The Philosophical Quarterly
The Scots Philosophical Association and the University of St Andrews
The Philosophical Quarterly Vol. 63, No. 250 January 2013
ISSN 0031-8094 doi: 10.1111/j.1467-9213.2012.00106.x

REJECTING THE PUBLICITY CONDITION: THE INEVITABILITY OF ESOTERIC MORALITY

By Ben Eggleston[1]

It is often thought that some version of what is generally called the publicity condition is a reasonable requirement to impose on moral theories. In this article, after formulating and distinguishing three versions of the publicity condition, I argue that the arguments typically used to defend them are unsuccessful and, moreover, that even in its most plausible version, the publicity condition ought to be rejected as both question-begging and unreasonably demanding.

> It would be natural to want the best theory … not to be self-effacing. If the best theory was self-effacing, telling us to believe some other theory, the truth … would be depressingly convoluted. It is natural to hope that the truth is simpler: that the best theory would tell us to believe itself. But can this be more than a hope? Can we assume that the truth *must* be simpler? We cannot.[2]

I. INTRODUCTION

A familiar criticism of act utilitarianism rests on the following claim: a society of agents who believe that morality requires them always to act so as to maximise overall happiness would very likely be a less

[1] Much of this article is derived from a chapter of my dissertation. I am grateful to David Gauthier, my dissertation adviser, for his support, guidance, and comments on successive versions of that chapter. I would also like to thank Maura Tumulty, Alice Crary, and Karen Frost-Arnold for their comments on an earlier version of this article, and referees for this journal for their comments on the penultimate version of this article. I would especially like to acknowledge the contributions of Dale Miller, who provided comments on multiple versions of this article and helpful advice at several points in my work on this article.

[2] D. Parfit, *Reasons and Persons* (Oxford UP, 1984), p. 24.

Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

happy one than would a society of agents who subscribe to any of several competing moral theories, such as common-sense morality. The reasons for this claim are familiar: in contrast to agents who simply accept common-sense morality, agents in the thrall of act utilitarianism would presumably spend too much time calculating as to the consequences of various possible acts, would surely find their calculations frustrated or falsified by their inability to predict the consequences of their acts, and would at least occasionally be caused by the complexities of their calculations to succumb to the temptation to engage, whether consciously or unconsciously, in self-serving rationalisations. (For example: who will really suffer if I fudge my tax return in order to save enough money to buy a huge television, thereby improving my quality of life and stimulating the economy?) Somewhat more subtly, but perhaps more seriously, such agents' disposition to calculate and to optimise at every turn would threaten to deny them access to certain fruitful forms of interaction enjoyed by agents who accept, as reasons for action, certain non-utilitarian considerations such as those of honesty, friendship, and fidelity to one's word. In sum, it can be expected that a society of act-utilitarian agents would do worse, in terms of achieving the act-utilitarian aim of maximising happiness, than would a society of agents who subscribe to other moral theories, be they non-utilitarian ones or other utilitarian ones, such as rule utilitarianism.[3] We can mark this thought by saying that act utilitarianism is *self-defeating* (understanding this term, which can have many meanings, to refer simply to the fact just stated).

One question that naturally arises at this point is whether a self-defeating theory is *ipso facto* unacceptable. But here I shall set this issue aside, in order to pursue a different one that arises a little further downstream. For even if it can be shown that self-defeat does not render a

[3] Bentham dismisses this point summarily, writing 'Dangerous, to endeavour to do what is most useful? The proposition … is a self-contradictory one' (J. Bentham, 'A Fragment on Government', in J. Bentham, *A Comment on the Commentaries and A Fragment on Government*, edited by J.H. Burns and H.L.A. Hart [London: The Athlone Press, 1977], p. 516). But most authors who discuss act utilitarianism assert or grant this point. Some regard it as a reason to reject act utilitarianism; others maintain act utilitarianism in the face of it. See, for example, R.M. Hare, 'Principles' (in his *Essays in Ethical Theory* [Oxford UP, 1989], pp. 49–65), p. 60; Parfit, *Reasons and Persons*, pp. 27–8; D. Brink, *Moral Realism and the Foundations of Ethics* (Cambridge UP, 1989), p. 257; T.L.S. Sprigge, 'Utilitarianism and Respect for Human Life', *Utilitas*, 1 (1989), pp. 1–21, at p. 18; W.H. Shaw, *Contemporary Ethics: Taking Account of Utilitarianism* (Oxford: Blackwell Publishers, 1999), pp. 144–5; and B. Hooker, *Ideal Code, Real World: A Rule-consequentialist Theory of Morality* (Oxford UP, 2000), pp. 142–3.

theory unacceptable, the fact of self-defeat in the case of a theory such as act utilitarianism (assuming it is a fact in the case of act utilitarianism) has a further implication that raises a new concern. The implication goes like this: if happier outcomes result from agents' subscribing to some other moral theory than act utilitarianism, then—given act utilitarianism's characteristic insistence on agents' bringing about the happiest outcomes—act utilitarianism itself will enjoin agents to subscribe to some other moral theory than itself. This raises a new question about the acceptability of act utilitarianism (and, in principle, some other theories): can a moral theory be acceptable if it enjoins agents to subscribe to some other moral theory than itself?

Before arguing for an affirmative answer to this question, which is my principal aim in this article, I want to clarify the difference between it and the earlier question of whether a moral theory can be acceptable if it is self-defeating. That earlier question is not concerned with what a moral theory requires of agents, if it is in fact self-defeating; rather, it asks about the significance, for the acceptability of the theory, of the phenomenon of self-defeat itself. In contrast, the question on which I want to focus does not ask about the significance of self-defeat itself; rather, it asks about the significance, for the acceptability of the theory, of a theory's responding to that phenomenon by enjoining agents to subscribe some other theory.

That act utilitarianism responds in this way has long been (like the issue of self-defeat itself) a leading concern of act utilitarianism's proponents and critics alike. Indeed it is pointedly articulated in one of the more memorable passages of Sidgwick's pioneering *The Methods of Ethics*—a passage that has become the *locus classicus* for this issue:

> on Utilitarian principles, it may be right to do and privately recommend, under certain circumstances, 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. … And so 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.[4]

These consequences, which Bernard Williams aptly says that Sidgwick 'pursued with masochistic thoroughness,'[5] have been echoed by subse-

[4] H. Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan and Company, Limited, 1907; reprinted Indianapolis, Ind.: Hackett Publishing Company, 1981), pp. 489–90.

[5] B. Williams, *Ethics and the Limits of Philosophy* (Harvard UP, 1985), p. 108.

quent writers on act utilitarianism.[6] And they raise the question of whether a moral theory must be regarded as unacceptable if it violates a requirement that it may seem natural to impose on moral theories—a requirement that may be thought of as the *publicity* condition.

In this article, I shall argue that the publicity condition is not a reasonable requirement to impose on moral theories. To advance this claim, I shall begin, in section II, by formulating and distinguishing several versions of the publicity condition and by assembling some evidence of the validity that this requirement is widely thought to have. That will set the stage for a critique, in section III, of some of the considerations that may seem to justify the publicity condition and, in section IV, for the presentation of two independently sufficient refutations of it. After addressing objections in sections V and VI, I shall offer some concluding reflections in section VII. Along the way, act utilitarianism will continue to serve as a central example. Several authors I quote refer to it simply as 'utilitarianism,' and I'll do likewise, after acknowledging here that other forms of utilitarianism, such as rule utilitarianism, would, in many cases, have to be discussed quite differently.

