The Theory of Institutional Design

Edited by
ROBERT E. GOODIN
Research School of Social Sciences
Australian National University

CAMBRIDGE UNIVERSITY PRESS
IN A FAMOUS PASSAGE, Kant wrote: "Enlightenment is man's emergence from his self-incurred immaturity. Immaturity is the inability to use one's own understanding without the guidance of another." He added: "Have courage to use your own understanding!" (Kant 1784, p. 54). Though Kant himself was at pains to deny it, this injunction is an enormously subversive political ideal. The most characteristic demand of political leaders has invariably been that their subjects submit their own understanding to the guidance of authorities, either because the authorities are wiser or because an ordered society requires artificial unanimity.

Kant's injunction presumes that the understanding of ordinary citizens is up to the task of deliberating and reflecting on political affairs without the guidance of others; this assumption amounts to the Enlightenment's article of faith. It is profoundly controversial. Politicians often operate according to the self-serving assumption that ordinary citizens must be pacified with malarkey, for they cannot face hard truths. They believe, in the words of H. L. Mencken, that "bosh is the right medicine for boobs" (Mencken 1956, p. 43). This belief, prevalent in political debate and ubiquitous in political practice, is anathema to the Enlightenment's article of faith.

The Enlightenment thus rejects an older view of politics, going back to Plato, according to which government necessarily relies on noble lies – myths or deceptions designed to secure loyalty and love of country.

I delivered drafts of this chapter to seminars at the University of Wisconsin, the University of Maryland, Monash University, the Australian National University, New York University, the University of Iowa, and Wake Forest University. I am very grateful for the many helpful comments and criticisms I received from participants in those seminars.
Machiavelli, like Plato (though for rather different reasons), argued that lies and secrecy are essential instruments of successful government. A successful prince, Machiavelli says, must learn how not to be good; he must accept that lies, like betrayal and violence, are necessary tools of government. (A backhanded contemporary acknowledgment of this is an anecdote related by the late Louisiana Senator Russell Long. When Long was in secondary school, he approached his uncle Earl, then the governor of Louisiana, and said that he had been assigned to debate the question of whether one should use truth in politics. What should he say? Earl asked which side Russell had been assigned. When Russell replied that he was to debate the affirmative, Earl thought for a moment, then said, "Hell yes! In politics you use anything you can get your hands on.")

The present essay examines one strand of the debate between open and closed government – between the Enlightenment ideal and the arguments represented by Plato and Machiavelli. The examination itself will not move on so enormous a scale. Rather, my method will be to approach the central issues by examining a single principle of political morality that derives from the Enlightenment ideal.

In the second appendix to his essay "Perpetual Peace," Kant proposes the following "transcendental formula of public law":

All actions relating to the right of other human beings are wrong if their maxim is incompatible with publicity.\(^1\)

The connection between this formula and the Enlightenment's article of faith is relatively straightforward. Publicity will enable citizens to submit "actions relating to the right of other human beings" – public policies, we may call them – to the scrutiny of their own understanding. If the Enlightenment's article of faith is correct, such public debate and

\(^1\)"Alle auf das Recht anderer Menschen bezogene Handlungen, deren Maxime sich nicht mit der Publizität verträgt, sind unrecht" (Kant 1795c, p. 381). Normally I shall use the English editions of Beck and Reiss (Kant 1795b, a). But here I am unsatisfied with Beck's translation, which reads, "All actions relating to the right of other men are unjust if their maxim is not consistent with publicity" (Kant 1795b, p. 129). Beck's "not consistent" makes the disharmony between the maxim and publicity a matter of logical or practical contradiction. But Kant's "sich nicht verträgt" implies only incompatibility, not contradiction. As we shall see in section 6.2.3, it is an important interpretive question whether Kant thinks his publicity test requires a demonstration of practical contradiction or merely contingent incongruence. I want to avoid begging this question through the translation. Reiss does better: "All actions affecting the rights of other human beings are wrong if their maxim is not compatible with their being made public" (Kant 1795a, p. 126). But he pluralizes "das Recht"; as in English, however, talk of "right" is more general (and more ambiguous) than talk of "rights."
scrutiny are highly desirable, and the inability of policies to withstand publicity is suspicious. If, on the other hand, ordinary citizens' understanding cannot be trusted, then good government, and indeed justice itself, will require noble lies. In that case Kant's formula will be false.

Kant believed that his "transcendental formula of public law" – the publicity principle, as I shall call it – would provide an easy-to-use test for the moral rightness of political action, an "experiment of pure reason" that we perform by asking of any political action (roughly), "Could I still get away with this if my action and my reason for doing it were publicly known?" If the answer is no, then the action is wrong. John Rawls, the most eminent contemporary Kantian political philosopher, adopts this "publicity condition" as a "formal constraint on the concept of right," and employs it in his argument in A Theory of Justice (Rawls 1971, pp. 130, 133).

Kant's test does not imply a moral requirement that every political action actually be publicized. It is a hypothetical publicity test: Kant speaks only of the "possibility" and "capacity" of a maxim's withstanding full public knowledge. His "experiment of pure reason" is a thought experiment.

It is quite possible, of course, that through thought alone we will be unable to answer the question of whether an action and its maxim could withstand publicity. Thus in many cases the thought experiment may yield no determinable result. Moreover, the test is purely negative: Kant is at some pains to emphasize that the bare fact that a maxim could withstand publicity does not by itself show that the maxim is right; the publicity principle says only that if the maxim could not withstand publicity it must be wrong.

Yet even within these limitations the publicity principle is a remarkably powerful proposition of political morality, and indeed of individual morality as well. As we shall see, on its face it rules out a variety of principles and policies, ranging from widely held views of the judicial process through the American Catholic Church's position on nuclear deterrence to utilitarian accounts of morality.

The publicity principle is not only a proposition of morality, moreover, but a principle of institutional design as well. Suppose that the

---

2The latter is the erroneous reading of Habermas, who glosses the test to demand that maxims of political action "were capable of, or indeed in need of, publicity" (Habermas 1989, p. 108).

3See Kant 1795c, pp. 381–82; Kant 1795a, p. 126; Kant 1795b, p. 130; likewise, Kant 1795c, pp. 384–85; Kant 1795a, p. 129; Kant 1795b, p. 133.

4Kant himself claims that his "transcendental formula" is both juridical and ethical, belonging to the doctrine of virtue as well as the doctrine of right. Kant 1795c, p. 381; Kant 1795a, p. 126; Kant 1795b, p. 129.
publicity principle were false; suppose, in other words, that on occasion just policies were incompatible with publicity. In that case, we would be forced to construct institutions capable of formulating and executing policies largely out of reach of public oversight, for fear that good government would be subverted by unwise or immoral public pressure. We would leave a good deal of latitude for official secrecy, and move very cautiously in constructing oversight and accountability institutions (and that fact might itself have to be kept confidential). If, on the other hand, the publicity principle is true, we would have an argument for augmenting public accountability and institutional openness: roughly, that the best way to make sure that officials formulate policies that could withstand publicity is by increasing the likelihood that policies will withstand publicity.

Kant’s formula, however, leaves many questions unanswered. What is meant by “publicity”? What is it to be incompatible with publicity? Does the publicity principle mean to rule out all state secrets, as it at first glance appears to? Most important of all, of course, is whether the publicity principle is true. What is the connection between publicity and morality? Is it merely the intuition that anything one can’t do openly has got to be wrong? I shall be arguing that the publicity principle is surprisingly difficult to defend; but, in the end, I shall offer a qualified and conditional defense (conditioned on the truth of some empirical conjectures that I am not in a position to defend but that strike me as more plausible than their denials).

6.1 The Publicity Principle in Action

My aim in this section is to exhibit the publicity principle in action— to show, by means of examples, that it is indeed a powerful tool for the moral criticism of public policy. What follows is an unordered catalogue of views and arguments that are facially ruled out by the publicity principle—“facially,” because subsequently we must see whether some of them may be rehabilitated. I choose examples from three arenas: the legal system, representing domestic issues; nuclear weapons policy, representing foreign policy and military issues; and moral theory.

6.1.1 Publicity and the Legal System

Should judges be candid about their reasons for deciding cases? Four Yale law professors, Charles Black, Alexander Bickel, Guido Calabresi, and Paul Gewirtz, have considered variations on an argument against judicial candor. For short, I shall call it the “Yale argument.”
Calabresi frames the argument clearly in a discussion of whether judges ought to strike, limit, or otherwise undermine anachronistic laws that the legislature has not gotten around to repealing:

[T]o admit this judicial power would be inevitably to admit its abuse; . . . we come closer to achieving the amount of judicial supervision we want by denying that it is permitted at all than we would by acknowledging what is going on and trying to control it by doctrine and language; . . . we decide better, in practice, by denying that it ever takes place at all. . . . [R]ecognition of the doctrine [of judicial revision of anachronistic laws] would have lawyers arguing its applicability where it should not apply. . . . (Calabresi 1982, pp. 174–75)

Charles Black argues similarly that a judge may want to insist on an absolute rule against torture while nevertheless permitting the torture of a terrorist who has hidden a hydrogen bomb set to go off in an hour and who will not otherwise reveal its location. The alternative to this judicial deception is admitting that the rule against torture may be balanced away, and that is too dangerous an admission. Analogously, he argues, for the late Supreme Court Justice Hugo Black's insistence that rights of free speech are absolute when Justice Black knew full well that they are not (Black 1961, discussed in Calabresi 1982, pp. 173–75). According to Charles Black, Justice Black was willing in his judicial practice to countenance the restriction of free speech rights. But he would not admit explicitly that free speech rights could be balanced away, for fear that if he did admit it lawyers would seize on his language to argue for restrictions in inappropriate cases.

The Yale argument likewise underlies Bickel's famous theory of the "passive virtues." The passive virtues consist in the judicious use of decision-dodging subterfuges by the Supreme Court to avoid placing its constitutional imprimatur on either side of certain politically divisive issues (Bickel 1962). According to Bickel, it would be catastrophic for the Court to admit that it wishes for political reasons to dodge constitutional issues, so instead the Court must appeal (disingenuously?) to technical doctrines of standing, ripeness, vagueness, abstention, and political questions which permit it to avoid reaching the merits of cases.

Gewirtz raises the Yale argument when he suggests that judges must often shape remedies with one eye directed toward the political resistance their actions will meet. The examples he has in mind are American federal court decisions ending school segregation, which the courts, fearful of white violence, coupled with remedies that ensured that segregation would persist for many years. The Yale argument points toward the conclusion that judges should deny that they are taking account of political resistance to their remedies, since to admit it would delegi-
mize the courts and create perverse incentives to lawlessness on the part of losing litigants (Gewirtz 1983, pp. 665–74).

I hasten to add that Calabresi and Gewirtz both reject the Yale argument, whereas Black and Bickel apparently accept it. But even Calabresi and Gewirtz reject it only on instrumental, cost-benefit grounds (weak ones at that, in my opinion). Gewirtz argues that “dishonesty always creates the risk of its detection, and, with detection, harm to the courts’ stature that may exceed any losses that result from candidly admitting limited power.”

