

WILLIAMS'S CRITIQUE OF INDIRECT UTILITARIANISM

By Jesse L. Rowland (Amherst '16E)

In *Ethics and the Limits of Philosophy*, Bernard Williams examines the “ways in which theory, and in particular utilitarian theory, may be related to the ‘morality of commonsense.’”¹ He is interested in the distinction between ethical theory and ethical practice, and he worries about how some utilitarians try to bridge the gap between utilitarian theory and nonutilitarian ethical intuitions. These utilitarians hold that our ordinary and intuitive ethical dispositions, which often motivate us to act in ways that conflict with utilitarianism, are indirectly justifiable because acting from these dispositions promotes more welfare than if we unsuccessfully attempted to perform the difficult utilitarian calculations that are necessary to determine the best utilitarian course of action. From the perspective of utilitarians, trying to act as a utilitarian is often not the right thing to do. Further, because this only works if we are truly committed to acting from our ordinary ethical dispositions, we cannot be made aware of the fact that these dispositions are only indirectly valuable. There is thus a strange disjunction between the theorist who reflects on the justificatory grounds of our ordinary dispositions and the ethical actor who implements and abides in those dispositions. Williams’s critique of this disjunction can be difficult to understand. In this paper, I try to illuminate his critique by expanding on several key points and by pressing on a number of contentious areas. After reviewing some of the more problematic features of the indirect utilitarian argument, I will turn to Williams’s critique and defend it from the potential utilitarian response that his critique does not apply to all forms of indirect utilitarianism. Along the way, we will also see how Williams’s assessment of indirect utilitarianism sheds light on an important aspect of his general indictment of ethical theory itself.

¹ Williams, Bernard, *Ethics and the Limits of Philosophy* (Cambridge: Harvard University Press, 1985), 106.

I. INDIRECT UTILITARIANISM

Ethical Intuitions

In the chapter “Theory and Prejudice,” Williams considers the possibility of grounding ethical theory in ethical experience. Such an ethical theory would systematize our foundational ethical beliefs, or *intuitions*, or extract an underlying general principle from these intuitions, in order to generate a test of ethical correctitude. Ethical intuitions are “spontaneous convictions, moderately reflective but not yet theorized, about the answer to some ethical question, usually hypothetical and couched in general terms,” which guide us through ethical life and inform our ethical decisions.² These intuitions are, as Williams puts it, only moderately reflective, and they are often the result of natural predispositions or inherited ideas about which kinds of actions, responses, or policies are ethical and which are not.

Utilitarian theory has some rather interesting things to say about our ordinary ethical intuitions. On the face of it, some of our strongest and most basic ethical intuitions seems to conflict with utilitarianism. “The value of justice, truth-telling, spontaneous affection, loyalty to your friends, a special concern for your own children, and so forth, might seem to involve an outlook that the thoroughgoing utilitarian would not endorse.”³ But if ethical theory is going to be based on, or at least account for, our ethical intuitions, it would not look promising for that theory if it told us that most of our deeply-rooted ethical beliefs were wrong. “It is important,” Williams writes, “that, for utilitarian theorists, systematizing everyday attitudes and dispositions does not necessarily mean replacing them. Theory may sometimes justify those attitudes; moreover, it may sometimes do this even if the attitudes are not themselves utilitarian in spirit.”⁴ Utilitarian theorists, such as Henry Sidgwick and R. M. Hare, have argued that individuals

² Williams, 94.

³ Williams, 107.

⁴ Williams, 106.

should act according to their ordinary ethical dispositions because, while they are not justifiable in and of themselves, they are justifiable *indirectly*.

The Argument

The argument for what Williams calls *indirect utilitarianism* is as follows. Ordinarily, individuals are not properly prepared or intellectually equipped to perform the complicated type of utilitarian calculation that is required to determine which actions, responses, or policies promote the most welfare. If they try to consult this type of calculation when they are confronted with an ethical decision in everyday life—perhaps because they have been persuaded by utilitarian argument that the ethically correct thing to do is to maximize welfare, and that in order to do so one must perform a complex multilateral role-reversal thought experiment⁵—they are then liable to make a mistake, which may result in more harm caused than if they simply avoided the utilitarian calculation altogether and instead acted according to their ordinary nonutilitarian dispositions. Since in these cases trying to act as a utilitarian causes more harm than good, the proper thing to do, from the point of view of the utilitarian, is to simply remain a nonutilitarian. When compared to the harm caused by these calculatory mistakes, nonutilitarian dispositions promote the most welfare. Therefore, according to indirect utilitarianism, nonutilitarian dispositions—such as honesty, justice, and loyalty—are justifiable because acting from them results in greater utility than if acting from utilitarian imperatives. In simpler words, from the utilitarian point of view it is usually not good for ordinary people to try and think like a utilitarian.