II. VERSIONS AND VIOLATIONS OF THE PUBLICITY CONDITION

II.1. Having only gestured at the general idea of the publicity condition in the introductory section, I endeavour in this section to formulate and to distinguish three versions of this requirement. We can approach the

[6] As one might expect, the same authors—if not always the same passages—can be cited here as were cited in footnote 3 in regard to the self-defeating character of act utilitarianism. See, for example, Hare, *Freedom and Reason* (Oxford UP, 1963), p. 44, Hare, 'Ethical Theory and Utilitarianism' (in *Essays in Ethical Theory*, pp. 212–30), p. 227, and Hare, 'Utilitarianism and the Vicarious Affects' (in *Essays in Ethical Theory*, pp. 231–44), pp. 232–3 and pp. 239–40; Parfit, *Reasons and Persons*, pp. 40–1; D. Brink, *Moral Realism and the Foundations of Ethics*, p. 257; Sprigge, 'Utilitarianism and Respect for Human Life', pp. 12–4, and Sprigge, 'The Greatest Happiness Principle', *Utilitas*, 3 (1991), pp. 37–51, at p. 37; Shaw, *Contemporary Ethics: Taking Account of Utilitarianism*, pp. 149–50; and Hooker, *Ideal Code, Real World*, pp. 143–4. Additionally, see A.M.S. Piper, 'Utility, Publicity, and Manipulation', *Ethics*, 88 (1978), pp. 189–206, at p. 196; M. Perkins and D.C. Hubin, 'Self-Subverting Principles of Choice', *Canadian Journal of Philosophy*, 16 (1986), pp. 1–10, at p. 8; and Hubin, 'The Moral Justification of Benefit/Cost Analysis', *Economics and Philosophy*, 10 (1994), pp. 169–94, at pp. 173–4.

It should not be thought, though, that there is unanimity on this point. For dissenting remarks, downplaying Sidgwick's worry, see J.J.C. Smart, 'Extreme and Restricted Utilitarianism', *The Philosophical Quarterly*, 6 (1956), pp. 344–54, at p. 348; and W.L. Langenfus, 'Implications of a Self-Effacing Consequentialism', *Southern Journal of Philosophy*, 27 (1989), pp. 479–93, at pp. 481–2. There are also complications within Brink's view; in addition to his text cited above, see his *Moral Realism and the Foundations of Ethics*, p. 261.

task of formulating these distinct versions by imagining some of the ways in which a moral theory may run afoul of the general idea of the publicity condition. A moral theory runs afoul of this general idea in a particularly flagrant way if the theory implies, in certain circumstances, that *every* agent in the group to which it applies (such as a particular society, or all rational creatures) ought not to subscribe to it. Borrowing a term found in the passage from Parfit's *Reasons and Persons* that I have chosen as the epigraph for this article, let us call such theories *self-effacing*. Then one version of the publicity condition may be formulated as follows:

> *The ban on self-effacing theories*: A moral theory is unacceptable if circumstances may arise in which it requires every agent in the group to which it applies not to subscribe to it.

So, to satisfy this requirement, all a theory needs to do is always (that is, in all circumstances) allow some agent or agents in the group to subscribe to it—even if it also sometimes or always (that is, in some circumstances or all circumstances) implies that some of those agents ought not to subscribe to it. If a moral theory violates this requirement, then it runs afoul of the general idea of the publicity condition in a particularly flagrant way.

But there are other ways in which a moral theory may run afoul of the general idea of the publicity condition. For example, even if a moral theory does not require *every* agent in the group to which it applies not to subscribe to it, it may still be thought to run afoul of the general idea of the publicity condition in some way if it ever requires even *some* agents not subscribe to it. Borrowing a term from Sidgwick, let us say that such theories are *esoteric*. This suggests another, more demanding, version of the publicity condition, which can be formulated by replacing the word 'every' in the ban on self-effacing theories with the word 'some':

> *The ban on esoteric theories*: A moral theory is unacceptable if circumstances may arise in which it requires some agent in the group to which it applies not to subscribe to it.

To satisfy this requirement, then, a theory must always (that is, in all circumstances) allow *every* agent in the group to subscribe to it. This requirement is obviously more demanding than the ban on self-effacing theories. It follows that self-effacing theories form a subset of esoteric theories: some esoteric theories are self-effacing, while the rest are, we might say, only partially self-effacing (which of course does not count as being self-effacing as defined above).

Perhaps no recent author is more responsible for drawing attention to the publicity condition—indeed, for causing it to be referred to in this way—than John Rawls. And it might be thought that the publicity condition as formulated by Rawls is equivalent to either the ban on self-effacing theories or the ban on esoteric theories. But although Rawls does imply an endorsement of the ban on esoteric theories, he does so in his discussion of what he calls the universality condition. There he writes that

> a principle is ruled out if it would be self-contradictory, or self-defeating, for everyone to act upon it. … Principles are to be chosen in view of the consequences of everyone's complying with them. (*ATOJ*, p. 132)[7]

What he conceives of as the publicity condition goes further:

> A third condition [after generality and universality] is that of publicity, which arises naturally from a contractarian standpoint. The parties assume that they are choosing principles for a public conception of justice. They suppose that everyone will know about these principles all that he would know if their acceptance were the result of an agreement. Thus the general awareness of their universal acceptance should have desirable effects and support the stability of social cooperation. The difference between this condition and that of universality [which, as just noted, implies the ban on esoteric theories] is that the latter leads one to assess principles on the basis of their being intelligently and regularly followed by everyone. But it is possible that all should understand and follow a principle and yet this fact not be widely known or explicitly recognized. (*ATOJ*, p. 133)

This is the last, and the most demanding, of the versions of the publicity condition that we shall set out. The distinguishing feature of this condition, which I shall refer to as *Rawls's publicity condition*, is expressed in Parfit's observation that it requires of a theory that 'it must be a theory that everyone ought to accept, *and publicly acknowledge to each other*' (*Reasons and Persons*, p. 43, emphasis added).[8]

[7] References of the form '*ATOJ*, p. ___' are to pages of J. Rawls, *A Theory of Justice* (Harvard UP, 1971). Incidentally, what Rawls means by 'self-defeating' seems to be somewhat more dire than the meaning of 'self-defeating' used above, but we need not pursue this matter.

[8] Similarly, Hare writes that 'the so-called publicity requirement by which Rawls and others set store' requires of a moral theory 'that it could be *openly avowed* without defeating its object' (Hare, *Sorting Out Ethics* [Oxford UP, 1997], p. 124, emphasis added).

The foregoing account of Rawls's publicity condition is from Rawls's *A Theory of Justice*. In his later 'Kantian Constructivism in Moral Theory', Rawls characterises his version of the publicity condition in still more detail. (J. Rawls, 'Kantian Constructivism in Moral Theory' [in his *Collected Papers* (Harvard UP, 1999), ed. by S. Freeman, pp. 303–58], pp. 324–7.) The revised edition of *A Theory of Justice* leaves unchanged the passages I've quoted. (*A Theory of Justice*, rev. ed. [Harvard UP, 1999].)

As Rawls and Parfit indicate, Rawls's publicity condition requires more of a theory than that it not be esoteric (not to mention self-effacing). So Rawls's publicity condition is the most demanding of the three versions of the publicity condition specified here, with the ban on esoteric theories being less demanding and the ban on self-effacing theories being the least demanding. We noted earlier that self-effacing theories form a subset of esoteric theories; those, in turn, form a subset of theories that violate Rawls's publicity condition. Although these relationships may be straightforward enough not to need a diagram in order to order to be fully understood, it will be convenient later to be able to refer to the following figure.

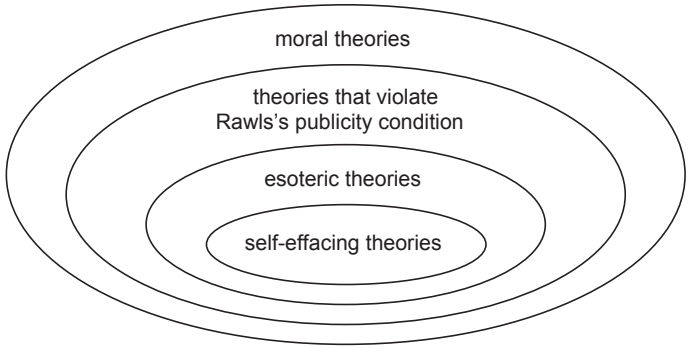

Figure 1

So in its least demanding version the publicity condition rejects only those theories in the innermost oval (the one for self-effacing theories); in an intermediate version it rejects all theories in the 'esoteric theories' oval, and in its most demanding version—Rawls's publicity condition—it rejects all theories except those outside the next-to-largest oval.

