Central Lancashire Online Knowledge (CLoK)

Title	How to Deal with Counter-Examples to Common Morality Theory: A Surprising Result
Type	Article
URL	https://clok.uclan.ac.uk/id/eprint/40953/
DOI	https://doi.org/10.1017/S096318012100058X
Date	2022
Citation	Herissone-Kelly, Peter N (2022) How to Deal with Counter-Examples to Common Morality Theory: A Surprising Result. Cambridge Quarterly of Healthcare Ethics, 31 (2). pp. 185-191. ISSN 0963-1801
Creators	Herissone-Kelly, Peter N

How to Deal with Counter-Examples to Common Morality Theory: A Surprising Result

ABSTRACT

Tom Beauchamp and James Childress are confident that their four principles—respect for autonomy, beneficence, non-maleficence, and justice—are globally applicable to the sorts of issues that arise in biomedical ethics, in part because those principles form part of the common morality (a set of general norms to which all morally committed persons subscribe). Inevitably, however, the question arises of how the principlist ought to respond when presented with apparent counter-examples to this thesis. I examine a number of strategies the principlist might adopt in order to retain common morality theory in the face of supposed counter-examples. I conclude that only a strategy that takes a non-realist view of the common morality's principles is viable. Unfortunately, such a view is likely not to appeal to the principlist.

<div align="center">*****</div>

One of the central pillars of the principlist approach to biomedical ethics, alongside a commitment to an irreducible plurality of *prima facie* principles, is a doctrine that can be labelled the global applicability thesis. This thesis states that Tom Beauchamp and James Childress's four principles—respect for autonomy, beneficence, non-maleficence, and justice—are not merely locally but globally applicable. That is to say, they are universally relevant to the sorts of ethical questions that arise in biomedicine. They are every bit as integral to the understanding and resolution of bioethical problems in Bangkok as they are in Boston, equally pertinent in both Phnom Penh and Paris.

Importantly, the global applicability thesis is undergirded by the claim that the four principles form a part of the common morality, understood as a system of general norms to which all morally committed persons everywhere subscribe. Certainly, the norms of the common morality will be specified differently in different cultures. Commitment to them may well show up, or be encoded

in, one set of rules and judgements in culture A, and a quite distinct set of rules and judgements in culture B. But for all that, the thought runs, they will be present in the more specific norms assented to by all those who count as morally committed. Those specified norms are not part of the common morality: only the general, unspecified norms that they express enjoy that privileged status.

In what follows I will first of all examine a number of strategies that the principlist might adopt when faced with supposed counter-examples to the global applicability thesis. Only one of those strategies, I will go on to argue, can be deployed without undermining common morality theory. Establishing this will lead to my central point: adoption of this one promising strategy for dealing with counter-examples will entail the acceptance of a rather surprising and novel view of the principles. That view will be non-realist in character. By this, I mean it will involve the claim that there is no determinate truth about precisely which set of principles underlies the moral judgements and rules to which we adhere.

It is important to note two things about what is being claimed here. First, my point is not simply an epistemological one. It is not the case that a consequence of the one successful strategy for coping with counter-examples, is merely that we cannot *know* which principles inform our more specific norms. Rather, I will maintain that, once we have used the strategy, we will find ourselves forced to hold that there really is no definite fact of the matter—*and so nothing to be known*—about which principles feed into our moral decision-making. This, incidentally, does not mean that everything is up for grabs, or that our moral commitments can be accounted for by reference to just any set of principles. Instead, it means that there will be a number of competing sets of principles, each of which is compatible with and equally capable of systematising our commitments, and none of which can be said to be the one unique set that, as a matter of fact, actually determines our moral deliberation.[1] That, I am aware, probably sounds like a retreat to relativism and, correspondingly, a step away from common morality theory. One of my tasks will be to show that it is not.