Calabresi’s argument is a more guarded version of Gewirtz’s:

[W]e are not comparing a certain harm-benefit with an uncertain one. We do not know the ultimate dangers of . . . use of subterfuges. . . . The burden must be on those who would argue for indirection. The choice must be for candor. (Calabresi 1982, p. 177)

The publicity principle, on the other hand, yields an immediate, and noninstrumental, argument against judicial subterfuge. If judges cannot openly state their real reasons for a decision, the publicity principle tells us that the decision is illegitimate. If Justice Black “was sometimes forced to twist the notion of free speech in order to square absolute protection with his opinion that certain speech-related activities could be regulated” (Calabresi 1982, p. 296 n. 11), he did wrong. Either he should have opposed regulating the speech-related activities in question, or, if he was convinced that the regulations were valid, he should have abandoned his professions of absolutism.

Interestingly enough, Judge Richard Posner has proposed a “publicity test” in the context of judicial decision-making. According to Posner, “a decision is principled if and only if the ground of decision can be stated truthfully in a form the judge could publicly avow without inviting virtually universal condemnation by professional opinion” (Posner 1985, p. 215). Posner is a leading federal appellate judge as well as one of America’s most influential writers on jurisprudence, and his resurrection of Kant’s idea underlines its centrality to the whole issue of judicial candor.

From the issue of judicial candor we may pass to the question of how transparent the legal system must in general be. Paraphrasing Thurmond Arnold with approval, Karl Llewellyn once wrote concerning the trial process:

an impressive ceremonial has a value in making people feel that something is being done; this holds, whether the result is right or

wrong; and there is some value in an institution which makes men content with fate, whatever that fate may be. (Llewellyn 1940, p. 610)

In a different context, Charles Nesson has recently argued that some rules of evidence that seem irrational from a strictly scientific or epistemological point of view can nonetheless be justified because they are “ways to promote public acceptance of verdicts” (Nesson 1985, p. 1368). According to Nesson, “The aim of the factfinding process is not to generate mathematically ‘probable’ verdicts, but rather to generate acceptable ones; only an acceptable verdict will project the underlying legal rule to society and affirm the rule’s behavioral norm [i.e., convince people that they should obey the rule].”

Both Llewellyn and Nesson seem to be saying that public compliance and deference to legal rules are more important than the objective rightness or justness of those rules. This is a view of great antiquity; it assumes that order, compliance, and stability are more fundamental social values than justice. The problem, however, is that it is quite likely that people will willingly defer to legal rules only when they believe that those rules are just. (Studies of convict popfor example, have found that convicts, even those who freely acknowledge that they committed the crimes, will typically accept the legitimacy of their punishment only if they believe that they were fairly convicted.) Hence, people must be made to believe that legal institutions are genuine purveyors of justice whether or not that is so. Paradoxically, institutions of justice aim to stabilize society rather than to do justice. But if people believe that these institutions do not aim at justice, then they will not respect those institutions, and the institutions will break down—paradoxically, they will not stabilize society. Hence the view that institutions of justice exist to stabilize a society rather than to do justice must remain esoteric, and most people must instead be made to believe in the justice of these institutions. Plato held such a theory and therefore proposed (in the Laws at 716a–b) that a public celebration of justice be promoted as a kind of official religion, even though a few philosophers understand that justice is less important than order and that the religion is false. The religion of justice is a noble lie.

6Nesson 1985, p. 1359. Subsequently, however, it becomes clear that Nesson, unlike Llewellyn, is not defending acceptable-but-wrong verdicts. Rather, he is suggesting (a) that even a correct verdict will be acceptable only when it is taken to refer to the underlying event rather than the legal evidence, and (b) that a verdict may be acceptable-and-right even though its probability is not high. In my view, Nesson’s actual argument is better than the less careful dicta quoted here.


8For a justification of this reading of Plato, see Luban 1994, pp. 321–29.
Llewellyn and Nesson seem to endorse this Platonic view that order is more important than justice and that the knowledge that order is more important than justice must be restricted to a few illuminati. Indeed, however, the Platonic view has wide acceptance among commentators on the judicial process. The often times repeated argument that overt judicial lawmaking is dangerous because it will make the general public lose respect for the judiciary (even though it is acknowledged that, realistically, judges cannot avoid lawmaking) makes sense only as a plea for judges to keep the political aspects of their craft esoteric.

It need hardly be said that all such views are inconsistent with the publicity principle. What would happen if everyone knew and accepted Llewellyn's point that a trial, accurate or inaccurate, is intended principally as a ceremonial to reconcile us to our fates? Then the ceremonial would no longer reconcile us to our fates. Similarly, if people generally knew that evidence laws were intended to make them accept verdicts rather than to keep the verdicts honest, verdicts would be less acceptable. If it were publicly known that institutions of justice exist for the sake of stability rather than justice, that knowledge would destabilize those institutions. None of these policies, it would seem, can withstand publicity. Thus, they all fail Kant's test.

Another example of a legal institution that facially fails the publicity principle arises from Meir Dan-Cohen's account of "acoustic separation" in the criminal law. By this Dan-Cohen means that some legal doctrines and rules are directed primarily at officials, and are not meant to be "overheard" by the general public - hence the term "acoustic separation." Dan-Cohen suggests that in many cases acoustic separation is a desirable feature of the law" (Dan-Cohen 1984). For example, proponents of a defense of duress point to "the unfairness of punishing a person for succumbing to pressures to which even his judges might have yielded" (Dan-Cohen 1984). If this argument is persuasive, the law ought to permit accused persons to defend themselves by pleading duress. However, if the availability of the defense is generally known, it invites obvious abuse. People will be less likely to resist pressures to commit crimes that they might in fact be perfectly capable of resisting. Similarly with many other defenses and mitigations in criminal law: they provide opportunities for judges to temper or mitigate a harsh response, and precisely for that reason it is better if they are not widely known, since general knowledge of their availability would weaken the deterrent effect of the law.

If the availability of such doctrines is transmitted only selectively, Dan-Cohen argues, we can have our cake and eat it too. The law can be merciful without losing its deterrent effects. "[T]he law's own violence and brutality may suggest a general rationale for selective transmission:
in some circumstances selective transmission can mitigate or serve as a substitute for the violent means that the law frequently employs” (Dan-Cohen 1984, p. 667).

As Dan-Cohen is acutely aware, however, acoustic separation seems antithetical to the publicity principle, which he understands to be an analytical component of the rule of law ideal as well as of political morality more generally. “Central to the rule of law is the requirement that the law be clearly stated and publicly proclaimed. The alarm likely to follow the realization that selective transmission may circumvent these requirements accordingly seems well founded” (Dan-Cohen 1984, p. 667). Violating the rule of law through selective transmission of legal doctrines contradicts the publicity principle.

6.1.2 Nuclear Deterrence and Publicity

Nations aim to possess nuclear weapons for a variety of reasons, but surely the most significant of these is to gain political advantage. Clausewitz defined war as the pursuit of politics by other means, and I will use the term *Clausewitzian theory* to refer to the view that nuclear weapons, like conventional weapons, may be used to gain political advantage.

Even at the dawn of the nuclear age, however, some military thinkers recognized that nuclear weapons are not simply more powerful versions of conventional force. In a famous 1946 book, nuclear strategist Bernard Brodie wrote: “Thus far the chief purpose of our military establishment has been to win wars. From now on its chief purpose must be to avert them. It can have almost no other useful purpose” (Brodie 1946, p. 76). I will use the term *pure deterrence theory* for the view that a nuclear arsenal exists for only one reason: to deter aggression in order to keep the nuclear peace. Most people consider an all-out use of nuclear weapons to be totally immoral. It is, in addition, irrational from a self-interested point of view, since an all-out nuclear attack may well precipitate nuclear winter. Even on a smaller scale nuclear war may be irrational from a self-interested point of view: if Israel, for example, used nuclear weapons on its Arab neighbors it would risk contaminating and poisoning its own population.

However, many people who are persuaded that nuclear warfare is immoral or irrational believe nevertheless that nuclear powers such as the United States must maintain the capacity for all-out use of nuclear weapons in order to deter other nuclear powers successfully. Indeed, especially if all-out use of nuclear weapons is totally immoral, one is morally required to prevent their use, and lacking anything like an effective Star Wars technology, deterrence – it is often argued – is the only
way to keep the nuclear peace. Though *using* nuclear weapons is immoral or irrational, *having* nuclear weapons may be moral and rational. This is the rationale behind pure deterrence theory. In contrast with Clausewitzian theory, pure deterrence theory regards nuclear weapons as having to do not with political advantage, but with nuclear weapons and nothing else.

It is sometimes maintained that strategies based on "counterforce" – attacking only an adversary's military forces – and "flexible response" – focusing on less drastic responses than all-out nuclear war – make the arguments raised by pure deterrence theory against all-out use of nuclear weapons moot. This is incorrect. Even strategies that claim to abjure city-killing or all-out use must nevertheless maintain the capacity and will for these responses, because every strategy of nuclear deterrence demands the capacity to retaliate against any level of violence (including city-killing) with a level at least as great. Without the effective capability to launch all-out city-killing thermonuclear war, even strategies based on limited and "counterforce" responses unravel, since the high trump card remains in the adversary's hands.

Thus, pure deterrence theory remains central to the discussion of nuclear weapons policy. Pure deterrence theory demands that a nuclear power engage in bluff, deterring adversaries by threatening to retaliate, but in reality intending not to retaliate should deterrence fail. This is because according to pure deterrence theory nuclear weapons can be retained only to deter and never to be used. Adversaries must nevertheless believe, or at least suspect, that a nuclear power will launch a retaliatory second strike despite the power's commitment to pure deterrence. The nuclear power must credibly threaten the second strike, and the only way to make such a threat credible is to convince adversaries that the power in fact does not believe in pure deterrence theory – that is, that the power does not believe that all-out nuclear use is inevitably immoral or irrational.

This conclusion follows from the so-called paradox of deterrence, stated by Jonathan Schell as follows: "One cannot credibly deter a first strike with a second strike whose raison d'être dissolves the moment the first strike arrives. It follows that, as far as deterrence theory is concerned, there is no reason for either side not to launch a first strike" (Schell 1982, p. 202). Bluffing the adversary into believing that one does not believe in deterrence theory is the only way out.9

9The raison d'être for a second strike can dissolve in the face of an enemy first strike either for moral reasons (e.g., one believes that while it is morally acceptable to threaten a second strike, it is immoral to carry out the threat) or for prudential reasons (e.g., the second strike lowers one's own utility, for example by inducing nuclear winter). Thus, the paradox of deterrence may be viewed as either a moral
From this it follows that pure deterrence theory cannot survive the publicity principle. Pure deterrence theory requires from us a policy of constructing the capacity for the all-out second strike *but intending never to use it*. Each nuclear power is morally required to bluff the others, but clearly a bluff cannot withstand full publicity.

One might object that the publicity principle is intended for use only within a single community, so that the inability of a policy of nuclear bluff to withstand universal publicity, outside as well as within a single society, does not really run afoul of the publicity principle. Unfortunately, this does not matter. Even full publicity within a single nation of the policy of nuclear bluff is impossible, for the obvious reason that an entire nation cannot go off and huddle out of earshot of other nations.