II.2. Although the publicity condition—in any of the versions just specified—applies to moral theories of all kinds, it arises most conspicuously in regard to utilitarian ones. Sidgwick anticipated this aspect of the debate that would ensue over utilitarianism, writing that his 'conclusions [the ones quoted above, in section I] are all of a paradoxical character' (*The Methods of Ethics*, p. 489) and that 'there is no doubt that the moral consciousness of a plain man broadly repudiates the general notion of an esoteric morality' (*The Methods of Ethics*, pp. 489–90).

Critics of utilitarianism have been only too ready to amplify this note of unease. Kurt Baier, for example, claims that

> An esoteric code, a set of precepts known only to the initiated and perhaps jealously concealed from outsiders, can at best be a religion, not a morality. … 'Esoteric morality' is a contradiction in terms.[9]

Similarly, D.H. Hodgson concludes that if certain assumptions are granted, 'it would mean simply that universal and correct application of act-utilitarianism could not persist, because it would involve rejection of act-utilitarianism by at least some persons,' and he implies that this result would discredit the theory.[10] Somewhat more bluntly, Williams affirms the 'capacity for utilitarianism … to annihilate itself' on this basis.[11] Nicholas Rescher, finally, claims that 'it would surely put the utilitarian in an untenable position to concede that his moral theory is not self-sustaining, that it enjoins him to teach and foster a moral theory at variance with itself.'[12] The publicity condition has probably been used to criticise utilitarianism more than any other moral theory.[13]

III. THE CASE FOR THE PUBLICITY CONDITION

III.1. It is only to be expected that a requirement that is frequently invoked to dispose of widely debated moral theories (such as utilitarianism) should itself become the object of extensive debate. So it should come as no surprise to find that, as Brad Hooker writes, 'The literature on the "publicity condition" is voluminous.'[14] In this section, I shall survey a representative sample of this literature with a view to rehearsing, and

[9] K. Baier, *The Moral Point of View: A Rational Basis of Ethics*, abridged ed. (New York: Random House, 1965), p. 101.

[10] D.H. Hodgson, *Consequences of Utilitarianism: A Study in Normative Ethics and Legal Theory* (Oxford UP, 1967), p. 46.

[11] B. Williams, *Morality: An Introduction to Ethics* (Cambridge UP, 1972), p. 95.

[12] N. Rescher, *Unselfishness: The Role of the Vicarious Affects in Moral Philosophy and Social Theory* (University of Pittsburgh Press, 1975), p. 79.

[13] It is an interesting irony that Rawls, even while construing the publicity condition in the strongest of the three versions earlier distinguished, seems to imply that utilitarianism may satisfy it. Referring to the set of constraints containing the publicity condition, he writes, 'I assume that they are satisfied by the traditional conceptions of justice' (*ATOJ*, p. 131). Presumably he counts utilitarianism among these, since he refers to it as '[d]uring much of modern moral philosophy the predominant systematic theory' (p. vii) and includes it on the list of the alternatives among which the parties in the original position have to choose (p. 124). Later, he notes that 'utilitarianism, as I have defined it, is the view that the principle of utility is the correct principle for society's *public* conception of justice' (p. 182, emphasis added).

[14] Hooker, *Ideal Code, Real World*, p. 85, n. 14. See also B. Hooker, 'Rule-Consequentialism', *Mind*, 99 (1990), pp. 67–77, at p. 72, n. 20.

displaying the inadequacy of, certain common putative justifications for the publicity condition.

III.2. One defence of the publicity condition is suggested by the claim of Baier's quoted above: '"Esoteric morality" is a contradiction in terms.' But as Samuel Scheffler explains in discussing the bearing of the publicity condition on utilitarianism and other forms of consequentialism, even if a definitional claim such Baier's could be substantiated, it would not endow that requirement with the force and scope that its defenders intend for it to have:

> If 'morality' is defined in such a way as to include the publicity condition, and if a thoroughgoing consequentialism dispenses with the publicity condition, then talk about the relative merits of consequentialist and non-consequentialist moral principles can simply be recast as talk about the relative merits of consequentialist principles on the one hand and moral principles on the other. By itself, no simple appeal to meaning is capable of showing that there is something wrong with consequentialism's apparent willingness to violate the publicity condition.[15]

Indeed it seems that an appeal to meaning is bound to be unconvincing except to someone already convinced of the propriety of the requirement in question.

III.3. It is notable that Rawls, who (as we saw) defends the publicity condition even in the strongest of the three versions distinguished in the last section, declines to deploy a definitional argument. Instead, he writes that

> There are certain formal conditions that it seems reasonable to impose on … conceptions of justice. … I do not claim that these conditions follow from the concept of right, much less from the meaning of morality. (*ATOJ*, p. 130)

Then, as if to emphasise the difference between his approach and Baier's, Rawls adds that 'by itself, a definition cannot settle any fundamental question' (p. 130) and that it is 'necessary that the conditions not be justified by definition or the analysis of concepts, but only by the reasonableness of the theory of which they are a part' (p. 131).

Rawls's approach, being more modest than Baier's, is not vulnerable to precisely the same reply as Baier's. But in its modesty it is vulnerable to an even simpler reply. As Scheffler writes,

[15] S. Scheffler, *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions* (Oxford UP, 1982), p. 47.

> once it is said that the condition is just something it 'seems reasonable' to expect an acceptable moral conception to satisfy, and that the adequacy of the condition must ultimately be assessed in the light of the moral conception it leads us to, the consequentialist can simply deny that the condition 'seems reasonable' to him. (*The Rejection of Consequentialism*, p. 47)

So although Rawls consciously avoids the lure of Baier's definitional argument, he does not manage to replace it with anything stronger.[16]

III.4. A third approach to defending the publicity condition begins with the following observation: a moral theory that violates the publicity condition is one that requires agents to cultivate and to maintain beliefs that the theory itself implies are false—for example, beliefs about what is the best theory of morality, or (what will be implied by such beliefs) beliefs about what acts are right and wrong.[17] This observation, when conjoined with the thought that any moral theory that can be so described must *ipso facto* be unacceptable, implies the unacceptability of any moral theory that violates the publicity condition.

To be sure, this thought has some intuitive appeal. For it is natural to think, especially from a philosophical point of view, that the only beliefs that we can have good reasons for cultivating and maintaining are *true* ones.[18] And so while we are accustomed in philosophy to encountering theories (moral and otherwise) that have implications that *we* regard as false, we may react with particular suspicion to a theory that recommends beliefs that *it* implies are false. Such a theory may seem not only mistaken in some way, as do so many philosophical theories, but also—and more seriously—guilty of some sort of philosophical bad faith.

But this assessment should look less appealing in the light of the following fact: such a theory (one that recommends beliefs that it implies are false) needn't be guilty of the logical sin of implying that those beliefs are

[16] Given the weight that Rawls puts on the notion of the 'reasonable' in his later work—especially in his lectures on Kantian constructivism (cited above, in footnote 8)—it might be thought that Rawls's earlier assertion of the reasonableness of the publicity condition is more pregnant with meaning than I acknowledge. But in my view, Rawls's later remarks on the notion of the reasonable fail to provide the ingredients for a fuller or further argument for the publicity condition.

[17] That an agent's acceptance of a theory involves not only her values and motivations, but also her beliefs, is emphasised by Langenfus, 'Implications of a Self-Effacing Consequentialism', p. 479.

[18] For a characteristically emphatic expression of this philosophical proclivity, see Moore's remark that 'What I am concerned with is knowledge only—that we should think correctly and arrive at some truth, however unimportant' (G.E. Moore, *Principia Ethica*, rev. ed., ed. by Thomas Baldwin [Cambridge UP, 1993], p. 115).

true (something that would be a logical sin because we have already supposed that the theory in question implies that those beliefs are *false*). Rather, the theory may recommend those beliefs for reasons that do not presuppose that those beliefs are true. For example, the theory may recommend those beliefs not for epistemic reasons, but for pragmatic, or practical, reasons—such as moral reasons.[19] As Parfit writes,

> there are two questions. It is one question whether some theory is the one that we *ought morally* to try to believe. It is another question whether this is the theory that we *ought intellectually* or *in truth-seeking terms* to believe—whether this theory is the true or best justified theory. (*Reasons and Persons*, p. 43)

Once these two questions are distinguished, then a theory that recommends beliefs that it implies are false may continue to seem (as Sidgwick said) paradoxical, but it cannot be dismissed as unacceptable on logical grounds.