Secondly, I should emphasise that I am not personally suggesting that a non-realist principlism constitutes the correct account of morality. I am merely arguing that if principlism and the common morality theory are to fend off the threat presented by the apparent existence of counter-examples, they will have to embrace non-realism. I recognise that non-realism will probably appear uncongenial to most principlists. But for all that, it may be their only hope.

Before expanding upon the non-realist picture of the principles, we need to examine some of the considerations that I think ought to incline us towards accepting it. As I have already suggested, we are pushed in the direction of a non-realist principlism by thinking about how we might respond to alleged counter-examples to the global applicability thesis (and thus to common morality theory). And it seems that there are a number of candidate strategies available to the principlist here, all of which will proceed by insisting that the apparent counter-examples are *merely* apparent.

Determining the common morality's constituents, and justifying the global applicability thesis

Before considering such strategies, however, we should briefly remind ourselves of the ways in which Beauchamp and Childress believe that the global applicability thesis might be justified. The account of potential justifications which features in the most recent editions of *Principle of Biomedical Ethics* (six through eight) is the fullest, and also the most satisfying, they have yet produced.[2] First of all, they hold that a conceptual justification is possible: the norms of the common morality—including, of course, the four principles—will most likely fall out of a rigorous *a priori* analysis of the very concept of morality, since those norms are, so to speak, internal to, or constitutive of, morality. I have previously made use of the following quotation from Philippa Foot, which was subsequently employed by Beauchamp and Childress in order to capture what is meant by certain norms' being internal to morality in this way:

> [T]here are … starting points fixed by the concept of morality. We might call them "definitional criteria" of moral good and evil, so long as it is clear that they belong to the concept of morality—to *the* definition and not to some definition that a man can choose for himself. What we say about such definitional criteria will be objectively true or false.[3]

The thought, then, is that a sufficiently full conceptual analysis of the notion of morality will deliver a set of very general norms which will qualify as the norms of the common morality: the norms to which anyone who is committed to the institution of morality must subscribe. What is more, Beauchamp and Childress are confident that amongst those norms we will find the four principles of respect for autonomy, beneficence, non-maleficence, and justice; the conceptual justification of

the common morality will, it is hoped, thus prove to be a conceptual justification of the global applicability thesis.

However, that is not the only method by which the common morality theory and the global applicability thesis can be justified. And that is probably just as well: conceptual analyses of notions as complex as morality are apt to be not only arduous, but also inescapably contentious—it seems unduly optimistic to suppose that every analyst will come away from the exercise having lighted upon the same set of constitutive general norms. Beauchamp and Childress's alternative method is empirical in character, even while it is, we might say, conceptually informed. The idea is that it is possible to pick out morally committed individuals by means of their adherence to some central norm that is clearly internal to morality in the way already explained. The other general norms shared by such individuals will then be the constituents of the common morality (divergence in more particular, low-level norms will, it is to be hoped, be explicable as adherence to localised specifications of higher-level norms, or principles). Beauchamp and Childress, it seems to me, identify a plausible candidate for the central norm, possession of which will, the thought runs, reveal individuals to be morally committed:

> In a proper methodology, an investigation would include only persons who have already been screened to ensure that they are committed to *some one* norm of morality that is reasonable to expect all morally committed persons to accept. We suggest that a reasonable such principle is *the principle of non-maleficence*, as it is unimaginable that any morally committed person would reject this general principle.[4]

Now, I have in previous work cast doubt on the likely efficacy of this empirical justificatory method,[5] in so far as it seems to rest upon the assumption that, where an agent subscribes to one principle (non-maleficence) that is internal to morality, she will adhere to them all. For the empirical justificatory method to bear fruit, it would seem to have to be the case not only that anyone who subscribes to all the norms internal to morality will adhere to a principle of non-maleficence, but also that anyone who adheres to a principle of non-maleficence will subscribe to all the norms internal to morality. And this assumption is one for which, so far as I can see, Beauchamp and Childress offer no argument. This point, however, is not my central focus here, even though it will inform some of what I later have to say about possible methods of dealing with putative counter-examples to the global applicability thesis.

Beauchamp and Childress's intention is only to lay out the ways in which conceptual and empirical ways of determining the common morality's norms might be carried out. They neither attempt nor claim to have carried out either sort of exercise. Despite this, their confidence that the four principles will form part of the common morality is very high. If we were able to show that they are not part of that morality, or even were unable to show that they are, then all support for the global applicability thesis would collapse. And there have been attempts to show that the four principles are not constituents of a common morality, and thus to produce counter-examples to the global applicability thesis.

A supposed counter-example to the global applicability thesis

One striking supposed counter-example, which I would like to use as a sort of test case in what follows, is provided by R.E. Florida, in his article 'Buddhism and the Four Principles'.[6] Florida insists that a principle of respect for autonomy is not to be found in Buddhist cultures at all, due to Buddhism's metaphysic of 'co-conditioned causality'. The thought seems to be that autonomy is not going to show up as a value, and so as something especially worthy of respect, in a culture that holds to a conceptual scheme in which there is no genuine separation between what we call 'individuals' (any apparent separation being at best an illusion).

Now, we could of course adopt an attitude of scepticism towards Florida's anthropological claim about Buddhist cultures, and insist that a commitment to autonomy *does* show up in the behaviour and moral judgements of morally committed Buddhists. I will examine the implications of such scepticism shortly. In the meantime, let us for the sake of argument take Florida's contention seriously, and suppose that respect for autonomy really is absent from the moral judgements of those who hold a metaphysic stressing the interdependence of all phenomena. The interesting thing about the Buddhist case, and the reason I have chosen to focus on it, is this: it would be very difficult indeed to argue that Buddhist culture is not committed to non-maleficence. The notion looms exceptionally large in Buddhist ethics, in the shape of the norm of *ahimsa*, or non-harm. Those who are committed to Buddhist ethics, then, are, by Beauchamp and Childress's lights, morally committed (at least on the unargued-for assumption—which as we have seen underlies the empirical justification of the common morality—that whoever adheres to one norm that is internal to morality adheres to them all). And yet, if Florida is right, subscription to the norm of non-maleficence is not universally accompanied by subscription to the norm of respect for

autonomy. Why then ought we to have confidence in the principlist's method for discovering the common morality's norms, or by extension in the global applicability thesis?

Possible strategies for dealing with the counter-example

How might the principlist respond to these points? I am not convinced that any of the responses open to her is especially attractive. For example, she might want to say that Buddhist cultures only fail to hold to a principle of respect for autonomy, because of their metaphysic—ditch the metaphysic, and those Buddhists whom we know to be morally committed (that is, those who do more than pay mere lip service to the principle of non-maleficence) will find themselves subscribing to a principle of respect for autonomy.

There are a couple of weaknesses in this response. First, it at least might seem to suggest that the Buddhist's metaphysic is straightforwardly false, in so far as it portrays that metaphysic as obscuring from her some important facts about morality (such as that autonomy ought to be respected). Granted, the metaphysic may indeed be false (I am, needless to say, not concerned here with whether it is or not). But if we are to show that it is, we will have to take a very lengthy detour into metaphysics before returning to ethics. This will be inconvenient at best, and may leave us open to charges of cultural insensitivity at worst. Secondly, the response functions by maintaining the truth of a particular counterfactual conditional, namely, '*If* the morally committed Buddhist were not to believe in a metaphysic of co-conditioned causality, she *would* adhere to a principle of respect for autonomy'. And it can be difficult to show that such counterfactual conditionals are true. It is at least difficult to justify this one independently of empirical data on the values of lapsed Buddhists (and who is to say that, as a result of their lapsing, they might not have come to have their views of morality skewed by adoption of a false metaphysic?). Finally, even if we allow the Buddhist to be morally committed, and the counterfactual conditional in question to be true, is that enough to support the common morality theorist's claim? That claim says that the morally committed all adhere to a certain set of norms; it is not the more diluted claim that they *would* adhere to those norms if the right set of circumstances were to obtain.