### 6.1.3 Morality

My final two illustrations of the publicity principle in action are drawn from the arena of what might be called general morality. The first is an issue in political morality, while the second is drawn from moral theory.

A frequent accusation directed at politicians is that they are hypocrites, who profess a high standard of rectitude but compromise it repeatedly. The eminent political philosopher Judith Shklar has defended public hypocrisy on the part of politicians, which she believes is essential to a democratic political order. According to Shklar,

> representative democracy must, like any form of government, maintain its legitimacy by reinforcing the ideological values upon which it is based. . . . That means that those engaged in governing must assume at the very least two roles, one of pursuing policies and another of edifying the governed in order to legitimize these plans. . . . There is nevertheless a built-in tension; for the disparity between what is said and what is done remains great. . . . Without ancestor worship or divine provenance to rely on, modern liberal democracy has little but its moral promise to sustain it. That is why it generates . . . the interplay of hypocrisy and vocal antihypocrisy. (Shklar 1979, pp. 13–14)

Hence Shklar admires Benjamin Franklin, who defended hypocrisy as a requirement of democratic politics:

He was a shrewd calculator who took it for granted that the politics of persuasion required hypocrisy. . . Here is hypocrisy as a conscious act in response to a situation that demands it. Persuasion is not natural; it requires a great deal of effort, and in a man as superior to his fellows as Franklin was, it takes exactly what he described. . . A democratic "social fabric" would "come undone" just as quickly as any other if everyone were always "wholly frank with everyone." (Shklar 1979, pp. 16–17)

Like the Platonic view of justice and legal institutions, Shklar's account endorses a double standard of behavior. A man like Franklin, who is "superior to his fellows," must engage in actions for reasons that he masks from his inferiors. Hypocritical action, however, cannot withstand the publicity principle: hypocrisy is the antithesis of publicity.10

In a sense, Shklar's argument translates Plato's concept of the "noble lie" into the political framework of modern democracy. But, like Plato's original notion, the argument concludes that some acts of political wisdom can never be publicized, and such a conclusion is inconsistent with the publicity principle.

Similar conclusions dog the moral theory of utilitarianism. A well-known problem with utilitarianism is that utility may best be served if people do not believe in utilitarianism. One way this can be so is that people may be happier if they do not believe in utilitarianism; in that case the utilitarian project of maximizing human happiness demands that people not believe in utilitarianism. As Bernard Williams writes,

Utilitarianism would do well then to acknowledge the evident fact that among the things that make people happy is not only making other people happy, but being taken up or involved in any of a vast range of projects. . . Now none of these is itself the pursuit of happiness. . . Happiness, rather, requires being involved in, or at least content with, something else. (Williams 1973, p. 112)

Peter Railton considers this to be an instance of the more general "paradox of hedonism," the fact that self-consciously pursuing happiness may actually make you less happy (Railton 1984, pp. 140–46).

Another way that utilitarian ends may better be served if few people believe in utilitarianism arises if, as seems likely, we are best motivated to do non-self-interested things such as utilitarianism may require of us if we believe in a different morality than utilitarianism – if, for example, we believe in Christian morality, or honor, or Cosmic Love. Williams has remarked that British colonial administrators in India, many of whom numbered themselves among the utilitarian faithful, decided

\[10\] I have criticized Shklar's argument in some detail in Luban 1982, pp. 1698–1703.
that the Indians would respond better to Hindu arguments than to utilitarian arguments, and therefore chose not to “convert” Hindus to the Benthamite creed. Such an elitist form of the theory has been termed “Government House utilitarianism” (Sen and Williams 1982, p. 16). In its Government House form, where the utilitarian elite inculcates antiutilitarian arguments to serve utilitarian ends, or teaches principles that are false from a utilitarian viewpoint while insisting that they are part of the Benthamite Gospel, it is clear that utilitarians embrace the use of the noble lie.

Nowhere is this clearer than in Sidgwick’s famous exposition of utilitarianism. Sidgwick worries that the complex, exception-laden rules of utilitarianism might be misunderstood or abused by simple folk; he proposes that in that case the simple folk should be deceived. I quote his extraordinary argument at some length.

Thus, on Utilitarian principles, it may be right to do and privately recommend . . . what it would not be right to advocate openly; it may be right to teach openly to one set of persons what it would be wrong to teach to others; it may be conceivably right to do, if it can be done with comparative secrecy, what it would be wrong to do in the face of the world. . . . These conclusions are all of a paradoxical character: there is no doubt that the moral consciousness of a plain man broadly repudiates the general notion of an esoteric morality, differing from that popularly taught; and it would be commonly agreed that an action which would be bad if done openly is not rendered good by secrecy. (Sidgwick 1966, pp. 489–90)

So far, Sidgwick is merely stating in more general terms an argument similar to Dan-Cohen’s defense of “acoustic separation,” including Dan-Cohen’s acknowledgment that such a defense is fraught with paradox. But Sidgwick next adds a twist to the argument:

We may observe, however, that there are strong utilitarian reasons for maintaining generally this latter common opinion. . . . Thus the Utilitarian conclusion, carefully stated, would seem to be this; that the opinion that secrecy may render an action right which would not otherwise be so should itself be kept comparatively secret; and similarly it seems expedient that the doctrine that esoteric morality is expedient should itself be kept esoteric. (Sidgwick 1966, p. 490)

The doctrine of acoustic separation must itself be acoustically segregated. From this, Sidgwick draws an explicitly political conclusion:

Or if this concealment be difficult to maintain, it may be desirable that Common Sense should repudiate the doctrines which it is expedient to confine to an enlightened few. And thus, a Utilitarian may
reasonably desire, on Utilitarian principles, that some of his conclusions should be rejected by mankind generally; or even that the vulgar should keep aloof from his system as a whole. . . . (Sidgwick 1966, p. 490)

This is Sidgwick's final twist. Not only is it best if utilitarianism is not widely believed, but it is best if the very fact that belief in utilitarianism has been suppressed by the "enlightened few" is not widely believed. This is no longer Government House utilitarianism. It is conspiratorial utilitarianism.

A utilitarianism that cannot succeed if it is publicized violates the publicity principle; and Sidgwick's extraordinary metadoctrine that the "doctrine that esoteric morality is expedient should itself be kept esoteric" not only violates the principle but, by keeping that fact secret, violates it at two levels simultaneously.

Rawls employs a different antiutilitarian argument based on the publicity principle (Rawls 1971, pp. 180–83). According to Rawls, if utilitarianism were the public morality of a society, members of that society would understand that their interests would be sacrificed by public officials whenever doing so served the larger social good. Each of them would thereby come to understand that from a social point of view their interests have no intrinsic worth whatever – as H. L. A. Hart put the point, each counts for one because none count for anything (Hart 1979, p. 830). Inevitably, Rawls argues, this realization would undermine their self-respect, and in so doing, would undermine their respect for the interest of others. As the phenomenon of fratricidal underclass violence graphically illustrates, individuals to whom the larger society conveys the message that they have no intrinsic worth and that their interest can be traded away are scarcely likely to be inclined to respect the interests of others. In that case, however, where will public support for the principle of utility, which requires some citizens to sacrifice their own interests to further the interests of others, come from? A utilitarian public morality, publicly acknowledged, will undermine its own support. Hence, utilitarian public morality cannot be publicly acknowledged: it must remain esoteric. If the publicity principle is true, it follows that utilitarian public morality is unjust.

Notice that the conclusion that utilitarianism is inconsistent with the publicity principle requires only a few plausible empirical premises: that nonutilitarians are generally happier than utilitarians, that people are best motivated to altruism by nonutilitarian beliefs, that people are likely to reject the elitist implications of utilitarianism, that by denying the absolute worth of individuals the principle of utility weakens the psychological underpinnings for the other-regarding altruism that utili-
itarianism requires. If we accept these premises, we are forced to abandon either utilitarianism or the publicity principle.

6.2 Parsing the Publicity Principle

For the sake of convenience, I shall continue to discuss the publicity principle in Kant's formulation. In this section, I shall consider possible meanings of three of the key concepts in that formulation: "maxim," "publicity," and the notion of a maxim being "incompatible with" publicity. My purpose is not primarily philological: I am interested less in what Kant actually meant by these terms than in what meaning we should most reasonably ascribe to them.

6.2.1 Maxims

The principle tells us that actions affecting the right (or rights) of others are wrong if their "maxim" is incompatible with publicity. "Maxim" is a semitechnical term Kant introduced in his *Fundamental Principles of the Metaphysics of Morals*: "A maxim is the subjective principle of volition."11 To talk about maxims is to focus attention on the subjective motivations of agents.

Restricting the publicity principle to maxims, as Kant does, has important consequences. The same action may result from more than one maxim; and some of those maxims may be less compatible with publicity than others. For example, hiring a white job applicant over a black job applicant may be done because the personnel officer reasonably believes that the former is qualified for the job and the latter is not; but it may also be done because the personnel officer is a racist. "Hire a qualified applicant rather than an unqualified applicant" is a maxim that could presumably withstand publicity, whereas "Hire a white applicant rather than a black applicant" is not.

By directing our attention away from actions toward their maxims, Kant means to focus our attention on the motives of political actors rather than on the nature of the actions they are performing. Since people's motives are often difficult for others to discover, Kant's restriction means that the publicity principle will be more useful as a principle of first-person deliberation by decision-makers than as a principle of third-person evaluation by observers. This way of reading it seems to

---

11Kant 1785/1949, p. 18n. Kant adds: "The objective principle (i.e., that which would also serve subjectively as a practical principle to all rational beings if reason had full power over the faculty of desire) is the practical *law*" (ibid.).
accord with Kant's own intentions: much of the argument of "Perpetual Peace" is a plea directed toward decision-makers.

Clearly, there are pluses and minuses to applying the publicity principle to the subjective motivations of decision-makers, rather than to the objective characteristics of their actions. On the plus side, Kant's strategy tells decision-makers that the bare fact that their actions may be publicly justifiable under some maxim is irrelevant if their actual maxim could not withstand the light of day. On the minus side, Kant's restriction may let decision-makers off the hook: since we are all capable of self-deception, decision-makers will undoubtedly persuade themselves that their subjective motivations are unimpeachable, even though it appears to others that they had acted for disreputable reasons. In such cases, it might do more good for the decision-maker to look at what he did, and not at the contents of his own soul.

Kant's focus on maxims, that is, on subjective rules of action, has another effect as well, and indeed helps clear up a puzzle about the publicity principle. The puzzle is the apparent incompatibility of the publicity principle with any secrecy whatever in public policy. Suppose, for example, that a government agency wishes to keep secret that weapons-grade plutonium is stored in Ravenna, Ohio. Obviously, if the action is described as "secretly storing weapons-grade plutonium in Ravenna, Ohio," it cannot survive the publicity test, for publicizing the action under this description reveals the very secret at issue.

The problem, of course, is that the description has too much particularity built into it. If instead the action is described simply as "secretly storing weapons-grade plutonium," it may well survive the publicity test, since most people would presumably agree that such a policy is desirable. By focusing the publicity test on maxims, that is, general rules of action, Kant invites us to deliberate at the level of policies, not at the level of specifics, which may indeed need to be kept secret.