Distinguishing these two questions also exposes a fallacy implicit in a superficially clever, but ultimately misleading, apparent dilemma for defenders of utilitarianism and other theories that violate the publicity condition. Here is how Williams constructs the alleged dilemma:

> [I]f utilitarianism is true, and some fairly plausible empirical assumptions are also true, then it is better that people should not believe in utilitarianism. If, on the other hand, it is false, then it is certainly better that people should not believe in it. So, either way, it is better that people should not believe in it. (*Morality: An Introduction to Ethics*, p. 98)

Clearly, the 'better' in the first premise—the one that alludes to utilitarianism's violation of the publicity condition—is a pragmatic 'better', not an epistemic one. And the reverse is true of the 'better' in the second premise—the one predicated on utilitarianism's falsity. Thus, no matter which way

[19] The distinction between epistemic and pragmatic reasons is, of course, familiar. When, in Hume's *Dialogues Concerning Natural Religion*, Cleanthes argues that we ought to believe in the existence of God because that is the best explanation of the order, complexity, and other apparent marks of design that we observe in the world around us, he is giving us an epistemic reason. When he claims that 'The doctrine of a future state is so strong and necessary a security to morals that we ought never to abandon or neglect it,' then he is giving us a pragmatic reason. (D. Hume, *Dialogues Concerning Natural Religion and the Posthumous Essays Of the Immortality of the Soul and Of Suicide*, ed. by Richard Popkin [Indianapolis: Hackett Publishing Company, 1980], p. 82 [ch. XII, par. 10]).

In pointing out that a theory of the kind under discussion may avoid logical inconsistency by appealing to pragmatic rather than epistemic reasons, I do not mean to imply that there cannot also be epistemic reasons for holding false beliefs. For example, there may be epistemic reasons for holding those beliefs that best enable one to acquire true beliefs—even if some or all of those enabling beliefs are themselves false. For further discussion, see J. Heil, 'Believing What One Ought', *Journal of Philosophy*, 80 (1983), pp. 752–65, at pp. 754–7, and J. Heil, 'Believing Reasonably', *Noûs*, 26 (1992), pp. 47–61, at pp. 47–8.

the 'better' in the conclusion is disambiguated, the conclusion will lose the support of at least one of its two premises, and will not remain standing.[20]

III.5. In this section, I have outlined and refuted three lines of defence for the publicity condition: that 'esoteric morality' is a contradiction in terms (section III.2), that the publicity condition is a reasonable one to impose on moral theories (section III.3), and that a theory is unacceptable if it recommends beliefs that it implies are false (section III.4). So several of the leading putative justifications for the publicity condition ultimately fail to provide it with solid support.

IV. TWO REFUTATIONS OF THE PUBLICITY CONDITION

IV.1. The publicity condition is not only, as I argued in the last section, in need of a sound justification. It is also vulnerable to direct attack, as a demonstrably unreasonable requirement to impose on moral theories. In this section, I offer two separate arguments, each of which I contend is sufficient to refute the publicity condition even in its weakest, and hence most plausible, version: the ban on self-effacing theories.

IV.2. First, the publicity condition is simply question-begging against any theory that violates it. To see this, recall what it means for a theory to be self-effacing: a theory is self-effacing if it sometimes requires every agent in the group to which it applies not to subscribe to it. But how, exactly, would a theory require this? Note that subscribing or not subscribing to a theory is not like stealing or not stealing a bicycle. Normally one can just choose, at will, to steal or not to steal; but normally one cannot not just choose, at will, to subscribe or not to subscribe to a particular theory. This is because subscribing to a particular theory is to be in a complex mental state, with both doxastic and affective aspects. Some meta-ethical theories emphasise the doxastic aspect of this state, while others focus on its affective aspect. Regardless of which (if either) of these approaches is correct, neither the doxastic aspect nor the affective aspect is normally a straightforward matter of choice. The doxastic aspect is normally not a

[20] Admittedly, the premises do support the following inference: that there is *some* sense of 'better' in which it is better that people not believe in utilitarianism. But utilitarianism itself already implies this, insofar as it admits that it is morally better that people not subscribe to it. It should also be admitted that Williams's remarks may well be enthememetic for an argument that cannot be dismissed so easily. For another reaction to Williams's blurring of the pragmatic and the epistemic, see Scheffler, *The Rejection of Consequentialism*, p. 51.

straightforward matter of choice because normally one cannot just choose, at will, to believe some proposition or other. For example, normally one cannot just choose, at will, to believe that the earth is flat. One can just choose, at will, to assert that proposition (as one might do in order to please or irk another person), but that is not the same thing. Similarly, the affective aspect is normally not a straightforward matter of choice because normally one cannot just choose, at will, to find certain kinds of acts or outcomes motivating. For example, normally one cannot just choose, at will, to feel motivated to bestow especially good treatment on people born in years that are divisible by 3. Because subscribing to a particular theory is to be in a complex mental state composed primarily of aspects that are normally not matters of choice, normally one cannot just choose, at will, to subscribe or not to subscribe to a particular theory. So, for a theory to require every agent in a group not to subscribe to it must not mean for it to require every agent in that group to simply choose to refrain from subscribing to it, as it might require every agent in that group to simply choose to refrain from stealing bicycles.

What, then, *does* it mean for a theory to require every agent in the group to which it applies not to subscribe to it? If subscribing or not subscribing to a particular theory is not something than an agent can simply choose to do, then what can it mean for a theory to require such subscription or non-subscription? The key to answering this question lies in seeing that although there may be no way in which agents can *directly* control their states of subscription and non-subscription (as we might call them), there are familiar ways in which they can *indirectly* control them. For example, on a large scale, educational and other social institutions can be set up in certain ways, to expose people to arguments for certain views (focusing on the doxastic element mentioned above) or to inculcate certain tastes, preferences, and motivations (focusing on the affective element mentioned above). On a smaller scale, people can decide to subject themselves to certain influences. For example, sometimes people go to church, or to therapy, in order to come to value things they do not yet value (though they want to value them) or to see certain things in a way they do not yet see them (though, again, they want to see them that way). It is also important to keep in mind that persons may exercise such influence over others and not just over themselves. In the scenarios discussed below, the specific causal mechanisms are either elided or simplified to focus on other issues, but the foregoing account may help to provide the broad outlines of the kinds of causal mechanisms that might often be at work.

With this background in place, we are in a position to say what it means for a theory to require every agent in the group to which it applies not to subscribe to it. It is for the theory to require some agent or agents to perform some act or acts that would cause (whether intentionally or not, and whether singly or collectively) every agent in the group not to subscribe to it, or—and this is a mouthful, but it's just another way of saying the same thing—by forbidding some agent or agents to perform the only act(s) that would cause it to continue to be the case that some agent in the group subscribes to it. For example, an agent may be situated such that one of the acts open to her would result in every agent's not subscribing to some theory, and that theory may happen to select *that* act as the one that she ought to perform. Or an agent may be situated such that only some of the acts open to him would save some theory from being effaced, and that theory may happen to forbid him to perform any of those. For our purposes the important aspect of all this is that if and when a theory violates the publicity condition, it does so in virtue of the *content of its prescriptions*. The publicity condition, then, amounts to a substantive constraint on the prescriptions that a theory may issue.

This understanding of the publicity condition has an important implication: the publicity condition, by discriminating among theories on the basis of the content of their prescriptions, begs the question of what the correct prescriptions of morality (or the prescriptions of the best theory of morality) are. As Railton writes,

> any such condition would be question-begging against consequentialist theories, since it would require that one class of actions—acts of adopting or promulgating an ethical theory—*not* be assessed in terms of their consequences. ('Alienation, Consequentialism, and the Demands of Morality', p. 155)

Similarly, Brink writes that

> Construed as a formal or conceptual claim … the publicity condition simply begs the question against teleological moral theories. Whether the true moral theory should be recognized, taught, or recommended as a decision procedure is itself a practical question the answer to which, the teleologist claims, depends on the intrinsic and extrinsic value that this sort of publicity produces. (*Moral Realism and the Foundations of Ethics*, p. 260)

Brink's conclusion sums up the essential point: 'The publicity constraint, therefore, must be construed as a substantive moral claim' (p. 260).

The point is not, of course, that a substantive moral claim cannot bear on the evaluation of a moral theory. On the contrary, such a claim may be eminently relevant to the evaluation of a moral theory.

But any such claim offered as dispositive of a moral theory, consequentialist or otherwise, ought to be embedded in a competing moral theory—or at least a sketch of one. For if such a claim is asserted as a freestanding and incontrovertible moral truth—as the publicity condition is when it is asserted as a freestanding requirement that it is reasonable to impose on moral theories—then it is bound to be question-begging.

IV.3. My second refutation of the publicity condition culminates in an equally simple claim: that this requirement is unreasonably demanding. The argument proceeds by showing that the range of moral theories that violate the publicity condition is wider than one might have initially expected. Indeed, although it has been claimed that only utilitarianism violates the publicity condition,[21] the argument below shows that nearly all moral theories violate it. This result, when coupled with the thought that only an unreasonably demanding requirement rejects so wide a range of the available moral theories, underwrites the conclusion that the publicity condition is an unreasonably demanding requirement to impose on moral theories.

To begin to see how wide the range of theories that violate the publicity condition is, consider a class of theories that we may call *disaster-avoiding theories*: those that, while not requiring agents to bring about the best outcomes they can, are just responsive enough to consequences to include a *disaster-avoidance provision* requiring agents to avert disasters if they can do so without suffering great costs themselves—with a disaster being understood to be an outcome that is much worse than every alternative (with the meaning of 'much' to be specified, of course).[22] This class of theories, while containing standard forms of consequentialism such as act- and rule-based versions,[23] also contains many theories that would not be regarded as forms of consequentialism, since the

[21] Piper, 'Utility, Publicity, and Manipulation', p. 192.

[22] A class of theories similar to this one, but not as quite broad, is defined by Kavka for a different purpose. G.S. Kavka, 'Some Paradoxes of Deterrence', *Journal of Philosophy*, 75 (1978), pp. 285–302, at p. 287. Closer to the thrust my argument, rule-conseqentialism's status as a disaster-avoiding theory is appealed to by Katarzyna de Lazari-Radek and Peter Singer in order to argue that Hooker must qualify his claim that his rule-consequentialism rules out the kind of esoteric morality that Sidgwick contemplates. K. de Lazari-Radek and P. Singer, 'Secrecy in Consequentialism: A Defence of Esoteric Morality', *Ratio*, 23 (2010), pp. 34–58, at p. 47.

[23] That act consequentialism requires the avoidance of disasters is entailed by the directly maximising character of that theory. On the entailment of a disaster-avoidance provision by indirectly maximising forms of consequentialism such as rule consequentialism, see Hooker, *Ideal Code, Real World*, pp. 98–9.

disaster-avoidance provision shared by these theories is so *weakly* responsive to consequences. A recognition of just *how* weakly consequentialist the disaster-avoidance provision is emerges from a consideration of three features of it. First, not only does the provision not require agents to bring about the best possible outcomes; it does not even require them to avoid the worst possible outcomes, except in those cases in which the worst possible outcome is *much* worse than every alternative. Second, it does not require agents to avoid the worst possible outcome whenever the worst possible outcome is much worse than *some* alternative, but only when the worst possible outcome is much worse than *every* alternative. Third, it excuses agents from this requirement whenever fulfilling it would require them to shoulder heavy burdens. So the disaster-avoidance provision is a very weakly consequentialist principle—weak enough, in fact, to certainly be included in such non-consequentialist moral views as common-sense morality.

The class of disaster-avoiding theories, then, is a broad one, including non-consequentialist theories as well as consequentialist ones. But, remarkably, *every* theory in this class violates the publicity condition. To see this, let T be some theory in this class. Now suppose that an agent finds himself in a situation in which he has only two options, with one option being not only worse than the other according to T, but also *enough* worse than the other for T to count it as a disaster. Since T is (*ex hypothesi*) a disaster-avoiding theory, T requires the agent to choose the second option (the disaster-avoiding one). But suppose also that the second option involves—either as a means to its intended result or as a side-effect—causing everyone in the group to which T applies not to subscribe to T. Then T, by requiring the agent to choose the second option, violates the publicity condition. Thus, every disaster-avoiding theory violates the publicity condition.[24]

IV.4. Already it is clear that the publicity condition is more demanding than it might have initially seemed, since it rejects non-consequentialist theories such as common-sense morality as well as consequentialist ones

[24] Although it must be admitted that situations with the structure just described are unlikely to arise in practice, this does not defeat the logical point stated in the text, which requires only their bare possibility. Illustrative examples, though conceivable, tend to be elaborate. See, for example, Perkins and Hubin, 'Self-Subverting Principles of Choice', p. 8. Perhaps the most helpful example is provided by de Lazari-Radek and Singer; see their 'Secrecy in Consequentialism: A Defence of Esoteric Morality', p. 49. Strictly speaking this latter example involves a person causing others to stop subscribing to a particular moral rule, not a whole moral theory, but it could be adapted to illustrate the latter sort of action as well.

such as various forms of utilitarianism. But one might still think that it is not *unreasonably* demanding, on the ground that it can be satisfied by certain moral theories that abjure consequentialist considerations altogether. One might think, for example, that Kant's moral theory satisfies the publicity condition, not only because of its rigorously non-consequentialist character, but also because Kant is explicitly credited with having developed a moral theory in the spirit of the publicity condition. Rawls, for example, writes that 'The publicity condition is clearly implicit in Kant's doctrine of the categorical imperative insofar as it requires us to act in accordance with principles that one would be willing as a rational being to enact as law for a kingdom of ends' (*ATOJ*, p. 133). So it would be telling indeed if Kant's moral theory could be shown to violate the publicity condition.

But Kant's moral theory can be shown to do just this, by way of an argument analogous to the one offered in reference to disaster-avoiding theories. Begin by supposing that some agent finds herself in a situation in which she has only two options, with one option being a textbook example of an act that violates the categorical imperative, such as lying. But suppose also that the second option involves (again, either as a means to its intended result or as a side-effect) causing all rational beings not to subscribe to Kant's moral theory.[25] Then, ironically, not even Kant's moral theory—which is thought to be the natural home of the publicity condition—turns out to satisfy it.[26] I further discuss Kantian approaches in section V.2, below.

IV.5. It might appear that the conclusion towards which we are driving is that *no* moral theory satisfies the publicity condition. Indeed this claim seems to be the upshot of Brink's statement that 'For any moral theory, there are possible circumstances in which its recognition and application would satisfy the theory worse than recognition and application of some alternative theory' (*Moral Realism and the Foundations of Ethics*, p. 261). But we cannot endorse a claim quite this strong. For there is one class of moral theories that satisfy the publicity condition—though before specifying it I should mention that it is a rather trivial one (and one that, there-

[25] As before (see footnote 24), although a case of this kind is unlikely to arise in practice, the bare possibility of one is sufficient for our purposes. For another account of how Kant's moral theory violates the publicity condition—an account based on a rather different argumentative strategy—see Brink, *Moral Realism and the Foundations of Ethics*, p. 261, n. 25.

[26] A further irony is that Kant himself alerts us to the possibility that the publicity condition may have surprisingly strong substantive implications, in his derivation from it of the rather strong claim that rebellion is always wrong. I. Kant, 'Toward Perpetual Peace', in I. Kant, *Practical Philosophy*, trans. and ed. by M.J. Gregor (Cambridge UP, 1996), p. 348.

fore, Brink and others may quite reasonably have regarded as not worth recognising). This class consists of those moral theories that specifically and absolutely forbid agents to act in a way that causes their effacement. Such theories satisfy the publicity condition either by requiring *only* that agents not act in a way that causes their effacement, or by imposing other duties on agents but making those other duties lexically subordinate to the duty of non-effacement.[27]

But no *other* theories than these satisfy the publicity condition. To see this, consider one last argument of the form already used in reference to disaster-avoiding theories and Kant's moral theory. Let T be some theory that not only imposes on agents some duty other than that of non-effacement, but also neglects to make this other duty lexically subordinate to that of non-effacement (either by requiring non-effacement but neglecting to give it lexical priority, or by neglecting to require non-effacement at all). This means that there are circumstances in which this other duty outweighs, trumps, or otherwise takes precedence over that of non-effacement. Now suppose that an agent finds himself in such circumstances and, moreover, in a situation in which he has only two options, with one option being in violation of this other duty (say, the duty not to kill innocents). Then T requires the agent to choose the second option (e.g., the one that complies with that duty). But suppose also that the second option involves causing everyone in the group to which T applies not to subscribe to T. Then T, by requiring the agent to choose the second option, violates the publicity condition. Thus, any theory that does not give lexical priority to non-effacement violates the publicity condition.

IV.6. Let us now sum up the results of the last few sections. We have seen that the publicity condition rejects not just consequentialist moral theories, but also many others: any moral theory that is just responsive enough to consequences to include a very modest disaster-avoidance provision (section IV.3), a paradigmatically non-consequentialist moral theory such as Kant's (section IV.4), and, in fact, any moral theory that does not lexically prioritise non-effacement (section IV.5). How these theories and sets of theories are related to each other is shown in the following figure.

[27] Perkins and Hubin make a related point, noting that a principle can avoid violating the publicity condition by 'logically requir[ing] the action of accepting the principle' ('Self-Subverting Principles of Choice', p. 8, n. 5). Also see Hubin, 'The Moral Justification of Benefit/Cost Analysis', p. 173. Similarly, de Larazi-Radek and Singer note that the only kind of consequentialism that can avoid being committed to esoteric morality is one that 'ranks openness as an overriding intrinsic value' ('Secrecy in Consequentialism: A Defence of Esoteric Morality', p. 36).

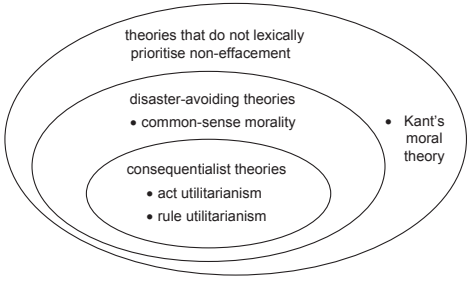

Figure 2

Now recall Figure 1, which shows how the various versions of the publicity condition are related to each other. Since it, like Figure 2, is a diagram of sets of theories, the two diagrams can be combined. The result is shown in the following figure.

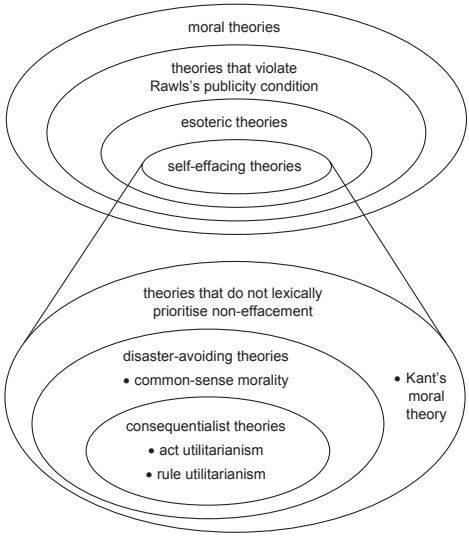

Figure 3

What this diagram illustrates, with its two lines connecting the first set of ovals to the second, is that the set of self-effacing theories *just is* the set of theories that do not lexically prioritise non-effacement (which is implied by our finding, above, that all and only theories that lexically prioritise non-effacement satisfy the ban on self-effacing theories). Since that version of the publicity condition rejects all of the theories in the innermost oval in the upper part of Figure 3, it rejects all of the theories in the lower part of Figure 3.

IV.7. The question that arises now is whether the publicity condition, being as demanding as the foregoing sections indicate, is a reasonable requirement to impose on moral theories. In essence, the question is this: is it reasonable to insist (as even the most plausible version of the publicity condition does) that a moral theory lexically prioritise non-effacement? To answer this question, let us consider what lexically prioritising non-effacement entails. It entails consigning all other values—whether standard consequentialist ones such as well-being or traditionally deontological ones such as being truthful and being respectful of others' lives and rights—to lexically subordinate positions. And this, I submit, is an unreasonable demand. Of course I cannot here offer a conclusive argument showing the importance of other values relative to non-effacement, but I can invite the reader to turn to any moral theory attracting widespread attention—any in Figure 2, at least—for then she will surely find an account of values that rejects the lexical priority of non-effacement (if it grants non-effacement any importance at all). Indeed the granting of lexical priority to non-effacement is such an extreme position that one might go so far as to say that not only does the publicity condition reject many reasonable theories, but also, the only theories it accepts are unreasonable ones.

V. ALTERNATIVE PERSPECTIVES: KANTIAN MORAL THEORY AND CIVIC REPUBLICANISM

V.1. It might be objected that the foregoing argument, however much it might be music to the ears of utilitarians and other consequentialists, is rather tin-eared about Kantian approaches and cousin approaches such as civic republicanism. These are substantive ethical views that take esoteric morality (and, *a fortiori*, self-effacing morality) to offend against something morally fundamental, and it might be objected that the foregoing argument fails to see, or at least fails to respond to the full force of, the reasons that might be given to justify their rejection of esoteric morality.

V.2. For Kantian approaches, the most immediately promising thought is what one might call the transparency principle: this is a meta-principle which says that in order for any first-order moral principle or procedure to be genuinely justifiable, it must be justifiable to each and every person —face to face, as it were. Perhaps the most familiar expression of this thought is T.M. Scanlon's claim that the principles of morality are those

that 'no one could reasonably reject.'[28] Although Scanlon's view differs from Kant's in certain key respects,[29] his claim effectively evokes the idea of justification *to* each and every person, as opposed to justification *simpliciter*. This idea is clearly relevant to the topic of esoteric morality because however justified (*simpliciter*) I might think I am in deceiving another person about the content of morality, it seems evident that I cannot justify this to her, face to face. And this, Kantians will say, reveals the immorality of such an action, regardless of how much good it might attain or how much harm it might avert.

This aspect of Kantian approaches is reinforced by Kantian conceptions of autonomy and of the nature of morality, which their expositors frequently contrast with consequentialist—indeed, specifically Millian—conceptions of these ideas. In regard to contrasting conceptions of autonomy, Onora O'Neill writes that

> Obligations not to deceive are more closely connected to Kant's rather than Mill's conception of autonomy. Kantian autonomy is a matter of acting on principles that can be principles for all, of ensuring that we do not treat others as lesser mortals – indeed victims – whom we disable from sharing our principles.[30]

In regard to contrasting conceptions of the nature of morality, Allen Wood mentions 'the idea (found in ch. 3 of Mill's *Utilitarianism*) that morality … is a mechanism of social coercion' and writes that

> Kantian morality, however—though the content of its duties may be socially oriented—is never about the social regulation of individual conduct. It is entirely about enlightened individuals autonomously directing their own lives.[31]

As this last phrase suggests, Kantian approaches to morality understand it as an aspect of practical reason: deciding what one ought to do.[32] Because of that, and because the capacity for engaging in such deliberation and autonomously acting accordingly is the source of the value that Kant finds in humanity,[33] deceiving another person about the content of morality is an especially egregious wrong. It might not result in the visible

[28] T.M. Scanlon, 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds.), *Utilitarianism and Beyond* (Cambridge UP, 1982), pp. 103–28, at p. 110. See also Scanlon, *What We Owe to Each Other* (Harvard UP, 1998), p. 4 and p. 153.

[29] Scanlon, *What We Owe to Each Other*, p. 6 and pp. 190–1.

[30] O. O'Neill, *A Question of Trust: The BBC Reith Lectures 2002* (Cambridge UP, 2002).

[31] A. Wood, 'The Final Form of Kant's Practical Philosophy', in M. Timmons (ed.), *Kant's Metaphysics of Morals: Interpretive Essays* (Oxford UP, 2002), pp. 1–21; see p. 9.

[32] See, for example, T.E. Hill, Jr., 'Kant's Theory of Practical Reason', in his *Dignity and Practical Reason in Kant's Moral Theory* (Cornell UP, 1992), pp. 123–46.

[33] See, for example, T.E. Hill, Jr., 'Humanity as an End in Itself', in his *Dignity and Practical Reason in Kant's Moral Theory*, pp. 38–57.

harms the casual observer sees in, say, slavery and murder; but in its interference with a central element of human dignity, it offends against morality in a uniquely fundamental way, as if diabolically constructed for just that purpose.

So described, this view of morality would appear to offer an unwavering condemnation of esoteric morality. Nevertheless, I stand by the argument of section IV.4, above, to the effect that even Kant's moral theory violates the publicity condition. That is, in response to the view of morality just described, I claim that one is entitled to deny that its proponents are correct in holding that practical reason, correctly deployed in deliberation and action, will never direct an enlightened and autonomous individual to violate the publicity condition. In support of this thought, one might not only recall the argument of section IV.4 but also note that even some Kantian theorists argue that in extraordinarily 'non-ideal' conditions, the demands of Kantian morality are not as stringent as they are in normal circumstances.[34] So I would claim that although one might expect Kantian moral theory to affirm the publicity condition in the sweeping way suggested above, it is actually the case that a careful application of the categorical imperative and other Kantian principles reveals a more nuanced approach that requires compliance with the publicity condition in some circumstances but requires violation of it in others.

Nevertheless, let me grant, for the sake of argument, that there is some form of a Kantian approach to morality according to which esoteric morality does indeed fundamentally offend against morality—that is, a Kantian approach that does indeed require compliance with the publicity condition as a matter of principle and hence regardless of circumstance. In response to the argument given in section IV about the potentially catastrophic costs of complying with the publicity condition, proponents of this kind of stringent Kantian approach may reply that they are well aware of those costs: those are the costs of being principled. Nevertheless, that argument still has dialectical relevance, because of the intuitive appeal that the publicity condition has, independent of its potential grounding in a stringent Kantian approach to morality. Many people who are unwilling to embrace a stringent Kantian approach to morality because of its high costs may find the general idea of the publicity condition appealing because they are not aware of its high costs. They, not

[34] See, for example, C. Korsgaard, 'The Right to Lie: Kant on Dealing with Evil', in her *Creating the Kingdom of Ends* (Cambridge UP, 1996), pp. 133–58; see esp. pp. 147–53. See also T. Schapiro, 'Compliance, Complicity, and the Nature of Nonideal Conditions', *The Journal of Philosophy*, 100 (2003), pp. 329–55; and 'Kantian Rigorism and Mitigating Circumstances', *Ethics*, 117 (2006), pp. 32–57.

unpersuadeable stringent Kantians, are the intended audience for this article's criticisms of the publicity condition.

V.3. Like Kantians, civic republicans (hereafter, 'republicans') might claim that the argument of section IV fails to respond to the full force of the reasons they have for regarding esoteric morality as offending against something morally fundamental. In particular, republicans—whom I will take to be represented primarily by Philip Pettit—might claim that esoteric morality offends against freedom, properly conceived. On this view, freedom is properly conceived not essentially in terms of the traditional liberal preoccupation with the absence of interference, but essentially in terms of the absence of domination (*Republicanism*, pp. 21–7; *LBL*, pp. 36–57; and 'Simple', p. 342).[35] Domination, in turn, is a relation in which one party has the capacity to exercise arbitrary power over the affairs of another party; this relation can obtain even when the first party does not exercise that power coercively, or does not exercise it at all (*Republicanism*, p. 52; *LBL*, p. 70 and p. 74; and 'Simple', p. 341), as in the case of a kindly master (*Republicanism*, pp. 63–4; and 'Simple', p. 352). Power is arbitrary, finally, when it is not forced to track the interests of the party over whom it is exercised (*Republicanism*, pp. 55–6; and 'Simple', p. 342).

It is natural to think that this view of the importance and meaning of freedom would find esoteric morality deeply offensive. After all, for a coterie to deceive the masses about the nature of morality is for that privileged group to exercise a power that many would regard as a paradigm example of domination. Indeed it is precisely such an arrangement that is referred to by Sen and Williams's evocative phrase 'Government House utilitarianism'.[36] Along similar lines, Sheldon Wolin writes that 'One might argue plausibly that secret doctrines are, by definition, incongruous with … the public world of democratic politics.'[37] One final indication of the alliance between republicanism and the publicity condition is the fact that Pettit makes a point of arguing that his view satisfies

[35] Throughout my discussion of republicanism in this section, short titles are used to refer to three works: '*Republicanism*', for P. Pettit, *Republicanism: A Theory of Freedom and Government* (Oxford UP, 1997); '*LBL*', for Q. Skinner, *Liberty Before Liberalism* (Cambridge UP, 1998); and 'Simple', for P. Pettit, 'Keeping Republican Freedom Simple: On a Difference with Quentin Skinner,' *Political Theory*, 30 (2002), pp. 339–56.

[36] A. Sen and B. Williams, 'Introduction', in Sen and Williams (eds.), *Utilitarianism and Beyond*, pp. 1–21, at p. 16; see also Williams, 'A Critique of Utilitarianism', in J.J.C. Smart and B. Williams, *Utilitarianism: For and Against* (Cambridge UP, 1973), 75–155, pp. 138–40.

[37] S.S. Wolin, *Democracy Incorporated: Managed Democracy and the Specter of Totalitarianism* (Princeton UP, 2010), p. 168.

Rawls's publicity condition (though he does not argue the claim being entertained here, that republicanism entails such a condition) (*Republicanism*, p. 170).

Above, in my discussion of Kantian approaches, I mentioned that one might argue that such approaches do not actually entail a sweeping condemnation of esoteric morality. The same possibility arises in regard to republicanism, because of three distinct elements of it. The first is its criterion for the avoidance of arbitrariness in the exercise of power: responsiveness to the people's interests. Pettit writes, 'I think of a rule of law as nonarbitrary to the extent that those who make the law are forced to track the avowable common interests—and only the avowable common interests—of those who live under the law' ('Simple', pp. 344–5). Thus, power can be restrictive without being arbitrary, as long as it is exercised in accordance with the rule of law. As Pettit writes, 'In the broader republican tradition … the constant refrain is that a nonarbitrary rule of law, while it is certainly restrictive, is not a straightforward offense against freedom' ('Simple', pp. 345–6). Thus, if the common interests of the people are served by some policy or state of affairs, it does not offend against freedom, properly conceived. Obviously the question for us is whether the common interests of the people might ever be served by esoteric morality. If this question is correctly answered in the affirmative, republicanism would appear to countenance certain instances of esoteric morality.

A substantive discussion of this question—whether the common interests of the people might ever be served by esoteric morality—would be too lengthy to pursue here. But we can notice more briefly that this possibility is unlikely to be denied even by those who condemn utilitarianism for violating the publicity condition. This is because they object to utilitarianism on the grounds that it countenances esoteric morality, and they do not typically diagnose this alleged defect of utilitarianism as stemming from utilitarianism's having a defective conception of the people's interests. (Instead, they diagnose it as stemming from utilitarianism's relentless consequentialism, including its taking an instrumental view of morality.) So, the structure of this standard complaint about utilitarianism seems to imply that the common interests of the people might, on occasion, be served by esoteric morality.

It might be thought that republicans can avoid this outcome because their conception of a people's common interests gives a central role to freedom, defined as non-domination. But it must be remembered that domination is defined in terms of arbitrary power, which in turn is defined in terms of a people's common interests, so it would be circular

for republicans to then appeal to non-domination in their account of interests. One might supplement republicanism with an account of interests that would somehow imply the unconditional rejection of esoteric morality, but that would take us beyond standard republicanism (and would presumably underwrite a direct argument against esoteric morality that would not need to be routed through republicanism or any other particular theory). So, republicanism's appeal to the common interests of the people is one element of it that prevents it from implying a sweeping rejection of esoteric morality.