Another response might run as follows. It might be urged that, *qua* morally-committed, the Buddhist does subscribe to a principle of respect for autonomy. It is just that her metaphysical beliefs about the interdependence of the world's constituents lead her to suppose (rightly or wrongly) that nobody really possesses the property of autonomy, and that therefore the principle

of respect for that property has, as a matter of fact, no application. In order to explain what I mean here, it might be useful for me to suggest a parallel case. Imagine the situation of a man who is genuinely committed to the principle of beneficence, but who has a firm belief about the scope of that principle. That belief is entailed by certain views he holds about moral status: he thinks (again, rightly or wrongly) that the principle of beneficence applies only to human animals, and therefore that we can have no obligations of beneficence towards non-human animals. Now suppose that all humans are wiped out, except our one man. He will presumably remain committed to a principle of beneficence, but will insist that the world is now such that there is no-one (leaving aside any duties of beneficence he might be thought to have towards himself) to whom the principle applies.

One advantage of this latter response is that it does not involve our insisting upon, and so having to prove, the falsity of any particular world-view. This is because we are not here, as we were in the previous response, concerned to say that a culture fails to be committed to one of the constitutive norms of morality because, so to speak, its field of moral vision is obscured by a faulty metaphysic. Instead, we are saying that the morally committed members of the culture *are* committed to the norm all right; it is just that their metaphysic leads them to believe that there is no-one who falls within its scope. And we can say this while leaving open the question of whether or not their metaphysic is accurate. (As an aside, we of course need to be careful here: we want to be able to say that people cannot hold just *any* old metaphysic and still qualify, provided they subscribe to certain general norms, as morally committed. This is because the holding of some types of world-view—types which, for example, reduce the scope of beneficence to members of one's own class, gender, or ethnicity—can be a moral fault in itself.)

It seems to me, however, that this response is constructed on rather insecure foundations. If commitment to a principle *P* does not make itself known in a person or group's behaviour or judgements, what reason could there be for positing that they adhere to *P*? It might be suggested that we could simply *ask* them whether they hold *P*. But this approach would be rather naïvely optimistic. If Beauchamp and Childress are right, very general principles do not, as it were, exist at the surface of people's moral lives. Rather, they occupy its depths—the surface is instead taken up with specified norms. And it might be as unrealistic to imagine that people can always articulate the principles that undergird the moral rules by which they live, any more than they can explicitly state the rules of grammar that inform the construction of the sentences they utter. (It is worth noting that the post-apocalypse man in my example might reasonably be said still to be committed to beneficence, just because such a commitment *did* show up in his behaviour at a time when he

thought it had application; no parallel situation obtains in the case of the morally committed life-long Buddhist.)

What, then, if we just want to be sceptical about Florida's anthropological claim that a value of respect for autonomy is absent in Buddhist cultures? If my last point is at all cogent, that claim cannot justifiably be based on studies that have simply *asked* those in Buddhist cultures whether they subscribe to the principle: such studies would be inconclusive at best. They must, therefore, be based on observation of the behaviour and judgements of morally committed Buddhists. And I doubt that that behaviour and those judgements would *at all* bespeak, say, a moral indifference to coercion (events in recent decades in both Tibet and Myanmar might seem strongly to support my belief here). But is it not the case that moral rules forbidding coercion are specifications of a principle of respect for autonomy? Buddhist cultures may lack certain specifications of the principle that are common in the West (perhaps no emphasis at all is placed on the notion of consumer choice, for example), but that does not mean that they possess *no* specifications of respect for autonomy.

We need to be wary here, though. Are we sure that what underlies rules against coercion in Buddhist cultures is something accurately describable as a principle of respect for autonomy? It seems at least possible that we are automatically, and perhaps erroneously, interpreting the moral aversion to coercion evident in such cultures as exemplifying a principle to which *we* happen to subscribe. Relatedly, we cannot in employing Beauchamp and Childress's empirical method simply assume that whoever subscribes to a principle of non-maleficence, and is therefore morally committed, *must* thereby be committed to a principle of respect of autonomy that explains their disapproval of coercive acts. This is because the method is intended to *determine* which general norms make up the common morality, and so ought not to proceed by assuming that those norms have already been established, and explaining all observed behaviour and judgements in terms of them.

Of course, it may be that the expression 'morally committed' is not to be understood in a non-technical sense, as applying to anyone who subscribes to *any* norm internal to morality. Instead, we might stipulate that it is a term of art intended to capture only those agents who are committed to *all* the norms internal to morality. But this does not really help us at all, since in advance of carrying out an exhaustive analysis of morality, we have no way of identifying such agents. We

have been given no reason to believe that they are the very same set of people who happen to subscribe to a principle of non-maleficence.

If my concerns about Beauchamp and Childress's proposed method are justified, then, it appears that we are thrown back upon a painstaking and extensive conceptual analysis as the only means of determining which general norms are constitutive of morality. And even if we were to embark upon and successfully conclude such an analysis, sceptical voices might still be raised against the global applicability thesis. That is, suppose that our analysis reveals that respect for autonomy really is amongst the values that are constitutive of morality, and that we take 'morally committed' as a term of art applicable only to those who adhere to all such values. If it is strongly argued that those in Buddhist cultures really do not, for all their opposition to coercion, subscribe to a principle of respect for autonomy, then we will be forced to say that they are not committed to morality. Our sceptical opponent might concede this point, and hold that people in Buddhist cultures are instead committed to some other norm-containing institution, call it morality*. Morality*, she may go on to say, has much in common with morality. However, being a slightly different concept, it differs in some of its constitutive norms, and is not to be thought of as in any way inferior, or indeed superior, to its related concept, 'morality'.

It seems to me that the principlist cannot accept this picture. It is true that, in one sense, it leaves the global applicability thesis intact. That is, the norms revealed through conceptual analysis to be constitutive of morality are globally applicable, in so far as, wherever one finds morality, one finds those norms. But this is not the sort of global applicability that the principlist wants for her norms, since those norms will not all be present in, for example, cultures where morality*, rather than morality, holds sway. In other words, the values constitutive of morality will not be globally applicable in the widest sense, just because morality itself will not be globally applicable (even though all cultures will have some system of norms of the same general *sort* as morality).

A more promising strategy: towards a non-realist model of the principles

There is one way that the embattled principlist can turn at this point in order to preserve the claims of common morality theory, and with them the global applicability thesis. It may, however, be a way in which she is reluctant to turn, insofar as it disrupts the sort of realist model of the principles with which principlism typically (and perhaps necessarily?) operates.

According to that model, the principles are, so to speak, *real existents* that inhabit the depths of our moral lives, and that genuinely inform and explain the lower-level norms to which we subscribe. Many people may not be able to identify or articulate them (recall that it would be inconclusive simply to *ask* people what principles are at work in their moral beliefs), but they are nonetheless wrapped up in and exert their influence on our more familiar, everyday moral rules. We can note first of all that it is this realist account of the principles that has led to many of the problems outlined above. So long as it is retained, the principlist will be dogged and discomfited by apparent counter-examples to the common morality theory, and with it the global applicability thesis. So, for example, the principlist notes that the anti-coercive norm of the Buddhist who protests against the military in Myanmar is consistent with a principle of respect for autonomy, and that she would be apt to regard any such norm that exists in her own community as expressive of, or flowing from, that principle. She concludes that the Buddhist adheres to that principle too, and her commitment to the common morality theory remains undisturbed. However, an objector asks the principlist how she knows that the anti-coercive norm flows from the principle of respect for autonomy, and not from some other principle. At this point, we are led straight into the sorts of difficulties I outlined in the previous section.

The escape route for the principlist lies in the rejection of the realist model. Such rejection would involve her holding that the principles are not real existents, and that they do not inform or feed into, and thus are not 'wrapped up in', our lower-level norms. Instead, she could say, they are simply a set of schemata for grouping and categorising those norms. With this shift in thinking, the standards for a correct identification of the principles that make up the common morality also shift. A principle *P* is now regarded as 'correct' if and only if (1) *P* is consistent with the lower-level norms theorised as falling under it; and (2) *P* affords an efficient and thus pragmatically justified way of systematising those norms. Note here that there is no *fact of the matter* about whether, say, the Buddhist's anti-coercive norm is undergirded by a principle of respect for autonomy or not. But we can be justified in thinking of it as so undergirded—we can *get it right*, according to our non-realist standards of rightness—provided that conditions (1) and (2) are met. We need not worry that, on this modified picture of the relationship between principles and lower-level norms, anything goes. An anti-coercive norm could not be consistent with just any old principle or set of principles. Nonetheless, it could be consistent with a number of *different* principles or sets of principles. The upshot of this is that there will be no single set of 'correct'

principles. And this may seem to force the conclusion that the non-realist route leads straight to a sort of relativism that is entirely inimical to the common morality theory.

This would, I think, be a mistaken conclusion. Recall what the common morality theory maintains. Its central claim is that we can identify a set of principles to which all people committed to morality subscribe. Certainly, once we have taken our bold step into a non-realist conception of the principles, we will have to modify our understanding of what it is to subscribe to a principle *P*. Such subscription will now be comprehended as someone's adhering to lower-level norms that can be systematised in accordance with *P*. But it need not be the case that there is only one way of understanding to which set of principles all morally committed agents subscribe.

So, when Beauchamp and Childress look out upon the world, and declare that the Buddhist in our supposed counter-example subscribes to a principle of respect for autonomy, they are (in our special non-realist sense) quite right. But if a common morality theorist in Myanmar says differently—if she subsumes the Buddhist's anti-coercive norm under a different principle, and indeed fails to recognise respect for autonomy as a principle at all—she is also quite right, so long as the anti-coercive norm is adequately systematised by the principle or set of principles she proposes. The point to understand here is this: both Beauchamp and Childress on the one hand, and the principlist in Myanmar on the other, are right (again, in our special non-realist sense) to hold that there is a set of universally applicable principles. No-one in this imagined situation has abandoned common morality theory. All that has been abandoned is the notion of the principles as real existents that inform our lower-level norms.

Of course, the deeper question is whether, in embracing a non-realist picture of the principles, principlism itself must be renounced. Is that doctrine, in other words, inescapably bound up with a picture of the principles as real existents that inform and explain lower-level norms? It seems only right that I should leave it to the principlists to answer that question.

<center>NOTES</center>

[1] Readers who are familiar with the philosophy of language may detect an echo of Quine's thesis of the indeterminacy of translation here. The parallel is deliberate and acknowledged. Quine WVO, *Word and Object*, Cambridge, MS: MIT Press; 1960: Chapter 2.

[2] Beauchamp TL, Childress JF, *Principles of Biomedical Ethics,* 8th ed. New York, NY: Oxford University Press; 2019: 449-58.

[3] Foot P, *Moral Dilemmas*, New York, NY: Oxford University Press; 2002:6.

[4] See note 2, Beauchamp, Childress 2019:451.

[5] Herissone-Kelly P. Determining the common morality's norms in the sixth edition of *Principles of Biomedical Ethics*. *Journal of Medical Ethics* 2011;37(10):584-7.

[6] Florida RE. Buddhism and the four principles. In: Gillon R, ed. *Principles of Healthcare Ethics*. Chichester: John Wiley and Sons; 1996:105-116.

Conflicts of interest

There are no conflicts of interest.