6.2.2 Publicity

What exactly is the publicity that Kant invokes in his publicity principle? Let us consider three possibilities: publicity as general knowledge, publicity as mutual knowledge, and publicity as critical debate.

**GENERAL KNOWLEDGE.** $X$ is public if everyone knows $X$. Of course, we should not mean literally everyone, since obviously there will be some people who don't follow the news or talk with others; or who are so geographically isolated that they do not find out much about the outside world. "Everyone" means tout le monde, "the general public," those who care about such things.
General knowledge is surely a necessary condition of publicity. It may not be enough to guarantee that an item is public in the politically relevant sense, however.

I have had experiences in which a colleague has approached me privately to tell me something quite significant— that he or she is getting a divorce, or has been offered a job elsewhere, for example. In each of these cases I assumed from the discreet way my colleague behaved that he or she had decided to confide in only a few people, and expected me to keep the information to myself— only to learn later that my colleague was in fact telling everyone. The knowledge of my colleague's predicament was general, in that everyone in the office knew about it. But, during the short period when we were all keeping it confidential from each other, it was not public. Our knowledge was more like parallel private knowledge.

This suggests that public knowledge is not merely general, but mutual.

**Mutual Knowledge.** $X$ is public if everyone knows $X$ and everyone knows that everyone knows $X$.

One could perhaps go further, and say that $X$ is public if

- everyone knows $X$;
- everyone knows that everyone knows $X$;
- everyone knows that everyone knows that everyone knows $X$; etc.

Stephen Schiffer has introduced the term "mutual knowledge*" (read: "mutual knowledge star") to refer to the limit of this sequence: mutuality all the way up (Schiffer 1972, pp. 30–31; see also David Lewis's very similar notion of "common knowledge": Lewis 1969, p. 56). One could do worse than define publicity as Schiffer's mutual knowledge* though I suspect that soon enough the iterations become for all practical purposes identical to each other.

In practice, examples such as my anecdote about my overly secretive colleagues occur only in small organizations; thus, when we are talking about public policies or other societywide items of knowledge, general knowledge is invariably mutual knowledge as well. If I read about a federal policy in the newspaper, I know about the policy and I know that the rest of the public knows about it as well.

**Critical Debate.** People may mutually know about something, however, without debating or discussing it; and we may think that the political significance of publicity lies in public debate rather than passive public awareness. This, I should note, is an idea stressed by Habermas in his influential study of the "public sphere" (Habermas 1989).
For Habermas, "the public" means "the critical-rational public," and much of Habermas's polemic against the contemporary mass media concerns the extent to which the media have become instruments of propaganda and pacification rather than of critical debate. Following Habermas, one might insist that publicity must mean not merely mutual knowledge, but rational public debate.

But a problem arises when this interpretation of publicity is coupled with Kant's claim that the publicity principle provides an easy test of the wrongness of maxims. How am I to determine whether a maxim is compatible with publicity if publicity is understood as rational public debate? How can I know a priori whether my maxim would be rejected by rational public debate? I could try to anticipate that debate in my own mind, and no doubt I should. But it is a mistake to think that debating the proposition in my own mind is anything over and above deliberating about it in the first place. At the end of my internal debate, I will arrive at what seems most reasonable to me, nothing more. In that case, the publicity principle has changed into the vacuous injunction not to follow maxims that, after full deliberation, seem wrong to the agent. This is hardly "an experiment of pure reason."

The problem is that the more content we pack into the notion of publicity, the more difficult it becomes to use the publicity principle as an independent test of maxims: all our questions about the wrongness of the maxim will simply translate themselves into equally perplexing questions about whether the maxim is compatible with some fancy notion of publicity. Perhaps the publicity principle is true only if the concept of publicity means something like critical-rational debate; in that case, however, the truth of the principle is purchased at the cost of its usefulness as a test.

Let me cast this caveat in slightly more general terms. Kant asserts not only that the publicity principle is true, but also that it provides an easy-to-use test of public policies. Much of the interest of the publicity principle rests in the prospect that Kant was right on both counts. In the examples I presented earlier, the test seemed easy to apply, and therein lay the fascination.

That being so, we must guide our interpretation of the publicity test by two hermeneutic principles: interpret it in the most plausible way, but also in the way most likely to preserve its usefulness – not merely salve veritate, but salve utilitas as well. Perhaps it is impossible to do both, but that should be our conclusion of last resort.

For this reason, I think that we should settle on mutual knowledge as our preferred interpretation of publicity. Mutual knowledge no doubt implies mutual discussion, but it does not imply full-fledged rational
debate. Interpreting publicity as mutual knowledge offers the best chance, I think, of salvaging Kant's claim that the publicity principle provides a useful test of the wrongness of public policies.

6.2.3 Incompatibility

Lastly, we must ask what it means to say that a maxim is incompatible with publicity, or cannot withstand publicity. Here, as with the concept of publicity itself, there are several possibilities: a maxim is incompatible with publicity if it is self-frustrating, arousing necessary and general opposition, unpopular, or politically suicidal. Of these possibilities, Kant explicitly discusses the first two, but the latter two are, as we shall see, plausible interpretations of what incompatibility might mean.

**SELF-FRUSTRATING.** Kant says that his test will rule out "a maxim which I may not declare openly without thereby frustrating my own intention" (Kant 1795a, p. 126). One of his examples illustrates what he has in mind. He asks whether a state may make a promise to another, reserving the right to break the promise "when its own welfare is at stake" (Kant 1795a, pp. 127–28). Kant answers that

if the ruler of a state were to let it be known that this was his maxim, everyone else would naturally flee from him, or unite with others in order to resist his pretensions; which proves that such a system of politics, for all its cunning, would defeat its own purpose if it operated on a public footing, so that the above maxim must be wrong. (Kant 1795a, p. 128)

Most, or perhaps all, of the policies discussed earlier fail the publicity principle because they would be self-frustrating if publicized. But it may be helpful to distinguish various ways publicity might make them self-frustrating.

First, a policy might be *performatively inconsistent* if the agent makes it public. If I announce that I am about to leave the house secretly, the fact of announcement – the performance – contradicts the content of the announcement. A policy by the CIA director to provide misleading intelligence to the president in order to influence the president's conduct of foreign affairs would be performatively inconsistent if the director made it public. (Note that if someone else makes it public there is no performative inconsistency, although the policy will fail.) In the nuclear deterrence case, maintaining a policy of nuclear bluff – that is, threatening to retaliate to a first strike but intending not to retaliate if deterrence fails – would enwrap one in performative inconsistency if it
were announced. One might likewise argue that Sidgwick was involved in a performative inconsistency by the very fact that he published his defense of esoteric morality.

Second, publicizing a policy might set in motion a chain of events leading to its defeat. This is, strictly speaking, the pure case of publicity making a policy self-defeating. Kant’s example illustrates this case. So does a Llewellyn/Nesson-like policy of tailoring legal procedure to reconcile us to trial results rather than to produce justice, for we saw that publicizing such a policy would do the opposite of reconciling us to trial results.

Third, publicizing a policy might set in motion a chain of events leading not to the policy’s own defeat, but to the defeat of some other important, usually related, policy. In that case we may say that publicizing the policy is collateral self-defeating (it leads not to the policy’s own defeat, but to the defeat of some collateral policy). The various examples of judicial candor illustrate this form of self-frustration. Take, for example, Calabresi’s argument about judicial sunsetting of anachronistic laws. His point is that if this judicial power were formally acknowledged in legal doctrine, the result might be abuse of the power. He is not arguing that doctrinal acknowledgment of the power would lead to the loss of that very power. Thus, he is not arguing that publicizing the power would be self-defeating. Rather, Calabresi holds only that publicizing the power would lead to other bad effects on the legal system—thus, that publicizing the power would be collateral self-defeating.

Similarly, Bickel’s theory of the passive virtues proposes that the Supreme Court manipulate doctrine in order to duck constitutional questions. The Supreme Court could hardly announce publicly that that is what it is doing. (“We deny standing to the appellant, not because she really lacks standing, but because we don’t want to reach the constitutional question”? – Hardly!) But that is not because announcing it publicly would compel the Court to decide the constitutional question it was trying to avoid. Rather, it is because announcing it publicly would create an immense scandal. Thus, publicizing that one is using the passive virtues is collateral self-defeating rather than directly self-defeating.

One special case of collateral self-defeating policies deserves special mention. In some of our examples, publicizing a policy would be collateral self-defeating because it would set up perverse incentives. In Dan-Cohen’s “acoustic separation” argument, the reason that publicizing the defenses of duress, necessity, and provocation would be self-frustrating is that public knowledge of the defenses might induce people to give in to pressure, or respond to provocations, that they might oth-
erwise have resisted. Knowledge of the defenses changes the structure of incentives in the society for the worse, and that is why publicizing them would be collateral self-defeating.

Other examples of perverse incentives are easy to come by. Late in 1991, the newly formed Russian Republic decided to remove price controls on many goods. But economists expressed concern about publicizing the date of the price rise (it was January 2, 1992), because suppliers might withhold needed goods from the market until the announced date in order to take advantage of decontrol. Thus, publicizing the date might create perverse incentives, leading to the collateral self-defeat of the policy of decontrolling prices. Similar concerns arise whenever governments decide on a currency devaluation.

NECESSARY AND GENERAL OPPOSITION. Kant writes that "a maxim . . . which I cannot publicly acknowledge without thereby inevitably arousing the resistance of everyone to my plans, can only have stirred up this necessary and general (hence a priori foreseeable) opposition against me because it is itself unjust . . ." (Kant 1795a, p. 126). Though the argument here seems somewhat question begging – isn't the injustice of such maxims precisely what he is trying to establish? – the passage suggests an interpretation of what it means for a maxim to be incompatible with publicity. It is incompatible with publicity if it "inevitably arouses the resistance of everyone," "stirs up necessary and general (hence a priori foreseeable opposition)," and thus "is itself unjust." A policy is incompatible with publicity, that is, if it is intrinsically loathsome – something akin to genocide.

This interpretation is eminently plausible, but it raises doubts parallel to our earlier doubts about the critical debate interpretation of publicity: if the test of incompatibility is a priori foreseeable opposition, intrinsic loathsome, then the injustice of the maxim will be as it were written on its face, and one would hardly need to apply the publicity principle to determine that it is unjust. That is, the publicity principle adds nothing over and above the direct contemplation of the maxim itself, and thus it is useless as an independent test of maxims.

UNPOPULAR. If it is nearly impossible to know whether opposition is necessary and general, perhaps the test should be that maxims are incompatible with publicity whenever one can foresee that they will be extremely unpopular. This, however, is not an acceptable interpretation of the publicity principle, for it would then say that any action that, if publicized, is unpopular is therefore wrong. Clearly, under this interpretation we would have a principle of demagoguery, not morality.

Recent American politics illustrates this point perfectly – unfortu-
nately. The clearest example concerns taxes. American voters insist on an undiminished delivery of governmental services and entitlements, but mercilessly punish any politician (such as Walter Mondale in the 1984 presidential election) who announces publicly that they must be paid for through tax hikes. (The attitude of the electorate has been caricatured by some observers as “It’s unfair to make the taxpayers pay for it – let the government pay for it!”) Given the demand for services and entitlements, tax increases are surely not wrong – but they could not pass the publicity test if it is interpreted as a popularity contest.

