becomes immediately clear in a strategic or game theoretic understanding of the problem of contracting for social order. We could all agree in the abstract that we do want order. But it does not follow that we could ever agree on the form that the order would take. Hobbes saw this grievous difficulty in his contractarian account but could ignore it as irrelevant in his account of why we should not rebel against any established government. We should not rebel for the simple reason that the costs to us of moving to a form of government more beneficial to us would outweigh the gains we would get from the change. This is, of course, a social scientific and not a merely logical or conceptual claim and it might well be false in many instances, as for example in Czechoslovakia in 1989. It seems also to have been arguably true in some instances, as for example in the US civil war. A Kantian account could perhaps escape this difficulty if it were true that there is only one rationally determinable best form of government. But this seems to be utterly implausible. Indeed, contingent considerations generally seem to argue against a purely rationalist resolution of the problem of social order.

CONSENT. The moral core of contractarian arguments is that we obligate ourselves by consent or that our order is justified by consent. Consider how this core fits with the various notions of consent in which contractarian theories have been grounded: actual, tacit, and hypothetical consent. Let us first address actual and tacit consent and then turn in the next section to hypothetical consent, especially in its contemporary variants. It is the failure or implausibility of arguments from actual and tacit consent that sets up the contemporary arguments.

A commonplace claim is that there have been some cases of very nearly genuine contractual agreement on social order when peoples have, through representatives, written and adopted constitutions. If this were true, then we could suppose that even actual-agreement contractarianism has a field of play. The outstanding early case of such a supposed social contract is the US constitution of 1787–88. The adoption of that constitution, however, lacked the essential features of a contractual agreement (Hardin, forthcoming). First, perhaps most obviously, it lacked an external enforcement agency so that it was binding, if at all, only morally or from endogenous constraints. Second, it also lacked the unanimous consent of an ordinary contract. Some state conventions voted only narrowly for it and two states at first voted against it. Prior to that, some states and some important interests were represented not at all or only briefly or poorly at the constitutional convention. Slaves and women were not represented at all; small farmers were represented if at all only indirectly, although they were the great bulk of the white male population; and the eventual anti-Federalists were poorly represented because many of them refused to attend or stay at a convention with an apparently illegal purpose. Both the Philadelphia constitutional and the subsequent state ratifying conventions were meetings of the elite. Hobbes (1651: 203) asserted, and most writers on promise-keeping would agree, that ‘no man is obliged by a covenant of which he was not the author’. Third, it set up future changes in the relations between the citizens who were party to it, changes that would not require even the degree of consent that adoption of the constitution itself required. Moreover, it continues to constrain generations who never agreed to it.

The most critical difference between that constitutional experience and an ordinary contract is that the actual problem of the creation of a new government lacked the strategic structure of such a contract. An ordinary contract governs an exchange. I do something for you – I build you a house – in return for your doing something for me – you pay me. It would, of course, be in my interest to have you do your something for me while I do nothing for you so that I walk away with your payment without having the burden of building your house. In the constitution of 1787–88, however, the people of New York who favoured the constitution could not have said it was their interest to have the people of Virginia adopt the constitution while they did not. The constitutional order could be of value to either only if both coordinated on it. The value of the constitution came from enabling Virginians, New Yorkers, and others to trade more easily, to interdict beggar-thy-neighbor foreign trade policies, and to defend themselves against foreign attack. The constitution resolved a massive problem of coordination on order for such purposes. It enabled the future resolution of countless exchanges by setting up an enforcer – government – that was external to the exchanges. But it was not an act of exchange itself.

An implication of its being a coordination rather than a contract is that the enforcement of the constitutional order
could readily be endogenous—it would not require an external enforcing agency. We follow the constitutional order because the costs of recoordination on an alternative make it not our interest to recoordinate. This might follow even when the new coordination would be substantially superior to the one it replaced. This is a very grand and complex instance of the general nature of coordination. But it has the central features of simpler coordination problems. For example, the convention that we all drive on the right or that we all drive on the left coordinates us all to our mutual advantage. Olaf Palme decided that Sweden's convention of driving on the left was too costly because it meant accidents on the open road with, especially, German tourists who were accustomed to the opposite convention. At substantial cost, the Swedish government changed its driving convention in 1967 (Hardin 1988). There was a similar concern in the United Kingdom, but there the costs of a recoordination seemed plausibly to outweigh the benefits of making it. Within either convention, however, there is little need for an external enforcing agency. Drivers are compelled by the convention itself to follow the convention in their own interest. Enforcement of acquiescence is endogenous to the system.

One response to these seemingly non-contractarian aspects of the US constitution is the claim that contract simply does not mean what it once did in the early contract tradition, and modern contracts are often like constitutions. Unfortunately, this response guts any claim that I am morally bound to the constitutional order because I consented to it or that the order is justified by my consent.

Alternatively, it can be argued, as Locke already did argue, that I have tacitly consented to that order by taking advantage of the benefits it offers me. This is an odd claim. It seems to imply that even the revolutionary who detests and wishes to destroy the constitutional order but who, on the way to doing that, must use weapons created under that order is morally bound to comply with the order. Hence, actual agreement and tacit agreement are implausible candidates for moralizing social order under a supposed social contract.

The crux of our problem is the moral nature of consent. We might suppose that consent is itself right-making, or we might suppose consent is merely an indicator of what it would be right or good to do. If we have the latter view, we might go further and suppose consent is morally binding: I morally must do what I have consented to do. Recent contractarians reject actual-agreement and tacit-agreement contractarianism. They want it to be true in some sense that what we would consent to under certain circumstances is right. The phrase 'under certain circumstances' suggests, rightly, that their concern is with hypothetical agreement. They also sometimes seem to want to hold that our consent binds us, although it is unclear why there is any further need for binding us morally in those theories in which we are already rationally bound. Turn then to the complex contemporary views.

CONTEMPORARY CONTRACTARIANS: RAWLS, BUCHANAN, SCANLON. A striking difference between the Hobbesian and Lockean views is the following. Hobbes supposed that moral obligations of many social kinds, such as the obligation to abide by ordinary contracts, is derivative from the prior existence of a powerful sovereign who can enforce such obligations. Locke implicitly supposed on the contrary that morality is prior to politics and that we can derive the moral status of a government from the prior morality of consent to it. For Locke, it would be wrong to break a promise even in the state of nature. For Hobbes, nothing is wrong in the state of nature if it conduces to survival. It is an odd quality of most contemporary contractarian thinking that it follows Locke in this respect. Much of it does so, however, not so clearly from Locke's influence as perhaps from Kant's (1785). (Contemporary contractarianism therefore has a metaphysical ring to it that irritates many of its critics, although the contractarians themselves are often irritated in turn by the accusation that they practise dreaded metaphysics.)

Buchanan (1975) argues that, in constitutional debate, we must reach unanimous agreement on principles for government. Hence, we start from here and now. That is to say, everything is not up for reconsideration de novo. The current distribution of property constrains what can be done hereafter to the extent that constitutionally determined expropriation would not receive unanimous consent. This makes Buchanan the theorist who is nearest to being an actual-agreement contractarian, although he writes of hypothetical agreement.

Rawls (1971) argues that we should seek principles of justice that a representative person could agree to from behind a veil of ignorance. That is to say, we should seek principles that we think are generally compelling independently of who we are, and not principles that would contingently happen to serve particular interests. Hence, Rawls can be said to be a hypothetical-agreement contractarian. However, there is an odd quality to his position that seems very uncontractarian. One person alone, a representative person, should be able to decide on the principles of justice for our society. If this is true, then presumably just any such person would reach the same conclusion on what those principles should be. And indeed, Rawls proposes what those principles are. But if this is possible, then the principles sound suspiciously rationalist. They are deducible from more or less pure reason, as Kant's moral principles are. Agreement with them is more nearly like agreement with laws of physics than like the kind of agreement we might make on some mutual accommodation. After being criticized on this point, Rawls (1985) came to defend his views with specific reference to Kant's rationalist theory, although he also has strongly defended his vision against dismissive claims that it is metaphysical.

Against Rawls's argument from behind the veil, or from what he calls the original position, which is partly defined by ignorance, note that John Harsanyi (1955) had argued nearly two decades before Rawls that we should establish social and moral principles from a supposition of individuals' ignorance of the advantages that would flow specifically to them from our chosen rules. His conclusion was that we would choose strictly utilitarian distributional principles because under these our expected individual benefits would be maximized. Rawls blocks this conclusion by
stipulating an unusually strong principle of risk aversion that many scholars find objectionable and even perverse. Even with his principle of risk aversion, however, Rawls's derivation seems likely to permit a somewhat distorted utilitarian maximization, although this is a hotly contested claim. Rawls's defence of his principle of fairness, on which his theory is based, depends heavily on his assumption of risk aversion. It seems likely that any effort to obtain principles of social justice from rationalist deduction must start from strong constraining assumptions, for example, either of Rawls's risk aversion or of Harsanyi's risk neutrality. Otherwise, we cannot adequately narrow the range of principles from which we wish to choose. They must seemingly be narrowed to a set that includes only one, unique, principle. But this means that our initial assumptions, which are themselves in no sense contractarian, heavily determine our contractarian conclusions.

Scanlon (1982) in essence addresses this problem of a purely rationalist deduction. He allows what might be a substantial multiplicity of principles or of distributions. Rather than stipulate what our principle must be, he stipulates that it must be one which we could not reasonably reject. This definition packs, of course, a lot of its concern into the notion of reasonable. Presumably, a fairly wide array of principles could be reasonable. If we actually have a social order that meets Scanlon's criterion of reasonableness, then it follows that we should not attempt to change it to meet another order that also meets the criterion. Hence, Scanlon's reasonableness is a more restrictive variant of Hobbes's conclusion that we should support any government that successfully maintains order. If our social order does not meet Scanlon's criterion, the criterion yields some constraint on what reforms we could seek, but it might not restrict our choices enough to make consensus possible. Hobbes thought that three quite different forms of government could be rationally supportable: monarchy, oligarchy, and democracy. But he also thought monarchies best, so that, if we were without government, we should plump for monarchy. His preference for monarchy turned, of course, on social scientific claims, not on analytical or normative claims, although he stated is as though it were a conceptual matter. Scanlon's reasonableness might yield no definitive conclusion without also invoking social scientific claims to narrow our range of choice.

Motivating moral principles. Most of the contractarians, classical and contemporary, have been driven by one of two main motivating moral principles: mutual advantage and fairness. The theories of Hobbes and Locke seem straightforwardly to be driven by a concern with mutual advantage. Hobbes openly stated this concern, although he granted that some individuals, such as religious fanatics and glory seeking aristocrats, might not see any advantage for themselves in an orderly society. Locke's concern with consent captures his concern with mutual advantage. One might read Rousseau as also concerned with mutual advantage. But it would be difficult to read Kant's contractarianism as essentially responsive to mutual advantage even though it might suggest an order that would be mutually advantageous. One might claim that his concern is some version of fairness, although it is not easy to articulate such a principle in his terms. Among contemporary contractarians, Buchanan is most emphatically concerned with mutual advantage. In his view, we start from here and we improve things for everyone — or at least we do not harm anyone.

Rawls greatly complicates the notion of distributive justice by grounding it in an odd combination of mutual advantage and fairness. Pure fairness might seem to require equality of distribution. What we distribute is, of course, what we have first produced plus whatever might be available to us without effort. But suppose we could produce so much more by distributing unequally that we could then make everyone, including those who would be worst off, better off than anyone would be under the perfectly equal distribution. As a rule, most modern economists suppose this to be the nature of our problem of production and Rawls grants that they might be right. He supposes that a rational person behind the veil would choose to be better off at the cost of some inequality rather than to be worse off in a state of equality. He therefore allows inequalities that make the worst off class of citizens best off. This is his difference principle. One might conclude, with typical economists, that distributive justice (taken as equality) trades off with productive efficiency. Rawls supposes, however, that distributive justice includes both concerns, equality and efficiency, and that the trade-off between them is internal to the conception of distributive justice.

It is a distinguishing central insight of Rawls that fairness cannot stand alone as a theory of the good or the right because it is fully consistent with egalitarian misery. Hence, inequality that produces greater overall welfare can trump pure equality. Those who dislike this assertion sometimes argue against it with an implicit dismissal of the possibility that inequality could have this effect or with claims that all we need is to correct aberrant psychologies to get people to produce for the general good rather than merely for selfish benefit.

In his theory, Rawls is concerned with fairness of an odd kind. Fairness in the allocation of the joint social product is at issue just because that product is joint. I might claim to 'deserve' a fair part of it because I contributed to its production. If market wages and profits really mirrored contributions, allocation according to desert would reduce to little more than what markets do. Rawls clearly thinks market wages and profits do not mirror contributions. We cannot causally relate your effort strictly to your share of the social product. Your wages are related to the supply and demand functions for your talents, rather than directly to your output. What you produce may stay constant while your wage changes, or vice versa. Rawls's theory of justice may therefore entail a distribution other than what the market would produce.

Yet, his theory does not simply correct for any supposed distortions that supply and demand might impose on your wage. Nor does Rawls want a simple desert model in which you get what you deserve as a result of your effort or whatever. Rawls does not tie your desert to what you produce. He ties it merely to the fact that you produce. The fact that you produce gives you a claim on a share of the joint product. The capitalist or well paid professional may think
she deserves what she gets from market relations. A Rawl-
sian and virtually any contemporary economist would
reject such desert claims and would say that the capitalist
or professional is merely in part very lucky to be in the
right place at the right time. Bill Gates is not in any plau-
sibly meaningful sense worth the four billion dollars a year
that he has recently been making, but that is what our
somewhat distorted market yields him. Still, Rawls's
theory retains an odd tie to desert or entitlement. Since
your entitlement depends on your producing something,
your entitlement is nothing if you cannot contribute any-
thing to the joint product. Hence, the severely disabled
and plausibly the citizens of other nations are not entitled
to a share of our nation's product, although we might
altruistically choose to give them something. Brian Barry
(1995) breaks this minimal tie to desert with his theory of
impartiality. Those who like desert talk might say he wants
to base desert in simple humanity, not in any specific
capacity or accomplishment.

CONCLUDING REMARKS. Contractarians are commonly
concerned to justify government, especially its coercive
power. We would not need a normative theory to explain
the success of government. The modern equivalent of the
theological theory would suffice: When we typically obey
government we can see that it is in our interest to do so. In
large part because this is in fact true, moral justification
may play little or no role in explanation. Indeed, given that
the contractarian moral justifications are the relatively
arcane and hotly argued theories of, mostly, philosophers,
it would be surprising if they were the reasons actual
people do obey governments.

It is one of the sometimes implicit claims of self-interest
theories that they can explain even what the actors appar-
ently do not understand. So for example, many people
readily assert that it is in individuals' interest to contribute
to public provisions of various kinds. Yet, rational choice
theory suggests that it is not in their interest individually
to do so and they commonly do not. Hence, people seem to
be motivated by interests that they do not understand and
that they often even deny. Perhaps one could suppose that
a hypothetical agreement is in some sense similarly binding
even on those who do not understand the issue. However,
the facts in the rational choice case are that behaviour often
fits the theory rather than the understanding.

Do the facts of behaviour fit the contractarian pro-
gramme? This is not generally a question that has been
addressed. For a social order to work seems to require
more that people acquiesce in the order than that they
consent to it. This is a fundamentally Hobbesian conclu-
sion that runs against the tradition of the social contract
theory of which he is often thought to be the father.
Hobbes supposed that acquiescence follows, of course,
from interests rather than from moral commitment.
Therefore he was happy with government by imposition,
as when a usurper seizes power or when a foreign power
invades and conquers, and he did not require contractual
agreement either for the establishment or the maintenance
of order.

This still leaves the questions whether government
would work better if it is consensual and whether it would
be more consensual if it is fair. Surely the answer to the
first question is typically that it would, although consen-
sual support of government can depend on many things
other than its fairness. For example, nationalist hysteria
can produce consensual support. The answer to the second
question might seem self evident to some, but there is
remarkably little evidence, although there are some very
modest results that show that fairness in the system of
justice leads to greater acceptance of its decisions (Tyler
1990). In any case, consensus per se is not requisite for fair-
ness nor fairness for consensus. Perhaps it should therefore
not be surprising that, even theoretically, contractarianism,
which originally began as a consent theory, has become
rationalist in its modern guises. Talk of consent in such
theory today is little more than metaphorical.

Given the virtually total unworkability of contractual
creation or justification of actual governments, one might
suppose that moral theorists concerned with government
would focus on how to make governments good in various
ways. At its best, social contract theory might be seen as an
ideal theory that commends standards to which we should
aspire. It might also commend reforms to which we should
be committed. But it is not easy to make a convincing argu-
ment that what we should first focus on is reforms that
enhance consent rather than reforms that enhance govern-
mental performance. (The move to broader consent by a
minority dictatorial government in Burundi led recently to
a bloodbath.) This is especially true if the consent that is to
be enhanced is hypothetical consent. Moreover, one might
suppose that enhancing political equality requires enhanc-
ing educational and economic opportunities, which might
well be prior to making consent workable.

RUSSELL HARDIN

See also BUCHANAN, JAMES M.; HOBBES AND
CONTRACTARIANISM; HUME, DAVID; JUSTICE, LOCKE, JOHN;
MODERN UTILITARIANISM; STATE OF NATURE AND CIVIL
SOCIETY.

Subject classification: 3a.

BIBLIOGRAPHY
Press.
Buchanan, J.M. 1975. The Limits of Liberty. Chicago: University of
Chicago Press.
Hardin, R. 1988. Morality within the Limits of Reason. Chicago:
University of Chicago Press.
Hardin, R. (forthcoming) Liberalism, Constitutionalism, and
Harsanyi, J.C. 1955. Cardinal welfare, individualistic ethics, and
interpersonal comparisons of utility. Journal of Political Economy
1983.
Hume, D. 1748. Of the original contract. In David Hume, Essays
Moral, Political, and Literary, ed. E. Miller, Indianapolis: Liberty
Kant, I. 1785. Grounding for the Metaphysics of Morals. Indianapolis:
Hackett, 1981.
modern utilitarianism. Utilitarianism is the view that one should do whatever will bring about the greatest amount of good. It was first clearly propounded in the eighteenth century by the philosopher Jeremy Bentham (1789). Leading figures in its subsequent development were John Stuart Mill (1863) and Henry Sidgwick (1874), both philosophers with a strong interest in economics. Throughout its history, economists have had a strong influence on the development of utilitarian thinking. Recently, work by the economist John Harsanyi (1953, 1955, 1977a) has been particularly influential. Important recent writings include Griffin (1986), the debate contained in Smart and Williams (1973) and the collection in Sen and Williams (1982).

The broad and imprecise statement that one should do whatever will bring about the greatest amount of good leaves plenty of scope for differing opinions among utilitarians, and there is no more precise formulation of their doctrine that all utilitarians would accept. Indeed, some would not accept even this broad statement. But it will be helpful to have a more detailed formulation as a benchmark for comparing alternative versions of utilitarianism. One formulation is the conjunction of the following principles.

Consequentialism. Of the acts that are available, one should do the one that will have the best consequences — that is to say, the one that will bring about the best state of affairs.

Personal good. The goodness of a state of affairs is its goodness for people.

Additivity. The goodness of a state of affairs for people is the total wellbeing of individual people.

Hedonism. A person’s wellbeing is the preponderance of pleasure over pain in her life.

This is now an old-fashioned sort of utilitarianism, and few modern authors would accept all these principles without reservation. But they provide a convenient scheme for classifying the points that are at issue in modern utilitarian thinking. This essay will take the principles of consequentialism, personal good, additivity and hedonism in turn, and discuss some of the questions raised by each. Many of the new ideas that have entered utilitarianism recently are responses to objections that were made against earlier versions of the doctrine. So this tour of the issues will also provide a picture of the present standing of utilitarianism, and its unsolved problems.

CONSEQUENTIALISM: DIRECT AND INDIRECT UTILITARIANISM. ‘Indirect’ utilitarianism has become popular in recent decades. The most familiar version of it is rule utilitarianism (see Harsanyi 1977b; Hooker 1996), but other versions are possible. For instance, there is character utilitarianism and virtue utilitarianism. All versions accept the principles of ordinary, direct utilitarianism, except that they deny consequentialism: they deny you should do the act that will bring about the best consequences. Instead, they first of all make a claim about which rules you should live by, or what sort of character you should have, or what virtues you should espouse: you should live by the rules that will have the best consequences — that which will promote total wellbeing — or have the character that will have the best consequences, or espouse the virtues that will have the best consequences. Then these theories say you should do whatever act is prescribed by the rules you should live by, or issues from the character you should have, or from the virtues you should espouse.

The attraction of indirect utilitarianism is that it seems to offer a way of overcoming some of the objections that are commonly levelled against the direct version. Here are four of these objections and the indirect utilitarian’s replies.

The first objection is that some implications of utilitarianism are intuitively simply wrong. For example, suppose the police are holding a person whom they know to be innocent, but whom the public believes to be guilty of a terrible crime. Suppose there will be riots if this person is not punished, leading to many deaths and great destruction. Direct utilitarianism will dictate that the innocent person should be punished, to avert the greater harm that will be caused by the riots. But intuition suggests it is wrong to punish an innocent person, even if there will be beneficial consequences, because it is unjust. On the other hand, a rule utilitarian can argue that on balance it best promotes total wellbeing to adopt rules of justice rather than rules of expediency, even if this occasionally leads to destructive riots. A just society is happier than an unjust one. So rule utilitarianism can support rules of justice, and agree with the intuition that it is wrong to punish the innocent person on this occasion.

A second objection is that even by a utilitarian’s own criterion, people must live by some rules. Otherwise society would collapse, and that would be bad for total wellbeing. For instance, people must normally keep promises. If they did not, no one would be able to trust anyone else, and most social interaction would become impossible. Direct utilitarianism says one should keep a promise only if keeping it will contribute more to total wellbeing than breaking it will. But if you break a promise whenever it turns out better for the total wellbeing to do so, your promises cannot be relied on. If people commonly followed the prescription of direct utilitarianism, the whole institution of promising would fail, and this would be very bad for total wellbeing. On the other hand, adopting the rule of keeping your promises (except perhaps when the consequences will be extremely bad) may well have better consequences than adopting some other rule about promises. If so, rule utilitarianism will tell you to adopt this rule, and keep your promises, and that will be good for total wellbeing.