⁵ Williams discusses the complex type of role-reversal utilitarianism requires in the previous chapter. He cites R. M. Hare and Roderick Firth as key proponents of Ideal Observer theory, which employs a utilitarian test that “postulates one omniscient, impartial, and benevolent observer—he might be called the World Agent—who acquires everybody’s preferences and puts them together. The test of what should be done (or, in indirect versions of the theory, of what practices or institutions should be adopted) then becomes what would be chosen by such an observer” (Williams, 83).

Intrinsic vs. Instrumental Value

Indirect utilitarianism is notable for Williams because it has strange implications for how ethical theory is supposed to relate to ethical practice. Specifically, ordinary ethical actors cannot be made aware of the fact that their nonutilitarian dispositions are only indirectly justifiable; they cannot be shown the justificatory argument above. If they are made to understand that their dispositions are only justifiable because deferring to those dispositions promotes more welfare than if they attempted the far-too-difficult task of trying to figure out which actions maximize welfare, then they would be thinking like a utilitarian, which is exactly what we do not want. They would no longer be able to act according to *intrinsically-valued* dispositions because they would now know that those dispositions are only *instrumentally* valuable. The value of ordinary dispositions, like being honest, or just, or loyal, lies in their being intrinsically-valued, and the better state of affairs is, for the utilitarian, the one in which ordinary people act intuitively rather than over-critically or over-reflectively.

Copycatting

Here we might wonder whether one could not simply *pretend* to value these dispositions intrinsically. If all that the utilitarian is concerned with is the welfare end-state, then it might seem like we could achieve the same result by imitating a person whose dispositions were intrinsically valued. In other words, a utilitarian could explain the indirect justification argument to an ordinary individual (who now comes to value nonutilitarian dispositions only instrumentally), and then further explain that, since what is valuable is an end-state identical to that produced by a person who truly values nonutilitarian dispositions intrinsically, it is important that the individual act exactly as if he truly valued those dispositions intrinsically as well, so that his actions would produce the same end-state. Unfortunately, instructing ordinary

people to become “copycats” in this way will not work. Williams gives a helpful example in an earlier discussion about how having certain virtues affects how one deliberates:

An important point is that the virtue-term itself usually does not occur in the content of the deliberation. Someone who has a particular virtue does actions because they fall under certain descriptions and avoids others because they fall under other descriptions. That person is described in terms of the virtue, and so are his or her actions: thus he or she is a just or courageous person who does just or courageous things. But—and this is the point—it is rarely the case that the description that applies to the agent and to the action is the same as that in terms of which the agent chooses the action.⁶

A truly virtuous person does not deliberate by asking herself what the virtuous things to do is, nor does she assess the value of that virtue in terms of some other intended desire. When trying to determine how to act, a truly compassionate person does not choose certain actions because people view those actions, and the people who engage in those actions, as compassionate. Rather, she acts the way she does because other people are in need of help. The same is true for nonutilitarian dispositions like loyalty. If a person is only pretending to be loyal because it serves some other purpose—it might, for instance, assist him in making new friends, as loyalty is certainly a trait that is desired in a friend—it is not *true* loyalty. Loyalty and being committed to our friends are only valuable if we are truly interested in the wellbeing of our friends, and only if we are not going to abandon them when a situation arises that seemingly calls for us to desert our loyal disposition. If we are only acting loyal, or honest, or just, because we know that the indirect utilitarian argument requires us to act this way, then we are not truly committed to those dispositions and we are liable to ditch them when we think they conflict with the best utilitarian action.

Moreover, even if we were diehard “method actors,” completely committed to never ditching these dispositions, the fact that we only value them instrumentally prevents us from

⁶ Williams, 10.

acting the same as a true believer. With the virtues, it is not possible to act according to a devised rule or principle that will allow one to fake a virtue, because “there is typically no one ethical concept that characterizes the deliberations of a person who has a particular virtue.”⁷ Similarly, with our ethical dispositions, we cannot articulate a rule that will allow us to fully imitate those who truly value them intrinsically. Acting according to ethical dispositions that are only instrumentally valued simply does not have the same effect as acting according to ethical dispositions that are intrinsically valued. “The dispositions help to form the character of an agent who has them, and they will do the job the theory has given them only if the agent does not see his character purely instrumentally, but sees the world from the point of view of that character.”⁸

Learning How to be a Utilitarian

Before we look at Williams’s own critique, there is simpler objection to indirect utilitarianism that we should consider, for the response brings out an important aspect of our ethical life, one that will be relevant later. Indirect utilitarianism gives us the following hierarchy of states (where the lower number promotes more welfare):

- (1) Successfully acting like a utilitarian, if one is capable
- (2) Acting according to nonutilitarian dispositions that are valued intrinsically
- (3) Attempting to act like a utilitarian and failing

The argument, as we have seen, is that since most people cannot perform the utilitarian calculations necessary to properly act like a utilitarian, they end up failing and causing more harm. The conclusion is that people should remain ignorant to utilitarian imperatives and simply act as they normally do, according to their ordinary dispositions. However, since properly acting like a utilitarian is the best state, promoting the most welfare, there would seem to be a strong

⁷ Williams, 10.

⁸ Williams, 108.

obligation to *learn* how to properly perform the complicated utilitarian calculations. In other words, if we are trying to promote the most welfare, and if being a utilitarian, although difficult, is the best way to do that, then we should take the time to learn how to successfully act as a utilitarian.

The problem with this objection is that for many people being an ideal utilitarian is not just difficult, it is impossible. And, importantly, even when a person is *capable* of performing the utilitarian calculations, it may nonetheless be impractical. Ordinary people face ethical decisions every day, and often these decisions require quick responses. Should you leave work early to console a friend whose parent just died? Should you step in when you see a stranger being bullied or harassed? Should you sacrifice your own life to save the person being mauled by a bear? While you may know how to make a utilitarian determination on the matter, you often do not have enough time to deliberate, following each possible choice-path to its end-state and weighing the utility of the respective outcomes. It is thus necessary that ethical actors be able to make quick, intuitive assessments when faced with practical decisions. This further shows the importance of having “strongly internalized” ethical intuitions that are not seen as merely instrumental.⁹ We are now, I think, in a position to better appreciate Williams’s critique.

II. WILLIAMS’S CRITIQUE

The Distinction between Theory and Practice

Williams takes issue with indirect utilitarianism because of how detached the theory is from those whose actions it is supposed to be justifying. He asks, “Is there anywhere in the mind or in society that a theory of this kind can be coherently or acceptably located?”¹⁰ If, as we have seen, it is important that people do not know that their dispositions are only indirectly justifiable,

⁹ Williams, 107.

¹⁰ Williams, 107–108.

then who is this theory *for*? Who is supposed to do the thinking and reach the conclusion that we should continue acting according to our innate and inherited ethical intuitions? As soon as we accept the utilitarian justification, we are no longer able to value these intuitions for their own sake and so no longer able to bring about the state of affairs that is sought in the first place. We are in a strange position, for it seems that the best state of affairs is, from the utilitarian point of view, either the one in which we do not question whether our intuitions are justified, or, if we do question the justification, the one in which we reach the incorrect conclusion that our intuitions are intrinsically good. “There is a deeply uneasy gap or dislocation in this type of theory,” Williams observes, “between the spirit of the theory itself and the spirit it supposedly justifies. There is a distinction that is supposed to bridge the gap or, rather, make us accept it: the distinction between theory and practice.”¹¹ Williams discusses two different forms of indirect utilitarianism, Government House utilitarianism and (for lack of a better name) Cool Hour utilitarianism, which handle this distinction between theory and practice differently, each placing it in different “locations.”

Government House utilitarianism identifies two different social orders: (a) the utilitarian elite, who are enlightened and thus fully capable of performing the utilitarian calculations, and who can devise theories that establish the justificatory grounds of nonutilitarian dispositions, and (b) the ordinary citizens who act from these nonutilitarian dispositions. For the Government House utilitarian, the social elite keep secret the fact that ordinary citizens’ dispositions merely serve utilitarianism, while these citizens continue living in ignorance of the true justification of their actions. Cool Hour utilitarianism, on the other hand, locates the distinction in two different psychological states, or times of reflection, within the same individual. Cool Hour theorists “distinguish between the *time* of theorizing and the *time* of practice, and use Bishop Butler’s

¹¹ Williams, 108.

notion of the ‘cool hour’ in which the philosophically disposed moralist reflects on his own principles and practice.”¹² A Cool Hour individual is supposed to embody both of the Government House social orders, but at different times. He navigates everyday ethical life by way of his ordinary nonutilitarian intuitions about how to act, but sets aside a period of time wherein he can sit back and reflect on the justificatory grounds of these intuitions.

Problems with Cool Hour Utilitarianism

There are serious problems with the psychological detachment that Cool Hour utilitarianism requires. “It is artificial to suppose that a thorough commitment to the values of friendship and so on can merely alternate, on a timetable prescribed by calm or activity, with an alien set of reflections.”¹³ As both practitioner and theorist, the Cool Hour utilitarian must hold both that (1) nonutilitarian ethical disposition x is *intrinsically* valuable, and that (2) nonutilitarian disposition x is only *instrumentally* valuable. But holding these two contradictory views at the same time is incoherent. As we saw earlier, once we have acknowledged the instrumental value of our dispositions, we cannot then continue to act as if we value them intrinsically. Without the social separation of Government House utilitarianism, the Cool Hour utilitarian must keep the justificatory secret from himself. This requires some sort of psychological detachment or “willed forgetting,” in order to secure the genuine commitment to nonutilitarian dispositions.¹⁴ During everyday life, the individual-as-practitioner cannot know what the individual-as-theorist knows. But how is this supposed to work in real life? What is the nature of the transcendental psychological state in which the theorist is completely isolated and removed from his ordinary dispositions?

¹² Williams, 109.

¹³ Williams, 109.

¹⁴ Williams, 109.

Williams believes that Cool Hour utilitarianism must rest on “a certain picture of the time of theory: it is an hour in which the agent leaves himself and sees everything, including his own dispositions, from the point of view of the universe and then, returning, takes up practical life.”¹⁵ The Cool Hour utilitarian must allow for the possibility of momentarily assuming an impartial view *of one’s* dispositions before coming back down to earth and re-inhabiting the ordinary view *from* those dispositions. But it is not possible to dissociate ourselves from the ethical world in which we live, because “any actual process of theorizing of that sort would have to be part of life, itself a particular kind of practice.”¹⁶ This sheds light on an important aspect of Williams’s critique of ethical theory itself. The philosophical project of ethics may be seen as an attempt to justify our ethical dispositions by assuming an impartial, “universalistic standpoint,” and then evaluating ethical life from that perspective.¹⁷ But if we were to actually succeed in completely transcending ethical particularity, we would have assumed a perspective that is completely devoid of ethical value. If, during the cool hour, we were to escape the world as seen from our intrinsically-valued dispositions, we would no longer be in a position to evaluate those dispositions.

Problems with Government House Utilitarianism

The indirect utilitarian might insist that, while Williams is right to condemn the Cool Hour form of utilitarianism, Government House utilitarianism avoids these problems. It does not require that the utilitarian theorist and the ordinary practitioner coexist within the same mind. However, the utilitarian elite must still live a life in which they make everyday ethical decisions. It might be thought that because the elite are capable of performing the utilitarian calculations they can just do so every time they face an ethical decision. But as we saw earlier, even if we are

¹⁵ Williams, 110.

¹⁶ Williams, 110.

¹⁷ Williams, 110.

capable of deliberating like a utilitarian, it is often the case that we do not have the opportunity to do so. While we are weighing each option, trying to determine which course of action will maximize welfare, our best friend is in the meantime having his face degloved by a grizzly bear. We need to be truly committed to our ordinary ethical intuitions in order to make the reflexive types of decisions that are often called for in life.¹⁸ We have seen that we cannot do this when we know that they are only instrumentally valuable, and so the Government House elite social class is still going to have problems when faced with ethical decisions that require quick judgement. Only unrealistically enlightened beings, who already omnisciently know the welfare end-state that would be produced by every possible decision and who are capable of comparing these end-states and making perfect utilitarian decisions instantaneously, would be able to function as the utilitarian elite. While these godlike beings may not have the need for ordinary ethical intuitions, the rest of us, including any reasonably-endowed utilitarian social elite, must often rely on such nonutilitarian intuitions. Government House utilitarianism, then, would require the same psychological dissociation that proved problematic for Cool Hour utilitarianism.

We have seen that Williams's critique applies to both Cool Hour and Government House utilitarianism. It is incoherent to value one's dispositions intrinsically while also understanding that they are only instrumental in bringing about a desirable utilitarian state. If indirect utilitarianism were correct, it would be hard to see how one could be an indirect utilitarian. The theory gives us a test for the correctitude of ethical dispositions, but it is a test that we cannot use because once we understand it we change the result of the test.

¹⁸ The grizzly bear example is of course an exaggerated example that is hardly the type of ethical decision ordinary people face in their daily lives, but it illustrates the immediacy of certain ethical decisions, which is, to varying degrees, certainly a feature of everyday ethical decisions.

WORKS CITED

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press, 1985.