There is one way, however, in which interpreting incompatibility as unpopularity makes sense. If we interpret publicity as critical debate, then the publicity principle could be read along the following lines:

All actions affecting the rights of others are wrong if, after a critical-rational public debate, their maxim is generally unpopular.

This formula is very close to Habermas’s theory of communicative ethics, and is also in the neighborhood of Dewey’s social philosophy. For reasons we have already examined, it is not going to be useful as an independent test of public policies; but it is certainly a serious contender for the palm of truth.

**Politically Suicidal.** Our last interpretation of incompatibility emerges quite naturally from pursuing the issue from the standpoint of *Realpolitik*. For the politician, the question of whether a maxim can withstand publicity will be heard as a question of what the political fallout of publicity would be. Does it mean angry letters from the voters? Does it mean not being reelected? Does it mean being impeached? Does it mean being forced to resign? Does it mean armed resistance? Does it mean the fall of the government, execution, assassination? From the politician’s point of view, a maxim is incompatible with publicity only if publicizing it would be politically suicidal, precipitating the loss of power. Richard Nixon discovered that his policies of endorsing political dirty tricks were incompatible with publicity, because he was forced from office.

We must reject this interpretation, for the simple reason that what is politically suicidal typically depends on how tough a politician is prepared to be. Machiavelli makes the point dramatically: a short dip into *The Prince* yields a nonstop tale of butcheries, strangulations, and atrocities used successfully by princes to secure their power. Agathocles the Sicilian and Machiavelli’s contemporary Oliverotto di Fermo both be-

---

12Kant himself explicitly rejects the unpopularity interpretation of incompatibility. Kant 1793, p. 79.
came rulers by gathering together the leading citizens of their cities and then treacherously ordering their soldiers to murder them all (Machiavelli 1532, chap. 8). Presumably, rulers willing to engage in such tactics would not resign from office simply because word got out that they had authorized the Watergate burglaries. Thus, a policy of burglarizing one's political opponents is compatible with publicity if one is willing to behave like Oliverotto.

This shows that the political suicide interpretation of incompatibility cannot be right. For it implies that the more ruthless a government is, and the more it is willing to hang onto power by injustice and violence, the more "compatible" its policies are with publicity. Political homicide is the tried and true alternative to political suicide. In short, the more unjust the government, the less likely its actions are to fail the publicity test.

To summarize, the political suicide interpretation of incompatibility must be abandoned. The unpopularity interpretation is quite unattractive, unless it means unpopularity after critical-rational debate, in which case the publicity test is unusable. If unpopularity means inevitable unpopularity, it is no different from the necessary and general opposition interpretation. Thus, only the first two interpretations of incompatibility survive. Kant himself accepted the disjunction of the two:

For a maxim which I may not declare openly without thereby frustrating my own intention, . . . or which I cannot publicly acknowledge without thereby inevitably arousing the resistance of everyone to my plans . . . is itself unjust and thus constitutes a threat to everyone. (Kant 1795a, p. 126)

Maxims are unjust if publicizing them results either in self-frustration or inevitable opposition. However, we have seen that the "inevitable opposition" interpretation cannot be applied as a self-standing test by decision-makers. Thus, we should interpret the notion of incompatibility in the publicity principle along the lines of the self-frustration interpretation (on any of the various modes of self-frustration).

6.3 Kant's Arguments for the Publicity Principle

In the first four paragraphs of the second appendix to "Perpetual Peace" Kant unveils the publicity principle amid a dense clot of arguments, argument fragments, and hints of arguments for its truth. Kant is seldom easy on his readers, and these passages are perhaps more gnomic than is usual even for Kant. Gnomic or not, his arguments form the natural starting point for our inquiry into whether the principle is
true: they remain, I believe, the most plausible arguments for the publicity principle.

6.3.1 The Public Law Argument

Start with the opening paragraph of the "Perpetual Peace" appendix (I insert bracketed numbers for ease of reference):

[1] If, in considering public law [Recht] as legal scholars usually conceive of it, I abstract from all its matter (as determined by the various empirically given relationships of men within a state, or of states with one another), I am left with the form of publicity [Publizität], which every legal claim [Rechtsanspruch] potentially possesses, since [2] without it there can be no justice (which can only be conceived of as publicly knowable) and therefore no right [Recht], which can be conferred only by justice.

[3] Every legal claim must have this capacity for publicity. . . . 13

Consider clause [1]. The argument seems to be that the only thing all public laws have in common is that they are public, hence publicity must be their formal attribute.

Kant draws this conclusion about public law "as legal scholars usually conceive of it." The traditional Roman concept of public law, as found on the first page of the Digest of Justinian, was thoroughly familiar to Kant. However, in this sense of public law, Kant's argument amounts to little more than a play on words, for "public law" does not mean "law that is public." It means "law that concerns relations between state and citizen (rather than between citizen and citizen)." Roman lawyers labeled such law "public" not because it is publicly promulgated, but because it concerns matters of public interest. Kant writes as though the publicity of public law follows from the very meaning of the term, but

13Kant 1795a, p. 125 (my emendation to the translation, based partly on Kant 1795b). Following Beck (Kant 1795b, p. 129), I have translated "Recht" at one point as "law" and at another as "right." The word occurs twice in the same sentence, the first time in the phrase "öffentliches Recht," which I believe must be translated "public law" rather than (as in Reiss) "public right," in order to retain Kant's focus elsewhere in "Perpetual Peace" on the role of law in international relations. This being so, it is of course taking considerable liberty with the text to translate the second occurrence of "Recht" in the same sentence as "right," as Beck and I do. However, the second occurrence cannot possibly be translated as "law," for then Kant's use of the verb "erteilen" (to confer or award) makes no sense: while justice can confer rightness on a claim to legality, it cannot confer law on it. It seems clear that Kant is exploiting the double sense of "Recht": he is arguing that only justice can confer moral rightness to positive law.
it does not, any more than the fact that lawyers place contract doctrine in the category of private law implies that contract law is secret.

Nor does the argument get much further if we read "public law" in Kant's own idiosyncratic sense of law "that need[s] to be promulgated generally in order to bring about a lawful condition [rechtlicher Zustand]" — Kant's definition of public law (Kant 1787, p. 123).\textsuperscript{14} For then the argument amounts to little more than a tautology. Law that must be promulgated generally shares the attribute of publicness. So what? How do we know that all, or even any, law must be promulgated generally? We need some such claim to get to the desired conclusion in passage [3], which tells us that every legal claim must be capable of publicity, not just that legal claims requiring publicity must be capable of publicity.

\textbf{6.3.2 The Rule of Law Argument}

On the other hand, many contemporary writers accept the proposition that every legal claim must be capable of publicity, because they regard public promulgation of laws as a defining condition of the rule of law.\textsuperscript{15} This suggests that, in its legal applications at any rate, the publicity principle may be a corollary of the rule of law ideal.

In that case, however, the argument for the publicity principle is only as strong as the argument for the rule of law ideal — and indeed, only as strong as the argument for the publicity of law within that ideal. It is far from plain, however, that the rule of law ideal should not sometimes give way. If a progressive South African government, stymied by conservative opposition, found itself politically unable to end apartheid, a secret proclamation directing officials not to enforce apartheid laws might improve justice in South Africa. As we shall see, plausible counterexamples to the publicity principle typically represent situations in which we are tempted to compromise the rule of law ideal in favor of other values.

Nor is it wholly obvious that the publicity of law is essential to the rule of law ideal. Consider first that, although law in contemporary liberal democracies is publicly promulgated, all this really means is that it is available in official publications that lawyers know how to find. In daily life, very few people actually know what the law is on any given subject. Only to those able to secure professional legal advice, therefore,

\textsuperscript{14}The phrase "bring about a lawful condition [rechtlicher Zustand]" means, I take it, something like "establish the rule of law." (I have altered Gregor's translation of rechtlicher.)

\textsuperscript{15}These include Fuller 1964, Rawls 1971, pp. 235–43, and Raz 1977.
is the law public in fact rather than fiction. Consider next that professional legal advice is so costly that it is available to only a tiny fraction of the population. Consider finally that this state of affairs is the consequence of deliberate public policies: Every liberal democracy has elected to distribute legal services primarily through market mechanisms, subsidizing legal aid only at negligible levels; and no liberal democracy permits nonlawyers to offer legal advice except in certain highly specialized circumstances – a state-mandated professional oligopoly of lawyers that serves chiefly to keep the cost of legal services high. In effect, Dan-Cohen’s acoustic separation principle is enforced worldwide through the simple expedient of pricing legal knowledge beyond the means of most citizens. More than any other form of government, liberal democracies are characterized by their principled commitment to the rule of law, but they pay only lip-service to the publicity of law. The only reasonable conclusion to draw is this: either no nation in the world has come anywhere near establishing the rule of law, or else the publicity of law is not an essential part of the rule of law ideal.

6.3.3 The Public Justice Argument

Next consider sentence [2] from the opening paragraph of the “Perpetual Peace” appendix: “Without it [i.e., publicity] there can be no justice (which can only be conceived of as publicly knowable) and therefore no right, which can be conferred only by justice.” Here Kant appears to be arguing not from the public character of law, but rather from the public character of justice “which can only be conceived of as publicly knowable.” Rawls, who assumes the publicity principle as a constraint on any construction of the concept of justice, evidently has this argument for the principle in mind.

Unfortunately, the argument seems at once question begging and incomplete. Question begging, because through most of human history justice has been conceived of as not publicly knowable, as an esoteric science revealed to the Chosen Few – the divinely sanctioned prophets, the philosophical souls. Some of the most sophisticated philosophy of the Middle Ages held that principles of justice contain an ineliminable kernel of revelation. Even some philosophers (Averröes is an example) who believed that the principles of justice can be discovered through natural reason held as well that only a very few individuals are up to

16See Rhode 1981. Rhode argues that standard bar arguments in favor of restricting legal practice to licensed lawyers rest on empirical premises that available data contradict.
the inquiry, and thus that the mass of mankind must be steered clear of the philosophically correct methods of scriptural interpretation. (This view is evidently a precursor of Sidgwick's theory of esoteric morality.) Contemporary philosophers who derive principles of justice from technical arguments in game theory or axiomatic bargaining theory may likewise agree that knowledge of justice is only for the mathematically trained few. In either form – revelation or restricted reason – proponents of esoteric justice argue that a profound inegalitarianism marks humanity. Principles of justice are only for the few.

The public justice argument is incomplete because even if justice must be regarded as publicly knowable, and laws must accord with the publicly knowable principles of justice, it scarcely follows that the laws cannot be secret. Perhaps, among the publicly knowable principles of justice, we find a principle permitting officials to act on policies that could not withstand publicity when discretion seems the better part of valor. Particularly when we consider the defects, vileness, and propensity for self-dealing that seems to characterize human nature and political life, the case for secrecy and dissembling in public policy can seem overwhelming, and not the least out of accord with publicly knowable principles of justice.