A second such element of it is its consequentialist structure. As Pettit explains, 'It is possible to think of non-domination either as a goal that the state should promote or as a constraint that it should honor.' The first view is consequentialist, allowing that the state might permissibly occasionally offend against the norm of non-domination if doing so is necessary to have non-domination be 'at a maximum.' The second view, in contrast, is a side-constraint view, requiring the state to have 'not … the slightest taint of domination,' even if this prevents it from maximising non-domination (*Republicanism*, p. 99). In the last paragraph of this section I'll return to the side-constraint view, but Pettit regards the consequentialist view as both historically more prominent and analytically more defensible. To support his characterisation of republicanism as historically displaying a 'fundamentally teleological outlook' (*Republicanism*, p. 100), he cites Machiavelli, Locke, and Montesquieu, quoting the last as allowing that 'there are cases where a veil has to be drawn, over liberty, as one hides the statues of gods.'[38] Then, in regard to which view is analytically more defensible, he opts for the consequentialist view, writing that 'it may be quite natural to tolerate a political failure to honour non-domination, if the failure represents the most effective means of increasing non-domination overall' (*Republicanism*, p. 102). The implications for republicanism's putative rejection of esoteric morality are clear: even if (contrary to my argument about common interests) non-domination is understood as never allowing esoteric morality, republicanism's promotion of non-domination would not entail the rejection of esoteric morality in all cases, since republicanism's promotion of non-domination does not entail the rejection of domination in all cases.

The third element of republicanism that prevents it from implying a sweeping rejection of esoteric morality is its potential for subordinating

[38] C. de Secondat Montesquieu, *The Spirit of the Laws*, trans. and ed. by A.M. Cohler, B.C. Miller, and H.S. Stone (Cambridge UP, 1989), p. 204. Quoted in *Republicanism*, p. 100.

the goal of non-domination to other goals, at least in sufficiently extraordinary circumstances. Pettit acknowledges 'the traditional assumption that freedom as non-domination is the only goal with which our political institutions need to be concerned,' but he disclaims any ambition of supporting that assumption himself (*Republicanism*, p. 81). He adds, in a separate discussion, that if, for some reason, the institutional arrangements required for the promotion of non-domination 'proved intuitively repulsive to our moral sense, then we might well wonder whether non-domination was an adequate political ideal' (*Republicanism*, p. 102). So, it is a possibility within republicanism that non-domination might, depending on the circumstances, be subordinated to other goals. Thus, even if (contrary to the previous paragraph) republicanism's promotion of non-domination entailed the rejection of esoteric morality in all cases, republicanism itself would not entail the rejection of esoteric morality in all cases, since republicanism does not prioritise non-domination above all other goals in all cases.

I have argued that there are at least three reasons why one might question whether republicanism really entails a sweeping rejection of esoteric morality. Still, as with Kantian approaches, one might argue that there is some form of republicanism (presumably not Pettit's) according to which esoteric morality is always prohibited. It would have to be a form of republicanism with a suitably publicity-protecting conception of interests, a conception of non-domination as a side-constraint rather than a goal, and an absolute prioritisation of non-domination above all other goals. Would such a stringent form of republicanism be plausible? I do not have space to investigate this question here. I would argue, though, that just as many of the people who find the publicity condition appealing are not drawn to a stringent form of Kantianism, so it is also the case that many of those people are not drawn to such a stringent form of republicanism either. My arguments about the high costs of affirming the publicity condition are relevant to their standpoints, even if those arguments can be shrugged off by adherents of stringent forms of republicanism as well as by adherents of stringent forms of Kantianism.

VI. IN PRINCIPLE AND IN ACTUALITY

One of my aims in the previous section was to argue that although there might be stringent forms of Kantianism and republicanism that unwaveringly repudiate esoteric morality, prominent forms of both theories

(though perhaps the latter more than the former) appear to countenance esoteric morality in some cases. This line of argument reinforces the claim, made in section IV, that the publicity condition is actually violated by a much wider range of moral theories than one might have initially expected. That is, on the assumption that one might have initially expected the forms of Kantianism and republicanism discussed above to unconditionally repudiate esoteric morality, their countenancing of esoteric morality in some cases reinforces my claim about the wide range of theories that actually violate the publicity condition.

In response to this claim, it might be objected that we also need to be mindful of two important aspects of the publicity condition: first, a theory's compliance with it depends on facts about the actual world, not just 'in principle' considerations of the kind stated above; and second, and relatedly, a theory's compliance with it can be a matter of degree rather than a simple binary matter of 'satisfies' or 'violates'. Based on these considerations, we might legitimately judge, for example, that given actual human psychology and other features of the actual world as we know it, some moral theories comply with the publicity condition much more fully than others. And if we can make such judgments, then we should distinguish 'in principle' versions of the publicity condition such as those discussed above from 'in actuality' versions that might also be formulated. Even if the former must be rejected for the reasons given above, the latter may be useful criteria for making comparisons among moral theories.

In response to this objection, I grant that the versions of the publicity condition discussed above might well be regarded as 'in principle' versions, and that it might be possible to formulate useful 'in actuality' versions that would need to be discussed and assessed separately. But I also maintain that the three versions of the publicity condition discussed above are, nonetheless, important versions of the publicity condition, and that it is therefore important to subject them to the kind of logical scrutiny presented by this article. In fact, such an examination may dovetail to some extent with the dialectical interests of proponents of 'in actuality' versions of the publicity condition, insofar as this article may be taken to show that if the general idea of the publicity condition is to have any useful role to play in the evaluation of moral theories, it must be formulated in 'in actuality' versions rather than in any of the 'in principle' versions discussed above. In any event, the possibility of useful 'in actuality' versions of the publicity condition is consistent with the soundness and importance of the foregoing criticisms of the versions of the publicity condition discussed above.

VII. CONCLUSION

This article has offered an examination of the publicity condition as a requirement to impose on moral theories. After distinguishing three versions of the publicity condition and documenting the importance accorded to this requirement in the literature (section II), we saw the inadequacy of several arguments that may seem to justify this requirement (section III). We then saw that this requirement is both question-begging and unreasonably demanding (section IV) and examined Kantian and republican perspectives on esoteric morality (section V). Finally, we reflected on this article's focus on 'in principle'—as opposed to 'in actuality'—versions of the publicity condition (section VI).

None of this is to say, of course, that the publicity condition expresses an ideal wholly lacking in attractiveness. On the contrary, a theorist who regards the publicity condition as an unreasonable requirement to impose on moral theories may find that fact lamentable. This was Sidgwick's view; Parfit writes that 'Sidgwick regretted his conclusions, but he did not think regret a ground for doubt' (*Reasons and Persons*, p. 41). Parfit, in turn, expresses much the same sentiment in the epigraph to this article. Brink, similarly, writes that

> Publicity is a plausible, but revisable, substantive moral commitment. A moral theory that violated publicity in the actual world would be less plausible for that reason. But the fact that there are merely possible circumstances in which a moral theory would require violation of publicity is not a fact peculiar to utilitarianism and is not itself, I think, an objection to utilitarianism or to any other moral theory. (*Moral Realism and the Foundations of Ethics*, pp. 261–2)

Langenfus, finally, puts the point as follows:

> [T]he situation where the vast majority of moral agents would (morally speaking) be precluded from having an explicit or conscious access to the *true* ground of moral obligation would, no doubt, be a disturbing fact. But, however disturbing this might be in terms of 'truth-seeking' ideals, on such assumptions, it would continue to be the only *morally* acceptable situation. ('Implications of a Self-Effacing Consequentialism', p. 488)

I join these theorists in regretting the implausibility of the publicity condition as a requirement to impose on moral theories.

Nevertheless, in this article I have sought to build on the existing literature on the publicity condition by giving arguments supporting an especially decisive rejection of the publicity condition. Specifically, not only do my arguments imply that the publicity condition is an unreasonable

demand to impose on moral theories as an absolute requirement; they also imply that the publicity condition has no significant role to play in comparing the relative merits of alternative theories, since the only theories that satisfy the publicity condition (as a matter of principle, at least—as noted in section VI) are themselves unreasonable ones. If this article can serve as a stimulus for further work that articulates more nuanced formulations of the publicity condition more suited to serve as reasons for theory acceptance and rejection, that would be no occasion for regret.

University of Kansas