6.3.4 Universalizability

Rawls remarks that "[p]ublicity is clearly implied in Kant's notion of the moral law" (Rawls 1971, p. 133n). The moral law, of course, is the categorical imperative, "Act only on that maxim whereby you can at the same time will that it should become a universal law" (Kant 1985/1949, p.38). If Rawls is right, the moral law may provide an argument for the publicity principle distinct from the rule of law and public justice arguments.

Kant conceived of morality on a legal model, and identifying morality as a kind of law is perhaps the single most characteristic feature of his moral philosophy. The categorical imperative says, "Act only on that maxim whereby you can at the same time will that it should become a universal law," not "Act only on that maxim whereby you can at the same time will that it should become a universal practice (or custom)." The point is plainer still when we turn to Kant's reformulation of the categorical imperative in the language of the "kingdom of ends": "A rational being must always regard himself as giving laws either as member or as sovereign in a kingdom of ends . . ." (Kant 1785, p. 50). Kant states explicitly that "[b]y a 'kingdom' I understand the union of different rational beings in a system by common laws" (Kant 1785, p. 50).
The test Kant intends in the categorical imperative is thus whether one could *legislate* one’s maxim, not merely generalize it.

Now perhaps one can enact secret legislation even in the kingdom of ends. But it would be unthinkably bizarre to attribute such a view to Kant. Remember that Kant defines public law as law that must be made public to establish a lawful condition (i.e., to be efficacious as law). He also, however, argued that all positive law, including the traditionally private law departments of tort, property, and contract, is “public” in this special sense.\(^7\) Kant’s philosophy of law is built around the requirement that law operates publicly. On his legalistic understanding of morality, the categorical imperative includes not just universalizability, but publicity (what we earlier called mutual knowledge\(^*\)) as well: “willing that my maxim should become universal law” means “willing that my maxim should become universal and publicly known law.” This yields a simple and natural derivation of the publicity principle from the categorical imperative. If I could not publicize my maxim, I surely could not “will that it would become a universal law,” for that should be tantamount to willing that it be made public. Hence, any maxim that cannot pass the publicity test will be equally incapable of passing the test of the categorical imperative. The maxim will therefore be wrong, just as the publicity principle says.

Without entering into the interpretive question of whether Kant intended to relate the publicity principle to the moral law – I regard the textual evidence as too ambiguous for a clear answer – let us note that Kant never so much as mentions the categorical imperative in his discussion of the publicity principle. Why? I think that the most plausible answer is that Kant did not distinguish between the universalizability and publicity components of the moral law, but instead simply combined them in the single complex he termed “uni-

\(^7\)In the introduction to Kant 1787/1991, Kant explains that the distinction between private and public law is simply another name for the distinction between natural law – law in the pre-social-contract state of nature – and civil (*bürgerliche*) law, in other words positive law. “The highest division of natural Right [Naturrecht] cannot be the division (sometimes made) into natural and social Right; it must instead be the division into natural and civil Right, the former of which is called private Right and the latter public Right. For a *state of nature* is not opposed to a social but to a civil condition, since there can certainly be society in a state of nature, but not civil society (which secures what is mine or yours by public laws). This is why Right in a state of nature is called private Right: of which the first is termed private law, the second public law. For the natural condition is not counterposed to the social, but rather to the civil: since in the state of nature society can indeed exist, but not civil society (which secures the ‘mine’ and the ‘thine’ through public statutes); thus law is called private law in the former” (Kant 1787/1991, p. 67).
versal law” (cf. Rawls 1971, p. 133). If Kant was not prone to partition the categorical imperative into universalizability and publicity components, then any moral argument he might fashion for the publicity principle would derive the principle from the categorical imperative 
taken as a whole. But Kant would be very unlikely to find that an appealing derivation, for it is blatantly circular. The publicity principle, remember, is preeminently a test of the moral wrongness of (public) laws. To test such laws by means of the categorical imperative amounts to asserting that a (public) law is morally wrong if one could not will it as a universal law, or as a piece of legislation in the kingdom of ends. That is singularly uninformative, and Kant would hardly trouble his readers with an argument that boils down to the proposition that a piece of legislation is unacceptable if it is unacceptable. Clearly, he would prefer to regard the publicity principle as an independent test of the morality of public laws. It is small wonder that under such circumstances Kant did not mention the categorical imperative in his arguments for the publicity principle.

This point, however, highlights a deep problem in the entire effort to argue for the publicity principle on the basis of Kantian morality. We have seen that Kant built publicity into the categorical imperative because at bottom he conceived of morality itself on the model of public law (in Kant’s sense of law that must be made public). It will hardly do to argue that law incapable of publicity is immoral, if the basis of the reasoning is simply that morality has been defined as a kind of public law. In that case, law turns out to be like morality only because morality has been modeled on law.

We may put the point in a different way. Obviously, anything that violates the publicity requirement thereby violates the publicity requirement and universalizability, that is, violates the categorical imperative. But universalizability does no work in this argument. Any property could be substituted for “universalizability” without affecting the derivation.

This is important because philosophers in the Kantian tradition have focused their interest and attention on universalizability, rather than on publicity or the quasi-legal character of morality. And this is for good reason: the universalizability requirement, far more than the publicity requirement, expresses the moral demand of impartiality and the conception of the underlying moral equality of all rational beings – ideas that are surely the principal attractions of Kantian ethics. To learn that universalizability plays no essential role whatever in the derivation of the publicity principle from the moral law surely removes most of the interest and importance of that derivation.
6.4. The Case against the Publicity Principle

In the end, I don’t believe that Kant himself makes a persuasive case for the publicity principle. Indeed, I shall suggest that no a priori argument for the publicity principle is likely to work. To see this more clearly, I now want to turn to the case against the publicity principle, which I present through a series of counter examples: policies that fail the publicity principle but that nevertheless seem on the whole the right thing to do. I group them into four categories: examples of mercy, examples of sunsetting, examples of Victorian compromise, and examples of shielding the righteous.

6.4.1 Examples of Mercy

1. When I was in graduate school, one of my fellow students was caught shoplifting from the university bookstore. The store detectives took him into the back room and, to his great relief, told him that they would let him go that day – but (they sternly added), if they caught him twice more they would press charges.

Whether the bookstore’s policy of pressing charges only on the third offense was a wise one is of course open to debate. It was, at any rate, a merciful one, and probably saved the store unnecessary fuss, so that it may have been an efficient policy as well. Plainly, however, it was a policy that was incompatible with publicity – and the store detectives were literal-minded dunderheads.

2. At a more sophisticated level, Dan-Cohen’s examples of acoustic separation in the criminal law are also counter examples to the publicity principle based on mercy. Devices whereby knowledge of certain defenses and mitigations is restricted only to officials – judges, let us say – allows us to retain the deterrent effect of criminal law while at the same time permitting judges to show mercy to defendants who have succumbed to forgivable and all-too-human impulses. As in the preceding case, publicizing merciful policies creates perverse incentives for the unscrupulous to take advantage of them.

3. One such device is particularly instructive. Nothing prevents a criminal jury from acquitting a defendant who is plainly guilty as charged, guilty beyond a reasonable doubt; and because of the ban on double jeopardy, the jury’s decision to acquit cannot be appealed. Juries thus possess the power to nullify the criminal law. Theorists typically applaud this power of jury nullification, suggesting that it is a device whereby the community’s evolving sense of justice, expressed by a representative group of ordinary citizens, can temper excessive harshness
or anachronism in the law.\textsuperscript{18} An example of tempering harshness would be acquitting an elderly, devoted husband who has mercy-killed his beloved but severely Alzheimer’s afflicted wife; an example of tempering anachronism would be acquitting individuals (heterosexual as well as homosexual) indicted under nineteenth century anti fellatio statutes.\textsuperscript{19}

Yet even those who applaud the power of jury nullification may well object to a judge or attorney explicitly pointing out to the jury that it possesses the power. Some American jurisdictions forbid lawyers from explicitly asking a jury to nullify, and judges will typically instruct the jury that if it is convinced that the state has proven its case, it must convict—leaving it for the jurors to figure out on their own that if they choose to ignore those instructions the sky will not fall and the defendant will go free.\textsuperscript{20} Though the argument is never put in entirely frank terms, the thought seems to be this: permitting lawyers or judges to call the nullification power to the jury’s attention might confuse jurors, making them believe that their assigned task is not to find facts and match them to law, but rather to decide whether or not they think the law is fair. The distinction between leaving nullification to the jury’s spontaneous act of conscience and inviting the jury to decide on the basis of conscience rather than law is crucial but subtle, so subtle (it is feared) that to call the power to the jury’s attention at all is to invite its abuse in cases when nullification is inappropriate. Nullification cannot survive publicity.

\textsuperscript{18}For a sophisticated treatment of jury nullification and similar issues, see Kadish and Kadish 1973.

\textsuperscript{19}Heterosexual cases are surprisingly common; they typically rise in the aftermath of a prosecution for rape, in which both parties agree that an act of fellatio took place, but in which the man insists it was consensual while the woman charges coercion. If the judge or jury is persuaded that the complainant had consented, the man will often find himself acquitted of rape but indicted for oral sodomy.

\textsuperscript{20}Thus, in the well-known trial of “subway vigilante” Bernhard Goetz, who had shot four black youths who approached him in a New York subway asking for money, Goetz’s lawyers made a back-handed appeal for jury nullification. When Goetz was convicted of a weapons charge, he appealed because the judge had instructed the jury that, “if it found that the People had proved each of the elements of the crime beyond a reasonable doubt, it ‘must’ find the defendant guilty.” Denying the appeal, the New York Court of Appeals said: “The jury’s function is to apply the legal definition of the crime to the evidence and to convict if it is satisfied that each element has been established beyond a reasonable doubt. While there is nothing to prevent a jury from acquitting although finding that the prosecution has proven its case, this so-called ‘mercy-dispensing power’ is not a legally sanctioned function of the jury and should not be encouraged by the court.” People v. Goetz, 532 N.E.2.d 1273 (1988).
6.4.2 Examples of Sunsetting

1. Consider again the 1992 Russian price decontrol. An old price regime was about to sunset; but to inform people when it would sunset invites them to withhold needed goods from the market. A good policy can be disastrous if it is publicized.

Other examples of the same type are easy to come by. Governments never announce the date of intended currency devaluations; and, in the United States, skeptics about the efficacy of gun control rightly warned that a ban on assault rifles would lead to an enormous run on gun stores before the date the legislation took effect. (Yet in this case it was impossible to keep the date secret.)

2. More abstractly, theorists of the Prisoner's Dilemma have long argued that a supergame consisting of several consecutive Prisoner's Dilemma games between the same players would lead to the same outcome as a single play — mutual defection — if the players know exactly how many games are in the supergame. Suppose, for example, that the players are told that they will play ten repetitions of the Prisoner's Dilemma. After game 9, they have only one game to play, and so they will both defect. Knowing that mutual defection is certain in game 10, after game 8 they will in effect have only a single uncertain-outcome game left to play, namely game 9. Thus, they will defect in game 9. In the end, the dominoes collapse in series and they will defect in all the games. In order to allow stable cooperation to emerge in Prisoner's Dilemma supergames, it is therefore necessary to conceal from the players how many times they will be facing one another. Only in that case will the possibility of being rewarded for cooperation and punished for defection in subsequent plays affect the players' calculations and allow them to cooperate rationally. As in the previous examples, knowing when a situation is going to sunset may alter one's behavior for the worse.

3. During the 1960s and 1970s the Chinese government wished to experiment for a limited time with a variety of different institutional arrangements on agricultural communes. They soon discovered, however, that residents of the communes would change their behavior in the waning months of the experiment in order to prepare for the return to preexperiment arrangements, thereby queering the experiments. Consequently, the government began instituting experiments without announcing that they would sunset after two or three years.21

21I owe this example to Stephen Holmes, who attributed it to Jon Elster.
6.4.3 Examples of Victorian Compromise

"Victorian compromise" is Lawrence Friedman's term for the typical Victorian manner of handling sexual impropriety: so long as you don't rub my nose in it, I will tolerate your escapades.\textsuperscript{22} If you force me to acknowledge that I know what you're up to, you will likewise force me to stop tolerating your vices. The pattern is a familiar one: it is the way many parents deal with their teen-agers' sex lives (the teen-agers understand without being told in so many words that they can sleep with their boy- or girlfriend so long as they enable their parents to continue pretending that they don't know what is going on).

It is also a pattern in politics, however. During the Cold War, both Americans and Soviets repeatedly violated each others' air space with spy planes. Both governments knew about the violations but understood that if the public knew, the government would be forced to respond with some form of hostility. So great were the dangers that both governments developed unspoken policies of tolerating the spy planes, provided that the other government did nothing to publicize its flights.

6.4.4 Examples of Shielding the Righteous

Another episode from the career of Russell Long: on one occasion, Long hoped to extract a hefty campaign contribution from a wealthy businessman. As quid pro quo for the contribution, Long offered to attach to pending tax legislation a provision from which the businessman would profit handsomely. The businessman balked at the size of the proposed contribution, however, and asked Long what he would get if he contributed a much smaller amount. Long answered: "Good government."

It is not exactly news that special interests and good government are often at odds, nor that politicians frequently owe their jobs to the former rather than the latter. That raises an important question of institutional design: suppose that politicians actually would prefer good government to policies designed principally for the benefit of predatory special interests (there is no reason beyond facile cynicism to suppose otherwise). Suppose further, however, that politicians are no more likely than anyone else to engage in kamikaze behavior, and that voting publicly for good government and against special interests is kamikaze be-

\textsuperscript{22}Friedman 1985, p. 585 (defining the Victorian compromise as "a certain toleration for vice, or at least a resigned acceptance, so long as it [remains] in an underground state").
havior. The conclusion that follows would seem to be this: we need institutions designed to permit politicians to do the right thing away from the scrutiny of special interests. U.S. Senator Robert Packwood explains:

Common Cause [an American political reform organization] has everything upside down when they advocate "sunshine" laws. When we're in the sunshine, as soon as we vote, every trade association in the country gets out their mailgrams and their phone calls in twelve hours, and complains about the members' votes. But when we're in the back room, the senators can vote their conscience. They vote for what they think is the good of the country. Then they can go out to the lobbyists and say: "God, I fought for you. I did everything I could. But Packwood just wouldn't give in, you know. It's so damn horrible."23

Along the same lines, Richard Neely has argued persuasively that workable democratic political systems must evolve sophisticated mechanisms for privately killing, rather than publicly defeating, bad legislation. For if legislation could be halted only by public vote, legislators "would either be required to vote for every predatory bill which comes up or anger so many powerful interest groups that their political life expectancy would be about as long as that of an artillery forward observer" (Neely 1981, p. 56). These bill-killing mechanisms include intricate committee systems, calendar and agenda manipulations, and bicameralism, which permits "the experienced manipulator's stock in trade . . . guaranteeing that the bill will pass both houses in different forms – thus never becoming law."24

Assuming that these arguments have merit, it follows that the politics of the back room, which seems plainly contrary to the publicity principle, is nevertheless indispensable in a workable democracy. At the same time, of course, the example is double-edged, and indeed reveals the basic tension underlying the publicity principle: Packwood's hypothetical acts of public-spirited virtue are possible only behind closed doors, but so was Long's effort to extort a contribution in return for special interest legislation. Backroom politics protects the virtuous and the vicious alike from public exposure.

24Neely 1981, p. 60. He adds: "In Norway the constitution establishes a unicameral legislature, but the pernicious qualities of such an institution are so widely known that the Norwegian unicameral legislature has divided itself into two separate houses, which by internal rules must each independently pass a given piece of legislation" (Neely 1981, p. 59).
6.5 A Defense of the Publicity Principle: A First Step

A hard-line defense of the publicity principle would reject the counterexamples. One might argue, for example, that the Chinese agricultural experiments really were immoral, precisely because they were so manipulative that they would fail if their conditions were publicized. One might reject backroom politics on the grounds that the sleazy transactions probably outnumber the public-spirited ones, or even that the gain in openness is worth some loss in public weal. One might observe that in fact telling players the number of Prisoner's Dilemmas they are about to play in a long series does not lead them to defect on every play, and thus that the game theorist's domino theory lacks descriptive bite; one might likewise doubt that publicizing the availability of criminal defenses such as duress or necessity would actually make any difference in the crime rate.

I think that this hard-line defense is unpromising, because I doubt that all of the counterexamples can be conjured away. Together, I believe these and similar counterexamples we might construct show that the publicity principle cannot be accepted in an unqualified form as a "formal constraint on the concept of right," let alone a "transcendental formula of public law." I shall argue, however, that the counterexamples themselves suggest a qualification of the principle that makes it much more plausible. I shall then offer an argument that, if not a complete defense of the qualified publicity principle, at least amounts to a sketch of how that defense might run.

The most persuasive of our counterexamples seem to be of two sorts. Consider:

1. The Russian price decontrol example. Suppose that I were a Russian farmer who knew a bit of game theory. I might think to myself:

Hmm! Price decontrols are for the good of the entire country, for in the medium to long run they will entice us farmers to bring more of our goods to market and alleviate scarcity. If the date of decontrol is announced publicly, however, then it will behoove each of us farmers to withhold our goods from the market until that date, even though our country risks collapse as food supplies dwindle in a harsh December. Too bad for the country. Furthermore, we risk flooding the market come Decontrol Date, driving prices down and annihilating our own profits from the price decontrol. Too bad for us. On the other hand, my individual strategy of withholding my crop until D-Date clearly dominates the strategy of sending the crop to market now, and so does the withholding strategy for all the other farmers. If the government announces D-Date, then it will plunk us farmers
into a multiperson Prisoner Dilemma, in which the only rational strategy is to withhold the crops and ruin the country.

If, on the other hand, the government keeps D-Date a secret until the last possible moment, none of us will have a rational motivation for withholding our crops (which may, after all, simply rot if we gamble on a prompt decontrol). For the good of the country, I hope that the government keeps D-Date under wraps.

In this case, the farmer has an *autopaternalistic* motive for wishing the government would keep Decontrol Date under wraps: he wants the government to remove a source of temptation (in this case, *rational* temptation) by concealing an important policy.\(^{25}\)

2. *The graduate student shoplifter.* Here the analysis is slightly different. We want the store detective to conceal the policy of charging shoplifters only on the third offense, but not (necessarily) to remove *ourselves* from the temptation to steal, as in an autopaternalism case. We may have no such temptation. Here I am not worried about you or me. I am worried about the Wicked Ones. I’m O.K., you’re O.K. – but what do we do about them? In this as in other examples of mercy, the problem is that if merciful policies are publicized they will provide perverse incentives to the wicked, who can’t be acoustically separated from the rest of us. The case is not much different, in fact, from our motivation for wishing our government would keep secret the location of weapons-grade plutonium: though we have no temptation to steal weapons-grade plutonium, we fear that the world contains plenty of people who do.

In both cases, the key to reconciling secrecy with the publicity principle lies in realizing that the publicity principle can be used to test second-order as well as first-order maxims: the maxims “Government may adopt policies that violate the publicity principle when that is the only way to prevent us from self-destructive behavior” and “Government may adopt policies that violate the publicity principle when that is the only way to forestall perverse incentives to the wicked” may themselves pass the publicity test. If the reasons for keeping a first-order policy secret can themselves be justified in a public manner, then the second-order policy of keeping the first-order policy secret is fully compatible with the publicity principle.

We may worry that this strategy will initiate a slippery slope permitting government secrecy any time government officials think they have a good reason for it. If second-order appeals to the publicity principle are acceptable, why not third- or even higher-order appeals? And if such

\(^{25}\)On the topic of autopaternalism, the principal source is Elster 1979.
higher-order appeals are valid, the publicity principle has become so weak and attenuated that it is hardly worth defending.

Consider, for example, the Iran-Contra affair. The U.S. Congress permits various government agencies to engage in covert operations, provided that congressional committees are secretly informed of them and are able to oversee them. The American public seems to accept the legitimacy of such covert actions, and thus the policy of secrecy coupled with independent oversight seems to pass the second-order publicity test.

However, members of the Reagan administration involved in the Iran-Contra activities concealed them from congressional oversight as well as from the rest of the world – and, obviously, kept secret the fact that they had unilaterally determined to conceal the operation from Congress. In the aftermath of the scandal, several Iran-Contra figures explained that CIA Director William Casey ordered Oliver North to lie to Congress because he did not trust Congress to keep the matter confidential (though it seems likely that this was merely an excuse, and that in fact Casey wished to conceal the operation from Congress because he anticipated opposition and outrage).

In effect, this explanation amounts to a third-order use of the publicity principle: the Iran-Contra figures believed that the (third-order) policy of keeping secret the (second-order) decision to conceal from Congress information about a (first-order) covert operation could be publicly justified by their concerns about congressional leaks. But – one worries – if this explanation is a legitimate appeal to the publicity principle, it is hard to see how the publicity principle constrains any kind of governmental secrecy or concealment at all.

I reply that while third- and higher-order appeals to the publicity principle represent coherent logical possibilities,\(^{26}\) they will almost never succeed in practice (and the Iran-Contra defense of lying to Congress in fact persuaded no one except right-wing fundamentalists who needed no persuasion). Thus the threat higher-order appeals seem to pose to the publicity principle is purely notional.

There is, after all, an enormous difference between second-order appeals to publicity and third- or higher-order appeals. The idea that policies of secrecy might be publicly acceptable on autopaternalistic or perverse-incentive-avoiding grounds seems straightforward and plausible. But the idea that a policy of concealing public justifications for

\(^{26}\) Except that on logical grounds an adherent to the publicity principle could not accept the $\Omega$-principle that allows all policies and higher-order policies of secrecy to be kept secret, including the $\Omega$-principle itself. For the $\Omega$-principle could not pass the publicity test.
policies of secrecy might itself be publicly justifiable has no real-world plausibility at all. Such a policy amounts to permitting officials not merely to make secret decisions, but to conceal the very fact that they have awarded this permission to themselves. Decisions to keep secrets are thereby removed from oversight by publicly established mechanisms of checks and balances. It is hard to imagine why someone who approved of such a policy would accept the publicity principle in the first place.

Higher-order appeals to the publicity principle will almost never succeed, whereas second-order appeals may succeed in a variety of circumstances. The task, of course, is to determine what those circumstances are. I have been arguing that autopaternalistic violations of publicity – permitting government to keep policies secret in order to avoid landing us in situations (such as prisoner’s dilemmas) where we anticipate we will behave self-destructively – are among those policies that will pass the publicity test at the second level. And so do violations of publicity designed to forestall perverse incentives to the wicked.

It is worth reflecting on why these examples violate the publicity principle at the first level but not at the second. In both cases, publicizing a maxim turns out to be self-frustrating, but not for reasons that seem to have anything to do with the morality of the maxim. That, precisely, is why policies of secrecy in these cases can pass the publicity test at the second level.27

The publicity principle thus seems to work as a test of morality only in the following circumstance: not merely is publicizing a maxim self-frustrating, but moral phenomena play a causal role in the chain of events leading to self-frustration. The arguments and examples we have looked at reveal two causal patterns that fit this description.

1. If the maxim is publicized, public disapproval plays a causal role in its self-frustration. Llewellyn’s and Nessen’s arguments that institutions of justice should strive for public acceptance rather than authentic justice would fail if publicized because the public disapproves of institutions of justice that do not strive for justice. In Kant’s treaty example, a nation that announced that it would violate treaties whenever it was in its interests to do so would fail to achieve useful treaties because other nations would mistrust it. (At some cost in precision, I am here regarding mistrust as a form of moral disapproval.)

2. If the maxim is publicized, public corruption plays a causal role in

27I came to appreciate this point thanks to the forceful objections that Peter Singer raised to an earlier version of my argument. Singer asked why the bare fact that a maxim would be self-frustrating if it were publicized should count as a moral objection to it, and made me see that I had no answer to the question.
its self-frustration. Here I have in mind Rawls’s argument against utilitarianism. If the utilitarian maxim became generally known, citizens would become demoralized – they would lose self-respect, leading in turn to the loss of respect for others that utilitarianism requires of its adherents. It is not necessarily true that the public would disapprove of utilitarianism: demoralized individuals are just as likely to blame themselves rather than the policy that has demoralized them. But a recognizably moral phenomenon – public corruption – plays a causal role in the self-frustration of utilitarian public morality.

Is there any formulation of the publicity principle that captures both these patterns? Let me suggest a feature they share. In both cases, public disapproval as well as public corruption, publicizing a maxim undercuts the legitimacy of the public institutions proposing to adopt the maxim. I mean the term “legitimacy” to refer to the institution’s capacity to garner empirical support based on its constituents’ moral commitment to it (or, in foreign relations, to garner its interlocutors’ respect). Let us, then, reformulate the publicity principle as follows:

All actions relating to the right of other human beings are wrong if publicizing their maxim would lead to self-frustration by undercutting the legitimacy of the public institutions authorizing those actions.

6.6 A Defense of the Publicity Principle: Second Step

This reformulation helps pare away counter-examples, but offers no affirmative case for the publicity principle. To my mind, the most persuasive affirmative argument for the publicity principle derives from considerations of popular sovereignty: an action or policy that cannot withstand publicity is one that cannot garner popular consent, and that is why the action is wrong.

But this argument raises the obvious question of what is so compelling about popular sovereignty. It seems clear that some issues are really better left to the experts, not the people – the Wise, not the Many. Those in public authority may be better informed, more intelligent, more moral, or wiser than the common denominator represented by the citizenry; and in those cases, Plato may have been right to insist that the virtuous ruler should do the right thing and then placate the ruled with noble lies. (Consider, once again, the case of a progressive South African government confronting an intransigently racist white electorate.)

I am persuaded that no a priori argument can dispose of this possibility. My response is rather to reformulate the question, translating it
out of the realm of a priori philosophical argument and into the realm of historical and political experience. Suppose we grant to Plato and Machiavelli the point on which both insist — that the Many are often wrong (ill-informed, incompetent, prejudiced) about the crucial issues of public life. Suppose we grant as well that some people, the Wise, may be right where the rest of us are steeped in error — indeed, not only may be right, but at countless moments in human history have been right in the face of popular error and thoughtlessness.

It simply does not follow that governments should be permitted to make policy in secret, placating the Many with noble lies. That conclusion would follow only if the Wise happen to coincide with those who occupy positions of official authority.

To put it another way, the empirical validity of the publicity principle turns not on whether the Many are ignorant or wrong-headed, but on whether their leaders are less ignorant or less wrong-headed. No doubt the Wise are few; and the leaders are few; but it hardly follows that the leaders are wise. Before we reject the publicity principle because the leaders know best, we must have reason to believe that the leaders know better. And to find that out, we must look carefully at the variety of mechanisms by which decision-making elites are actually selected. If actual selection mechanisms choose randomly between the Many and the Wise, or affirmatively disfavor the Wise, then the foolishness of the Many is irrelevant: the few in official positions have no reason to suppose that their policy brainstorms are any less foolish. It would be absurd to insist that officials should disregard the publicity principle merely because the Many view matters through a glass darkly. The officials may view matters through the same leaded glass.

These observations do not, of course, offer a defense of the publicity principle. Rather, they offer (in the most general terms, at that) a research program for assessing whether and when the publicity principle is defensible. How well do various selection mechanisms succeed in empowering the Wise? — political party systems, democratic elections, civil service examinations (such as the 2,000-year-long Imperial Chinese system), nepotism and hereditary officeholding, palace intrigue, the business/government revolving door, grooming through hautes écoles, . . . In those systems that succeed in selecting the Wise, the publicity principle is false; in those that offer only random or worse chances of selecting the Wise, the principle is true (in a sense that I shall explain shortly). Plainly, this is a difficult research program to execute or even to formulate operationally. How one defines “the Wise” may itself beg crucial questions.28

28One might object that the inquiry not only begs crucial questions, but contains
Is this all that can be said? Must we end with an evanescent question? I think not. Though I am insisting that the justification (or lack of justification) of the publicity principle turns on empirical inquiry, there is no reason not to offer an educated guess about what that inquiry will reveal. Let me sketch an argument in support of the conjecture that in most if not all selection systems the publicity principle is true.

It is important to realize that the character of those who seek and attain office in a political system depends in large measure on the moral expectations attached to the offices. Commentators on recent trends in American politics have worried that the voting public demands a level of personal rectitude, or even asceticism, that may be inconsistent with the extroverted vitality and "fire in the belly" needed for national office. Court watchers have groaned that intellectual achievements may disqualify candidates for the Supreme Court; numerous organizations have argued that current ethics-in-government laws discourage capable individuals from accepting appointment to government offices.

These are examples of a more general point: the "moral job descriptions" of leadership posts serve in effect as recruitment devices for those posts; they serve also as filtering devices on the kind of person who survives on the job. A moral principle vesting dictatorial powers in an office may attract mainly megalomaniacs; a moral principle licensing unfettered political ruthlessness may guarantee that only the ruthless survive in politics; a moral principle permitting noble lies may attract mainly liars, and not necessarily noble ones.

One possibility we cannot rule out is that a principle of political morality that could be exercised wisely by the right individuals will tend to recruit and retain precisely the wrong individuals, that is, individuals who are incapable of exercising it wisely. I wish to suggest that precisely this is the most likely result if we reject the publicity principle and adopt instead a principle permitting noble lies -- a principle in which government officials are licensed or encouraged to deceive the Many for the Many's own good. That principle makes sense if we can count on the superior wisdom and virtue of officials; but it seems much more plausible that those who would be attracted to public office by the unfet-

an in-built circularity, since political wisdom surely includes the ability to gain and retain political power in a given selection system. I reject this latter proposition, however. Our inquiry, remember, is whether existing selection mechanisms will pick officials who can make better (wiser, more virtuous) policy than the hoi polloi - indeed, policy so much better that its rationale must be concealed from the hoi polloi. There is absolutely no reason to presume that wisdom in this sense includes, or even correlates with, the ability to gain and retain political office. That, in fact, is what the inquiry is all about.
tered discretion to deceive are precisely those who are neither wiser nor more virtuous than the rest of us.

Viewed in this light, the publicity principle should be accepted as a principle of rational skepticism on the part of officials.\textsuperscript{29} Suppose, that is, that I am an official contemplating a certain policy. Suppose in addition that I believe that my policy would violate the publicity principle: if it were publicized, widespread public disapproval would undermine my legitimacy and frustrate my pursuit of cherished political goals. My first impulse is to conceal my policy, because I think that I know better than the wider public.

At this point, rational skepticism should lead me to reflect on how I came to occupy the office I now possess. Is it really because I'm such a smart cookie? Or is it that I knew somebody who knew somebody? Perhaps I made it on my own; but — the voice of rational skepticism reminds me — the skills that I used to make it on my own were maneuvering skills in turf wars coupled with opportunism. (In politics you use anything you can get your hands on.) In my nasty little corner of the bureaucratic world, wisdom and virtue may well be counteradaptive traits; if so, the fact that I outmaneuvered the Wise doesn't show that I am wise. Quite the contrary. So even if the Many are neither wise nor virtuous, the smart money says that I am one of them rather than one of the Platonic Few. Rational skepticism — the smart money — says that if publicizing my maxim would undercut my legitimacy, it is probably because my maxim is wrong.

One might object that supporters of noble lies need not assume that policymakers are wiser or more virtuous than the public at large. They need assume only that policymakers who have devoted long, patient hours to learning the intricacies of a complex issue know a lot more about it than the man or woman on the street. The relevant distinction is neither wisdom nor virtue, but accumulated experience and time on the learning curve. A former U.S. official who had worked on the sea treaty told me that it took him three years on the job to really understand the issues. Why, then assume that vox populi represents vox Dei? If the public must be deceived for its own good, that is simply because the public is ill informed, not that it is less wise or virtuous than the policymaker.

My reply to this objection is that there is a crucial difference between being sufficiently knowledgeable to set policy about complex issues and being sufficiently knowledgeable to render intelligent consent to it.

\textsuperscript{29}I owe this formulation to Michael Smith, whose astute but sympathetic criticisms of an earlier version of my argument improved it greatly.
What takes three years to learn may take only three minutes to explain. In fifteen years of law teaching, I have never encountered a legal or public policy issue that couldn’t be explained to a bright twelve-year-old in half an hour. That is not to say that the twelve-year-old would decide the issue intelligently, but it is to say that she would accurately understand why the issue is in dispute and what the chief arguments are on all sides. By supporting deception, opponents of the publicity principle assume not only that the public is less informed than policymakers, but that the public is for practical purposes ineducable. That, I think, really does amount to a presumption that policymakers are wiser and better than their fellows. Rational skepticism should lead us to doubt this presumption. If a policy would excite across-the-board moral condemnation, the reasonable conclusion is that, even if the public doesn’t know best, it probably knows better.

In that case, the publicity principle – like the Enlightenment ideal of popular sovereignty that it represents – would be justified by a version of Winston Churchill’s bromide about democracy: it is the worst form of government we can imagine, except for all the others.

References
Bickel, Alexander. 1962. The Least Dangerous Branch: The Supreme Court at the Bar of Politics. New Haven, Conn.: Yale University Press.
Fuller, Lon. 1964. The Morality of Law. 2nd ed. New Haven, Conn.: Yale University Press.
Habermas, Jürgen. 1989. The Structural Transformation of the Public Sphere:


