Informed Altruism and Utilitarianism

Abstract

Utilitarianism is a consequentialist theory that assigns value impartially to the well-being of each person. Informed Altruism, introduced in this paper, is an intentionalist theory that relegates both consequentialism and impartiality to subordinate roles. It identifies morally right or commendable actions (including collective actions such as laws and policies) as those motivated by a sufficiently informed intention to benefit and not harm others. An implication of the theory is that multiple agents may perform incompatible actions and yet each be acting rightly in a moral sense.

Key terms: utilitarianism; consequentialism; impartiality, altruism; intention

Introduction

A strength of the classical Utilitarianism established by Bentham is that it is a moral theory for a fallible species in a recalcitrant world. It recognises that having to pursue imperfect, and imperfectly predictable, outcomes is the normal condition of our moral decision-making; and it aims to guide us towards the best outcomes achievable within our limited knowledge and limited powers. It is a naturalistic theory, in the sense that it requires us to accept no extra-human authority or controversial phenomenology, and finds the test of morally right action in the achievement of consequences describable in wholly natural terms, such as the prevention of pain or the satisfaction of desires. Its main variants, such as Act and Rule Utilitarianism, whatever their other differences, share these characteristics.

In this paper, I want to defend a theory that retains some of these advantages, yet differs from Utilitarianism in two fundamental respects: first in taking intention, and not consequence, as the index of right action;[1] and secondly in rejecting impartiality, as between one's self and others, in favour of a sufficiently informed and unconditional altruism.

Much of the paper will be devoted to justifying the first of these, the rejection of consequentialism. The second difference may seem more elusive, since Utilitarianism itself presupposes the possibility of altruism, in the form of a capability of attaching importance to the well-being of others when making decisions. Indeed almost all moral theories presuppose this possibility. But the distinctive demand of Utilitarianism is that, in moral decision-making, we count our own actual or predicted utility as of *equal weight* with that of every other person, subordinating to this impartiality of aim any priority derived from our relations to ourselves or to specific others, unless such priority can itself be justified by some rule of action supposed to maximise (impartial) utility. In the reciprocal relations of an entire community, where every citizen is simultaneously an ego and an 'other', impartiality of treatment and distribution readily emerges as a basis for law and policy within a utilitarian system. But it seems unlikely that many individuals in their moral decision-making can adopt the intellectually difficult and psychologically challenging presupposition of impartiality, where the agent is required to hold in balance her own and the other's interests, as she understands them, and improvise a decision procedure that will express that balance. I will argue that foundational impartiality (as distinct from the second-order impartiality that is mandated in specific institutional contexts by a fixed relation to a set of others, such as a

[1] On this first point, the theory has a close competitor in Foreseeable (or Probable) Consequence Utilitarianism: see section **III** below.

physician's patients or citizens under the law) should be abandoned, in favour of a simpler identification of morality with sufficiently informed altruistic intention. Egoistic intention, though not inherently reprehensible and not irrational except when self-defeating, is not, on this view, part of morality.

These two differences from the classical utilitarian view bring the theory closer, I believe, to everyday moral decision-making. Most people, when trying to act as decent moral agents, do not find the justification of their actions in a foundational impartiality in the service of utility, but in a direct attention to the well-being of others. Nor do most people regard the actual consequences of their actions as the definitive index of whether they have acted rightly in a moral sense. Everyone knows that an action can sometimes be the wrong one in a practical, actually-consequential sense. Rather, they are satisfied if they have tried hard to act 'for the best' (a loosely defined objective, which has something to do with helping and not harming other people); and they know that in an uncertain world their hopes of achieving 'the best' will sometimes, however hard they have tried, be disappointed. Such disappointments do not, in general, cause people to lose that part of their self-esteem that is invested in being decent moral agents, but they may sometimes blame themselves when, in spite of their awareness of good intentions, they come to believe that they could, and should, have tried harder to work out what *was* likely to be for the best.[2]

[2] It is true that there are morally serious people for whom it is not sufficient to act as just described. They need a more definite assurance of acting rightly. They may, for example, aim to discover the morally correct principles of action, in personal relations, or in politics, or in the workplace, and then apply these principles of action in all circumstances. Then, since these principles are the correct ones, the uncertainty of the world makes no difference. They will always have acted rightly whatever happens, and need never blame themselves. This second group of people are not necessarily morally superior to the first. Their assurance of right action may have been purchased with unwarranted confidence in the correctness of their principles.

The fact that a moral theory plausibly accounts for a great deal of everyday moral decision-making does not prove that it is a good theory but, within a naturalistic framework, it counts in its favour. An extra-naturalistic theory of morality, such as one that derived moral conclusions from abstract principles of reason or divine commands, would have less need to worry if it entailed substantial, or even comprehensive, rejection of actual, historically observable, purportedly moral behaviour. Such behaviour might simply show that human beings were chronically wicked or deluded or confused, in need of external correction by God or by philosophers. A naturalistic theory, one that looks for an understanding of morality to the facts of nature, and primarily (as in the case of Utilitarianism) to those of human experience, would have more need to worry, since it would face the additional historical task of explaining how the point of morality, correctly understood, could have come to be so radically misconstrued in the purportedly moral intentions and actions of so many human beings. Utilitarians, recognising this potential challenge, generally begin by emphasising the strong intuitive appeal of their consequentialist conception of morality, its conformity with many everyday moral judgements, while reserving the right to correct everyday moral thinking in so far as it conflicts with utility. But utilitarianism, I will argue, is not in this respect the most persuasive version of a naturalistic moral theory. An alternative view may have even greater intuitive appeal.

I

We can call the alternative theory, for short, *Informed Altruism*. In this first section, I set it out in seven parts or stages, intended to form a roughly sequential exposition. 1-3 are general, meta-ethical claims about the nature of morality and its function in human life, while 4-7

state the substantive content of the theory, and offer a basic apparatus for the evaluation of morally significant decisions. The argumentation in this section will seem rudimentary, but I hope in this initial exposition to show that the claims of Informed Altruism are not *prima facie* less plausible than those of Utilitarianism.

The remainder of the paper develops and defends some aspects of the theory, aiming to show that it can hold its own against Utilitarianism in offering a credible account of important aspects of our moral experience. Section **II** elaborates and defends its 'intentionalism'. **III** argues its superiority to both 'actual consequence' and 'foreseeable consequence' Utilitarianism. **IV** argues that its intentionalism protects it from various objections commonly levelled at consequentialist theories. Finally, section **V** summarises the main differences and resemblances between Utilitarianism and Informed Altruism as theories of individual and collective moral action.

Here, then, is the seven-part exposition of the theory, noting (where this is not too obvious for comment) how it differs from Utilitarianism.

1 Historical Naturalism

Classical Utilitarianism offers a morality derived from the universal sovereignty of pain and pleasure over our perceptions of outcomes as good and bad, desirable and undesirable, and thus (respectively) eligible to be pursued and prevented by rational beings. But these natural facts about pain and pleasure offer equally plausible reasons for non-moral principles of action: for ruthless individual selfishness, or collaboration with others for competitive advantage, to the detriment of others outside the collaborating group. There is plenty of historical evidence that a rational pain/pleasure calculus can motivate people to actions,

individual and collective, that virtually all moral codes would condemn. The facts of pain and pleasure do not, by themselves, point us towards Utilitarianism or any other moral theory. Reflection upon these natural facts needs to be reinforced with some substantive moral objective that requires us to take some actions, and not others, in light of them.

Note that the objection to Utilitarianism here is not that, in basing a moral theory on natural facts, it improperly derives an 'ought' from an 'is.' Following Parfit, we should accept that some natural facts can provide reasons for, or count in favour of, actions. (Parfit 2011, I, 31-38 *et passim*). We can properly say that these actions have been derived, through the mediation of reasons, from the natural facts, though not logically entailed by those facts.[3] If the natural facts of pain and pleasure ceased to obtain, these actions would become actions lacking point or reason. Rather, the objection to Utilitarianism is that one cannot explain why a particular *class* of actions, the class of morally right actions, is derived from, and given point by, the natural facts of pain and pleasure, without referring to a distinctive purpose served by this class of actions. If this class of actions has no distinctive purpose, then there is nothing to block the derivation from the natural facts of pain and pleasure of actions we would never want to call morally right, such as the pleasurable infliction of pain on others.

There is a better way –a way more consistent with the actual history of moral thought and moral systems– of understanding how morality arises from natural facts. According to this alternative view, *morality* is best understood as a natural, historically-developing human

[3] Searle (1969) argues that many institutional and linguistic facts, such as contracts, promises, etc., do logically entail derivation of 'ought' from 'is'. Parfit (2011, II, 310-315) criticises this view on the grounds that it is not contradictory to 'recognise' –in the sense of accepting the existence of– the rules inherent in contracts, promises etc. yet to deny them any normative force whatever (314). I believe (in brief) that Searle is right to insist that institutional facts are as capable as other natural facts of giving us reasons for, or counting in favour of, some action, even if they may be overridden by other reasons; Parfit is right to insist that these are not (in any case) matters of logical entailment, if this is understood to mean that to accept the existence of some fact is to be logically committed to take some action.

faculty. To make this claim clearer, we can compare the historical development of morality to that of technology. Like technology, morality has a distinctive purpose, which gives it its importance in the life-history of the human species: it has arisen because we are purposive animals of a particular kind. Just as technology is possible because human beings can reflectively (and not merely instinctively) deploy non-human objects as tools, in order more effectively to achieve practical ends, so morality is possible because human beings can reflectively (and not merely instinctively) take the good of others as their objective, in order that human lives other than their own, or in addition to their own, go better. If human beings were a species incapable of acting in this reflectively altruistic way, what we think of as morality could not even begin: no credible conception of morality can avoid some appeal to the idea that we are *able* to regard and treat others altruistically. Like technology, morality can make progress (imperfect and sometimes reversible though it may prove): its development is susceptible to cultural enhancement, entrenchment and disruption.[4]

2 Moral codes and the diversity of moral thought

If we are prepared to accept the morality/technology analogy, it is no more surprising that there is diversity in moral thought than that there is diversity in technological invention. The existence of a relatively advanced culture which accepts slavery as part of the natural order of things is no stranger than the existence of a relatively advanced culture, such as that of the Incas, which has not invented the wheel. The moral faculty is a work in progress. Actual historical *moral codes* are developments of the moral faculty which arise under particular

[4] The idea that morality has some determinate general purpose (in the sense of an operative function of which human beings are at least intermittently conscious), such as 'to ameliorate the human predicament, specifically by tending to countervail such ills as are liable to result from the limitedness of "human sympathies"' (Warnock, 1971, 134) is not, of course, a new one. Warnock in common with many mid- twentieth-century philosophers proceeds by 'conceptual analysis', but he also, if less systematically than, e.g. Dewey in his *Ethics* ([1932]; 1985), attempts to situate an understanding of morality within an anthropological/historical framework (e.g.2-11).

historical conditions. Some well-known examples are Christian virtue, the Five Constants of Confucian ethics, the Golden Rule, 'American values', and 'common decency'. Like technological developments, these codes can have limited applicability, or be too strongly influenced by mistaken empirical beliefs or irrelevant assumptions to be effective, but in general they are useful. When we make moral judgements, we often invoke some code (for example, 'your act is an offence against common decency'; 'your proposal is incompatible with American values'), without bothering to reflect critically on the merits of the code itself. Codes typically address commonly-occurring decision situations in particular cultures, and offer guidance which is easy to understand.

Differences in moral code from one culture to another do not undermine the claim that there is a definite substantive content in morality. If the purpose of the moral faculty is to enhance human life by activating our capability of altruistic choice, checking or transcending the struggle of competing egoisms, its codification, even when successful, will sometimes vary across different societies. David Wong, who characterizes morality in a different (but not, I think, radically incompatible) way as the regulation of 'conflicts of interest', reminds us that some –but not all– such conflicts are universal. A moral norm forbidding X to torture Y to satisfy a whim, to borrow Wong's own example, would have universal applicability. (2006, xii).[5] But other conflicts of interest are specific to particular societies, and therefore some moral norms will be distinctive to them: the co-operative norms that emerge among plains nomads, for example, might be quite different from those of a settled agricultural community.

[5] Wong offers "A is to do X under conditions C" as the abstract formula of a moral norm. That might seem to imply that all local norms are derivable from one or more universal principle, adapted to varying conditions. But Wong adds that "Some criteria for adequate moralities will be local to a given society. They neither follow from nor are ruled out by the universally valid criteria." (xiii). That more radically relativist claim might also be not incompatible with the present theory, but the issue is too complex to pursue here.

There is no obvious justification for invalidating moral codes completely, unless -like those of, say, the Nazis- they have plainly parted company with the distinctive purpose of morality, and so leave us with a choice between saying that they are not moral codes at all, or saying that they assert a false morality. They would be like a purported technology that retards the pursuit of practical ends, such as a shipbuilding technique that causes ships to sink, or a medical treatment that worsens the illness. We could say, indifferently, that these were failures of technology, or failures to *be* technology.

3 The moral life, and the possibility of happy egoism

It might be objected that to conceive morality as a natural human faculty like technology misses the point of morality. According to this objection, morality is not just one biologically possible way for human beings to spend their time and energy, but has compelling normative power. Why *ought we to exercise* this faculty? Only when that question is answered can we begin to understand morality.

The answer to this objection is that our most significant natural faculties have purposes, not in the Paleyan sense that they are each externally designed to achieve some end, but in the sense that the life of human beings is, all things considered, significantly enhanced by their exercise. A reflective person can perceive that we have, in general, good reason as a species to develop as effectively as we can technology, morality and other natural faculties (artistic creativity, linguistic communication, scientific inquiry, etc.).

But we should concede that, as Sidgwick (1874, 473) gloomily recognised, morality does not have compelling rational power. If we understand this, we can avoid the impossible quest for a conception of morality which does have compelling power, and yet has some substantive

content, such as permits us to say that the code obeyed by Nazis should not be regarded as a kind of morality. When a mode of reflection has some substantive aim, and is not simply a formal system like those of mathematics, it is always possible to choose to be guided by some other aim. Many human beings try to lead a *moral life*. But a life lived in indifference to, or even deliberate refusal of, the exercise of the moral faculty might not be unhappy, or unsuccessful.[6] Moreover, morality would have no point if we could not also have self-regarding needs and desires, to which others, acting altruistically, could lend aid. It is therefore possible for a rational, though immoral, agent to prioritise only these self-regarding needs and desires.

4 The altruistic aim

If the moral faculty, like the technological faculty, has some distinctive purpose, it must be possible to summarise that purpose in the form of a general action-guiding aim. (For technology, it would be something like: 'use material objects as means to increase effective exploitation of natural resources for human benefit'.) The most plausible statement of purpose for morality is as follows. In so far as the moral faculty prevails in the decisions of rational agents, they evince *informed altruistic intention*: that is, they take as their unconditional aim the good of other human beings,[7] and the avoidance of their harm, and they conform their practical reasoning as well as they can to that end.

[6] I don't mean to suggest that some such choice of 'life' would, or easily could, be made once for all by any individual. The notion of alternatives lives used here is a simplifying device for exposition. In the case of most of us, if not all of us, our lifetime of practical reasoning will display a mixture or patchwork of altruistic and egoistic action.

[7] We might wish to add, 'and other sentient beings', most obviously animals. But it is simpler for present purposes to defer this question.

By ‘good’ here I mean the opposite of harm; or what Parfit calls life ‘going well’. This is not a perfectionist theory, although –as with Utilitarianism– it would be possible to argue that for life to go as well as it can, human beings must approach some conception of perfection. I assume we can set this idea aside.

The point of ‘unconditional’ is not to suggest that morality only exists in the intentional actions of supremely self-sacrificing individuals. No-one acts morally all the time, just as no-one deploys technology all the time. Each of us is aware of herself as a potential object of another’s altruism or egoism: sometimes, we ourselves act egoistically, and sometimes we act to protect ourselves from others’ egoistic actions. (See the discussion of ‘servility’ later in this section.) Rather, the point is to represent the principle, when it *is* being acted upon, in its purest form.

Various arguments might be advanced for the imposition of what one might call ‘target-conditions’ on altruistic action. Some, for example, would argue for limiting the altruistic aim to the good of those who *deserve* to experience good. Much then depends on how these who favour limiting the aim in this way would understand *desert*. If their arguments failed, or did not even attempt, to show that applying a desert condition would subserve the altruistic aim itself (all things considered), then Informed Altruism would reject its application. We are familiar with this controversy in the stand-off between utilitarian and retributive theories of punishment, and it recurs in Parfit’s hope to find a proof that ‘no one can ever deserve to suffer’ (2011, II, 569). We must leave this unresolved question for another occasion.

Altruism, then, is unconditional, in the sense that no other person is in principle excluded from being its object, or assigned a diminished value within it, and its extension is limited

only by our capacity to exercise it as effectively as we can. If Informed Altruism might be called 'impartial', it is only in the sense that it recognizes this unconditional eligibility of others. It does not, unlike Utilitarianism, require impartiality in the stronger sense, of an at least notionally possible calculation in which the utility of each person, including myself, must be counted as = 1 in order to meet the moral standard. Utilitarianism can, in some versions, countenance departures from impartiality in our individual actions (for example, a rule-utilitarian ideal code might include a rule requiring parents to prioritise the welfare of their own children; a sophisticated act-utilitarian theory might recommend this rule as a decision procedure), but only on the condition that there is reason to believe that such departures will tend to serve an impartial utility better than case-by-case attempts at impartial utility judgements. Informed Altruism requires no such condition.

5 Terms of moral appraisal (I)

Morally right, or commendable,[8] actions are those motivated by a sufficiently informed, reasonable and reflective intention to act in accordance with the altruistic aim. That is to say, an intentional action is morally right or commendable if and only if (a) the intention is to benefit or prevent harm to others, and (b) the decision to perform this particular altruistic intentional action is informed by such acquisition and deployment of knowledge and understanding as can reasonably be expected in the circumstances.

6 Terms of moral appraisal (II)

Morally acceptable actions are those which, whether or not actually motivated by informed altruistic intention, are in their reasonably predictable consequences consistent with it (that is: they are actions that a person with that intention might have performed), and are not

[8] I explain the distinction here, and its relation to the next category of the morally acceptable, at 18-20 below.

motivated by its repudiation. Morally right and commendable actions are *ipso facto* morally acceptable, but morally acceptable actions also include morally neutral everyday actions such as walking along a street (but not: walking along a street intentionally barging into people).

7 Moral conflicts and the non-fulfilment of an altruistic aim

Moral conflicts arise when the altruistic principle cannot be ideally fulfilled, because possible altruistic actions are incompatible (for example, an action which benefits A prevents benefit to B, or harms B). Moral conflicts are a normal feature of morality. Usually, though not always, there is a preferable course of action, which various secondary principles can help us to determine. The unfulfilled altruistic objective, nevertheless, does not lose its unconditional character; we should always regret its non-fulfilment.

Like Utilitarianism, Informed Altruism would not, therefore, fail as a moral theory because of cases in which we harmed A as a necessary condition of helping B. The altruistic aim to help and not harm A would not be like a *prima facie* judgement, which further investigation might cancel. It would remain morally in force even if its fulfilment was unavoidably blocked. Such failures of ideal application, when unavoidable because of the way the world is, and because of human limitations, leave a morally necessary residue of regret. Not to regret would be to deny the theory, and to begin to habituate oneself to others' harm.

A few points of exposition and clarification can be added right away. These may help to head off some possible misunderstandings, and begin to situate the theory more clearly in relation to some current preoccupations of moral philosophy.

(1) A key feature of Informed Altruism is its *epistemic requirement*, summarised above in the condition of 'sufficiently informed, reasonable and reflective intention'. Although it is not a directly consequentialist, but an intention-focused theory (and the implications of this distinction will become clearer later), the intentions that it commends are those informed by sufficiently careful thought about likely consequences of actions. The intending altruist must pursue and deploy relevant knowledge and understanding, to whatever extent is reasonably required in light of her own capabilities, the complexity and abstruseness of the decision situation, and the seriousness of the outcomes at stake. There is no precise formula for weighting these variables: and moral appraisal is, at least sometimes, a matter of degree.[9] The criterion of sufficiently careful thought should not be understood as requiring separate extended deliberation prior to any action, however trivial; like Utilitarianism, the theory can accommodate useful dispositions to habitual actions,[10] provided that the altruistically-intending agent could, if pressed, justify such dispositions as sufficiently informed by reasonable knowledge of likely consequences.

(2) Among the more important things the altruistically-intending agent needs to keep in mind are her own and others' intellectual fallibility, and the difficult of predicting longer-term events. These considerations tend to favour giving greater weight -though again, there can be no precise formula- to what is salient and immediate in a decision situation, compared to what is remote and speculative. We can call this the *salience guideline*. As an example, the moderate priority that common-sense morality expects parents to give to their children in reflective moral decision-making is most plausibly justified by epistemic considerations, and

[9] Compare the 'scalar' consequentialism of Norcross 1997. But the scale is applied in this case to the sufficiency of the trying (cf. the discussion of Mason below) and not, as in Norcross, the approach to optimality of the outcome.

[10] Compare, e.g., the motive utilitarianism discussed by R. M. Adams (1976, 467-470).

especially by the salience guideline. We know these human beings relatively well, we have a relatively extensive power to affect them, we may be able to judge their interests better than we can judge those of other people, and above all they are in our moral sights, so to speak, more immediately and continuously than others.[11]

(3) The epistemic requirement and the recognition of personal fallibility should make us alert to the fact that individuals intending to benefit others often make serious mistakes, and do more harm than good. Bentham is right when he emphasises that, in general, people are better judges of their own interests than are others; that increased happiness is often better achieved by individuals being left to pursue their own satisfactions, than by their attempting to confer benefits on one another. (1834, II, 288). Human beings are also, in general, better placed to report accurately harms and benefits experienced by themselves than by others. But Informed Altruism can accept these generalisations about human experience. Altruism is not intrusive control, but having an eye to other people's happiness and acting, or refraining, accordingly. The epistemic requirement puts the agent on guard against thinking, without good reason, that she knows what will be best for others. It authorises not only acting upon the other person, but leaving her alone, or actively empowering her freedom, according to the best judgement and insight we can achieve.

(4) From the unconditional character of the altruistic aim follows the moral wrongness of intentionally pursuing an outcome for others less good than one we know we could have

[11] That these reasons, rather than the purely biological relation, are decisive is suggested by an example in Sidgwick. He imagines that he and his children, following a shipwreck, are thrown on a desert island, where they find an abandoned orphan. (Sidgwick 1874, 325.) Assuming that the orphan's vulnerability is no less than that of his children, it seems intuitively clear that he ought to extend care to this child which is, at least broadly, comparable to the care he extends to his biological offspring. Further persuasive arguments for what I am here calling a salience guideline are developed in Jackson 1991.

achieved.[12] But we know there are outcomes good for others that we could not have achieved, because of our moral and intellectual limitations: on these failures, we and others can pass graduated judgements that are not equivalent to an intolerable (and demoralising) demandingness. Graduated judgements do not legitimise self-exculpation. In so far as we choose to live the moral life, we aim to do the best we can: it would contradict that aim to assign for ourselves a 'satisficing' threshold of actual achieved benefit to others with which we should rest content (even assuming –which is doubtful– we could find a formula to express such a threshold). Of course, it would always be morally better if we all tried even harder to be effective informed altruists. And 'trying to do well by Morality', as Elinor Mason argues (2019, 50-74), is more than a kind of muscular exertion of the will: there are psychologically complex pitfalls –self-deception, ambivalent intention, instrumentalization of displayed altruism for half-conscious purposes of self-advancement– to be negotiated. The epistemic requirement extends to our own capacity for self-deception.

(5) Finally, Informed Altruism does not imply self-exclusion from the category of objects of altruism. The altruistic aim of morality is not a mysterious phenomenon, and a morally competent person is assumed to be aware, in broad terms, of its expressions in her own society, including the moral codes that attempt (but sometimes fail) to implement it. It does not therefore endorse the self-harm Hill (1991) criticises as 'servility', and illustrates by such cases as that of the 'Uncle Tom' who "displays an attitude that denies his moral equality with whites" (9), or the 'Deferential Wife' who "believes that the proper role for a woman is to serve her family" (6). In these cases, misguided by psychological deficits or by defects in the current moral codes, the servile person wrongly supposes that she alone (or persons like her)

[12] Of the kind that worries Ben Bradley, in his critique of Satisficing Consequentialism (Bradley 2006).

cannot be an appropriate object of the same altruistic aim as others, and conforms her actions to this false belief. She fails to grasp the comprehensive scope of the moral faculty and its freedom from target-conditions. As Hill remarks, such servility is morally objectionable because it expresses "a certain [mistaken] attitude towards one's rightful place in a moral community" (9). In a community whose dominant moral code expresses a conception of the equal worth of all, which in turn generates a discourse of universal rights, servility can be appropriately characterised in Hill's terms as "failure to understand and acknowledge one's moral rights" (9), where one's moral rights are precisely the implications of one's membership of a community of equals.[13] It is also morally wrong for anyone else to endorse the servile person's misconception, or for a community to require servility of any of its members.

However, as Hill recognises, a person who fully understands her moral rights may refrain from *asserting* them, in order to prioritise the welfare of another, and such actions may be morally commendable.[14] Where A's assertion of a right would conflict with B's interests, and no impact on third parties is involved, there is a moral case for not asserting the right. An example would be the forgiving of an offence by another that one might legitimately resent (so far as one's moral rights are concerned).[15]

[13] A solipsistic individualism is not dependent, however, on a mistaken self-exclusion from a set of universal rights. Where the dominant moral code is different, as in a caste system, or other hierarchical culture, a conception of equal rights would be absent, but there could still be servility in the solipsistic sense.

[14] "In order to avoid servility, a person who gives up his rights must do so with a full appreciation for what they are. A woman, for example, may devote herself to her husband if she is uncoerced, knows what she is doing, and does not pretend that she has no decent alternative" (Hill 1991, 15).

[15] Hill would, I think, add a further, Kantian claim: that we have an obligation to ourselves, because we have an obligation to respect the moral law, and we ourselves are among the persons who are protected and respected by, as well as bound by, the moral law. (Hill 1991, 12-16, 19-24, *et passim.*) This claim again forces us to consider the distinction between recognising myself as among the persons protected and respected by the moral law, and actually asserting one of my rights under the moral law. I can surely, provided this does not involve conspiring to violate the rights of third parties, exercise autonomy –without violating some supposed obligation to myself– in choosing freely to waive my recognised right, in order to benefit others. This is not the same as the "pretending to approve of violations" of my moral rights that Hill attributes to the servile person (14). Rather, it

II

The most important feature of Informed Altruism which distinguishes it from Utilitarianism, and from other consequentialist theories, is its intentionalism: it identifies morally right or commendable actions as those grounded in a reasonable and informed intention to benefit and not harm others. On this view, an intentional action is morally right or commendable if and only if (a) the intention is to benefit or avoid harm to others, and (b) the decision to perform this altruistic intentional action is informed by such acquisition and deployment of knowledge and understanding as can reasonably be expected in the circumstances.[16] An aspect of this epistemic requirement is *epistemic modesty*, an awareness of one's own moral and intellectual fallibility in understanding others and the world.

On this view, the intentional action that emerges from the process of decision may sometimes actually lead to bad consequences, or worse consequences than a possible alternative action, but if conditions (a) and (b) are both met, it remains a morally right or morally commendable action.

Without (a), an action cannot be either right or commendable. At best it can be morally acceptable, if its objectives are such as altruistic intention *could* aim at, and it is not enacting

is a further intentional act in pursuit of the very aim that is codified in any fair and established system of rights. My obligation to myself and to the moral law cannot be an obligation not to do this. "Giving my fair share and no more" is a morally acceptable but not a morally commendable policy.

[16] Or in a situation of urgent decision, at least by such attentiveness as is realistically possible to the situation as it presents itself.

a counter-altruistic intention. Offering to share one's taxi with a fellow traveller with the intention of halving one's costs is morally acceptable. Doing so with the intention of making a third traveller feel excluded is not. Doing so for both reasons is a morally mixed action. Morally mixed actions are very common. For example: one buys a lottery ticket, partly in order to make a small contribution to charity (morally right), partly with the aim of winning a fortune (morally neutral). If one wants the fortune partly in order to found a hospital, partly in order to stay in five-star hotels, and partly in order to make a rival feel envious, this opens up the morally neutral aim into a morally mixed plan of actions.

Where there are alternative actions available, each of which might benefit another or others according to a sufficiently informed judgement, we can distinguish between 'morally right' and 'morally commendable'. If the agent chooses the option for which there is the most compelling altruistic case, her action will be morally right. If she makes a different choice, yet there is an arguable altruistic case for this actually chosen option, her action may be morally commendable. If she chooses an option which is only weakly altruistic in preference to one for which there a much stronger case, her action will not only not be morally right, but may not even be morally commendable. An example from ordinary life would be this: Anna went to visit her friend, who (she knew) would benefit from her help with a school project, instead of visiting her grandmother, who (she knew, or should have known) was lonely and depressed that evening. Anna acted in a morally commendable way, in light of her altruistic attention to her friend, though arguably hers was not the morally right choice. If she knew, or could reasonably be expected to have discovered, that her grandmother was seriously ill, her actions were not even morally commendable. If she knew, or could reasonably be expected to have discovered, that her grandmother was in immediate need of help in summoning medical assistance, her actions were not even morally acceptable.

It's important to emphasise that condition (b) refers to the knowledge and understanding possible for the actual historically situated individual. Two individuals may therefore take actions which are both morally right, but are in practical conflict. Suppose, for example, that in signing the Munich agreement with Hitler, Neville Chamberlain (a) intended to benefit and prevent harm to others as effectively as he could and (b) took every reasonable step within his powers to gather and deploy relevant information to that end – bearing in mind that this epistemic requirement will have been ratcheted up to the highest level given the scale and gravity of the human risks at stake. On these two, admittedly demanding, suppositions Chamberlain's action was morally right.[17] An anti-appeaser such as Churchill who advocated a contrary policy might also have satisfied conditions (a) and (b), and in that case, his action in campaigning against the appeasement strategy was also a morally right action. Each acted morally rightly in his given historical position.

This analysis leaves us with two morally right actions by individuals amid the political turbulence of 1938, tending towards different and partly incompatible sets of potential consequences. But that is what one should expect, since different individuals have different decisions within their power, different capabilities of understanding, and different access to deployable relevant information. 'Morally right action' is not, on this view, the same as 'consequentially best action as judged from a wholly objective standpoint'. In order to make robust moral judgements, we do not have to track down the consequentially best course of action (supposing that we could discover what this would have been), and dismiss others as morally wrong. But we can, on this view, aim moral criticism at those who lacked altruistic

[17] Of course, many historians would dispute the truth of (b) in Chamberlain's case: and because of the importance of the epistemic requirement in such a critical case, this would be a *moral* criticism, not merely the awarding of low marks for insight.

intention, or were negligent of the need to pursue and deploy relevant knowledge and understanding as well as they could in such compelling circumstances.

If, improbably, we were able to transcend the decision situations of individual agents, and identify from a wholly objective standpoint the 'consequentially best' course of action, the conclusion that we could then state could not be the basis of a moral judgement. It would merely be a remarkable feat of historical analysis. We cannot reasonably require Chamberlain himself, as a moral requirement, to have done *other than the best he could do* to pursue and deploy relevant knowledge and understanding.

We may perhaps plausibly suppose that another person, differently placed, epistemically speaking, could have rightly taken a different course of action, and that it would have been better if Chamberlain had deferred to the judgement of such another person. Supposing that Chamberlain could reasonably have been expected to reflect on whether to take the decision himself or to defer to another's judgement, the question whether he was morally right to take the decision himself would then turn on whether he took every reasonable step, given the information available to him, to decide whether his or another's judgement should be relied upon.

The weighting of the various contextual variables within condition (b) –the powers and capabilities of the agent, the complexity of the circumstances with their attendant probabilities, and the seriousness of the issues at stake– cannot be specified systematically. But for us even to consider the attribution to someone of a morally right or commendable action, it must be assumed that their capabilities are such that they can be seriously attempting to practise the moral faculty, and can recognise at least the most salient facts

relevant to an intended altruistic action. We are not obliged to find an action morally right if it is the best achievable by a person not so capable, such as a psychopath or a brainwashed Nazi.[18] And when an agent holds public office, there is a particularly demanding standard of capability in respect of condition (b), including such sophisticated expectations as the collation and synthesis of information and analysis from specialist advisors. It may be morally right, if in office, to take the best decision one can take given one's powers and capabilities, but morally wrong to have remained in office taking the decisions, if one knew, or should have known, that one's capabilities were likely to lead to errors into which someone else would be less likely to fall.

III

Examples such as the one just discussed have sometimes been framed within a conflict between Actual Consequence Utilitarianism (ACU) and Foreseeable Consequence Utilitarianism (FCU).[19] Both views have undergone a variety of formulations, but we can see the nub of the issue if we take ACU as holding that

[18] Such cases would normally be precluded by condition (a). But one can imagine ingenious counter-examples in which it is supposed, e.g., that an altruistically-intending youth has no access to any source of knowledge and understanding other than Nazi propaganda, which prompts him to kill Jews as a 'necessary evil' in order to save the lives of Aryans. (Assume that he acts voluntarily and not under pressure of threats.) We might conceivably wholly exculpate this person because of his brainwashing, but could not possibly want to say that his action was morally right. How can we explain and justify this intuition? The guideline of giving greatest weight to the salient and immediate facts (see 15 above) is crucial here. Our default assumption in a culture in which altruism is reasonably well entrenched is that an ordinary good person will always reject saliently (and grossly) counter-altruistic actions, such as killing a person at your mercy who presents no immediate threat to anyone. The action of the young Nazi could only be judged morally right if we were willing to overturn this default assumption, and think it reasonable for him to be guided by the force of a sincerely-held political argument –which *ex hypothesi* this naïve agent is incapable of evaluating critically. Either the young Nazi must be judged morally incompetent, i.e. as lacking the basic capacity for moral decision; or, if he is taken to be morally competent, his action must be condemned as a failure to satisfy condition (b).

[19] A key statement of ACU is Moore 1907, 190-195. The conflict was revived, and fought at least to a temporary standstill in the 1970s and 1980s. See Singer 1977; Temkin 1978; Gruzalski 1981; Ellis, 1981; Strasser, 1989.

> The performance of an act [where 'act' includes rule, policy, system, etc.] is made to be morally right by the fact that the actual consequences of that act's occurring or existing are better from the point of view of utility than the consequences of alternative acts which might have been chosen. Only these actual consequences can confirm or disconfirm the moral rightness of the act.

FCU, on the other hand, holds that

> The performance of an act [where 'act' includes rule, policy, system, etc.] is made to be morally right by the fact that the consequences of that act's occurring or existing, as reasonably foreseen at the time of action, are better from the point of view of utility (when compared using a calculation of probability x expected utility) than the reasonably foreseen consequences of alternative acts which might have been chosen. The moral rightness of an act can therefore, in principle, be evaluated at the moment of choice.

This is not, it is important to emphasise, a debate about the merits of competing decision procedures, or about how much care you need to take over your decision procedure in order to be morally in the clear, though these are important questions.[20] Rather, the question is whether actually achieving the best result is the criterion of moral rightness.

If ACU is right, we face the unsettling conclusion that any reliable evaluation of the moral rightness of an act can only be made retrospectively, when its actual consequences have

[20] See, e.g., Feldman 2006.

become fully apparent. At most, we can make initial, conditional evaluations, which may need to be revised, and further revised, as further consequences emerge. For the agent herself, this means that, in making the decision, she not only cannot know whether her chosen act will lead to her intended outcome: she also cannot know –however good her intentions and however industrious her practical reasoning– whether she is acting rightly or wrongly in a moral sense, and must wait on an indefinite series of subsequent events to find out. A further objection to ACU, noted by Singer (1977, 72) is that, notwithstanding its robustly empirical appeal to the evidence of actual consequences, it requires an epistemically asymmetric comparison between these consequences and alternative 'consequences' which have not occurred. It is, then, arguably just as speculative as FCU, in which we compare alternative sets of possible consequences, none of which has (yet) occurred.

In defence of ACU, Jack Temkin convincingly argues that these are merely epistemic inconveniences which cannot invalidate the ontological claim that actual consequences provide the objective determinant of moral rightness (1978, 413-414). It is unclear, indeed, that anything could invalidate this claim: if you affirm it, and are prepared to accept its implications, you have a logically impregnable theory. Still, these implications are not merely epistemic. All else being equal, it seems to be a weakness of a moral theory that it requires an indefinite deferral of moral evaluation, not only by observers but by the agent herself. We like to think that when morally mature adults perform intentional actions, they can normally know –allowing for the possibility of culpable self-deception– whether they are, or are not, intentionally acting in a morally right, commendable or acceptable way. If they normally did not know, there would be little point in encouraging people to act rightly, commendably or acceptably, or in imposing on ourselves a duty to do so.

The defender of ACU might concede that, though an act with actual bad consequences is always a morally wrong act, the agent in some cases may not be *blameworthy* for it.[21] That would account for the cases in which a well-meaning agent, intending and expecting a good outcome from what, had her expectations been fulfilled, would have been a morally right act has the unhappy experience of discovering from the actual consequences that her act was morally wrong after all. The separation of blameworthiness from moral wrong, with blameworthiness assigned the role of meeting the intuition that moral duty bears some kind of contemporaneous and not deferred relation to action, preserves the appearances and is not an impossible theoretical model. But this separation is itself uncomfortably counter-intuitive, an adjustment to our ordinary moral discourse to be made only if there is no better alternative. 'You did what was morally wrong, but we do not blame you' is the kind of statement we generally reserve for cases of crime or other serious moral offence in which there is a powerful excusing condition, such as involuntary intoxication, insanity or extreme duress. It sounds quite different if said to some morally capable and well-intentioned person who has taken every reasonable step to act for the best.

FCU avoids these problems by requiring us instead to compare, at the time of choice, two or more sets of consequences which might happen, estimating their probability as well as the utility they will provide if they do occur. Epistemically, this may be no easier than, later on, comparing actual with might-have-been consequences. But our knowledge that we are uncertain of outcomes, and can only do our best, is incorporated by FCU into our moral discourse. To act rightly, on this view, is to act on our best reasonable prediction of consequences and utility.

[21] See Moore 192-193, Ellis 329-331, and Gruzalski 174.

FCU is open, however, to this objection by its ACU-defending opponents: that it is not really a consequentialist, or a utilitarian, theory at all, because it separates the morally significant act from the only kind of evidence that under a consequentialist theory can confirm or disconfirm its rightness. Any theory in practical reason needs to state the conditions of satisfaction of an intentional action. On a purportedly consequentialist theory of morality, the condition of satisfaction of the intentional performance of a morally significant action must take the form of some consequences. Consequences that have not occurred, and may never occur, utility that has not been enjoyed, and may never be enjoyed, cannot be what makes an action right *in a consequentialist theory*: yet it is these speculative items that are weighed in judgement by FCU. They are not conditions of satisfaction: they are hopeful intentions, which may be more or less well-informed. What FCU is really saying, according to this objection, is that it is not consequence that matters morally when an act is performed: what matters morally is the intention with which it is performed, and the more or less well-informed hopes and fears that shape the decision to perform it.

It is the assumption that the moral judgements of FCU can be corrected by the actual utility of outcomes (an understandable assumption, given the appearance of the words 'consequence' and 'Utilitarianism' in its full title) that exposes FCU to these criticisms. If FCU, in the light of this problem, renounces consequentialism, it dissolves into an intentionalist theory, such as Informed Altruism, which would assert the following:

> The performance of an act is made to be morally right by the fact that, at the moment of choice, the agent intends to achieve the good of other human beings, and the avoidance of their harm, and conforms her practical reasoning to that end as effectively as is reasonable to expect in the circumstances. Where a moral conflict

> presents itself, the act is morally right if the agent has made every reasonable effort (i.e. such as is reasonable to expect in the circumstances, and taking account of the salience guideline) to prefer the act which she judges will serve best the objective of the good of others and the avoidance of their harm. The moral rightness of an act can therefore, in principle, be evaluated at the moment of choice.

A number of further qualms commonly voiced about FCU can be dispelled at least as well by Informed Altruism as by FCU itself. One is that FCU leads to the disturbingly permissive-looking conclusion that actions may be 'subjectively right' yet 'objectively wrong'.[22] Another is that it creates an incoherence in our moral discourse, when the moral rightness of an act motivated by an expectation meets the intuitively obvious moral wrongness of that very act's actual consequences (for example, when innocents are killed –a consequence that it could not be morally acceptable to have intentionally caused). The answer to both of these concerns is that the actual consequences simply do not affect the moral rightness of the act. If an act intended to murder someone accidentally saves the life of the intended victim, the victim is *lucky*, not the agent more morally right than she thought she was; if an act intended to save a life accidentally destroys one, the victim is *unlucky*, not the agent less morally right than she thought she was. The epistemic requirement –that the agent acquire and deploy knowledge and understanding as well can reasonably be expected in the circumstances– defuses the suspicion of a merely 'subjective' rightness, where 'subjective' implies 'justified by criteria only available to the agent'. Whether the agent fulfils the epistemic requirement, in performing an action directed towards an altruistic end, is, at least in principle, legible to others: assessment of the act by oneself is not a solipsistic deed immune to external

[22] See Lewis 1969, 35-38; cf. Singer 73-74.

correction. Saying that a right action is 'subjectively' right is just a way of reminding us that the decision to perform it has been made from a specific agent's historically situated point of view: and that is a characteristic of any action. Mason comes close to this position when she claims that "an agent acts subjectively rightly when she tries to do well by the standards of what *really is* morally appropriate. In other words, she has to get morality right" (9). [Italics in the original.] Lest this imply that conscious intellectual grasp of some true moral system is a necessary condition for moral rightness in action,[23] I would prefer to say "she has to have sufficiently informed altruistic intention"; but in either case, the notion that an action can be right merely relative to an agent's personal value system or set of priorities is avoided.

Informed Altruism has several other advantages derived from its rejection of the view that actual consequences affect the moral rightness of an act. It dispenses with the need for 'conditional' and revisable judgements of right and wrong, and for consequential 'moral luck'. It respects several widely-accepted ethical axioms, of which Brian Ellis reminds us in his critique of both 'prospective' and 'retrospective' utilitarian theories: 'ought' requires 'can'; every action is either right or wrong;[24] X did wrong if and only if X ought to have acted otherwise; a right action is praiseworthy.

Informed Altruism also avoids another objection Ellis levels against both utilitarian theories: that if they are construed as claiming that, in order to act rightly in a moral sense, individuals

[23] This implication would debar from moral rightness simple acts of thoughtful kindness, performed by people who do not think about morality from one year's end to the next. On my account, they are exercising a human faculty, even if they no more reflect upon it as they do so than a person using a spade to dig a garden thinks about technology. Mason cannot, I believe, mean to deny moral rightness to many such acts. Though she compares the necessity of grasping morality in order to act rightly to that of grasping the game of chess in order to play it (38), which suggests a degree of system-awareness, many of her remarks emphasise that the possession of a moral (or any other) aim can be "in the background", tacit and inexplicit, offering the test that "the agent, if asked, would agree that it is one of her aims" (51).

[24] Assuming we allow for graduated, as distinct from all-or-nothing, judgements. The key point is that the judgement is in principle determinate.

should act so as to bring about whichever 'total realizable situation' has maximum (actual or) prospective utility, then they give no practical guidance to anyone. The combination of individual actions in shaping outcomes is alone sufficient to frustrate the practical application of such a criterion of moral rightness. As Ellis writes (334-335):

> My main criticism of comprehensive prospective utilitarianism is that it is utopian. In principle, it requires us to consider and evaluate the possible consequences of all possible combinations of actions, whether done by ourselves or by others, decide which set of actions has the greatest prospective utility, and to act accordingly. And the success of this strategy presumably depends on everyone else's making the same evaluations… ordinary human beings cannot do these things, therefore, it cannot be the case that they ought to do them.

In defence of Utilitarianism, it may be protested that many utilitarians –Singer (67-70) cites passages from Hutcheson, Bentham and Mill– have not set up this counsel of perfection, by which anything less than the utopian optimal choice would stand condemned. Rather, they have commended acts and policies which can reasonably thought to *tend* to increase utility, and to commend them the more emphatically, the more they do so tend (or may be reasonably thought so to tend). Informed Altruism recognises more explicitly than this non-utopian version of Utilitarianism that actions which are *intended* (after every reasonable step has been taken to acquire and deploy relevant knowledge) to create benefit may nevertheless, unluckily, fail to do so. And it entails that such actions are morally right even though the outcome is unlucky.

IV

Some recent philosophers have criticised Utilitarianism (or more broadly, consequentialism) on grounds that might be thought also to implicate Informed Altruism, and we can consider a couple of these criticisms. Again, I will try to show that Informed Altruism has the advantage in responding to these arguments.

James Lenman's 'argument from Cluelessness' maintains that it is futile to judge an action by the overall goodness of its consequences, since the consequences of any causally effective action will ramify massively and inscrutably through time, defying any hope of an ultimate cost/benefit calculation that could form the basis of an adequate justification. Immediate good consequences of an action can be outweighed by subsequent bad consequences that would not have occurred had the action been omitted.

Identity-affecting actions make this particularly clear.[25] If we assassinate some psychopathic tyrant to prevent him from causing the deaths of millions, we may, for all we know, be causing the birth, descended from among those millions, of still more prolific mass-murderers who would otherwise not have existed. Attempts to avoid this conclusion by arguing that later consequences will tend to balance out between good and bad, or that the impact of our action will in some unexplained manner taper away, are at best exercises in self-reassurance for which there is no empirical warrant. We are thus clueless with respect to the ultimate utility brought about by our actions.[26] Note that both FCU and ACU are vulnerable to Lenman's objection, since it implies that we can neither hope to foresee the final reckoning

[25] And as Parfit reminds us, almost any action can be identity-affecting (Parfit 1984, 351-379).

[26] Lenman 2000. Similar concerns over inscrutable ramifying consequences emerge in Norcross 1990, Frazier 1994, Howard-Snyder 1997 and Kagan 1998.

nor live long enough to wait for it. A final reckoning may, humanly speaking, never arrive. We can call this the long term epistemic objection.

In one way, these considerations strengthen the case for an intentionalist theory, by weakening the claims of a consequentialist one. However, although Informed Altruism denies that actual consequences are morally decisive, its claim that the moral faculty has developed historically because it serves to make human life better might seem be thrown into doubt by the long term epistemic objection. Moreover, since the point of an informed altruistic intention is to bring about a consequence, an ultimate cluelessness in the assessment of consequences might be thought to entail a similar cluelessness in the assessment of intentions.

However, unlike ACU, Informed Altruism requires the agent to fulfil an epistemic criterion expressed in terms of what is reasonable for her in her circumstances. Moral assessment of the rightness of her action is relative to her efforts, her capabilities and the seriousness of the issues at stake. Whereas the risks of unforeseen bad consequences of a well-intended action do not, as the argument from Cluelessness demonstrates, conveniently taper away over time, the ability of an agent to take reasonable steps to inform her decision *is* limited by the gradual recession of foreseeable effects into the future. What limits the reasonable expectation on the altruistically-intending agent is not, we should note, a purely chronological matter. A morally right action in personal relations, such as caring for a sick friend, could not plausibly require an effort to take account of possible events in the twentieth-ninth century, but a morally right action by someone professionally responsible for the safe disposal of nuclear waste, where scientific knowledge of potential far-future consequences is accessible and her possession and use of it is a reasonable expectation, might well be dependent on just such prevision.

As Lenman points out, appeals to currently-accessible information in justifying our decisions are very weak, if we apply to present-day decisions the expectations of an ultimate-outcome consequentialism. From the point of view of *that* theory, these appeals show, at most, that we have reason to choose A rather than B on the basis of such predictive indicators as we have. But these indicators, even when they seem intuitively compelling to us, are flimsy indeed, relative to the unknown benefits and calamities lurking in the vast inscrutable ramifying future to which our actions will contribute. There is no reason to suppose that an assessment of consequences from, as it were, a God's eye view across all the remaining centuries of human history would ratify any of our decisions. Thus an ultimate-outcome consequentialism holds human beings to an unachievable moral standard. We do not have access to this God's eye view, and ought to revise our moral thinking in recognition of that fact. Informed Altruism, in contrast, is an agent-focused theory: it views morality as a natural human faculty, exercised as best they can by transient individuals and communities. We can assume that this faculty will continue to exist, and that future individuals and communities will continue to exercise it in some way –probably as imperfectly as we do, more or less. As Dale Miller, who also sees the long term epistemic objection as fatal to ACU, remarks,

> whatever types of beings the universe may contain, moral standards are only meant to apply to humans – a proposition that might follow from a conception of morality as a human construct or institution…. [I]f no possible humans can satisfy a moral standard then it cannot be a valid one. (2003, 61.)

The scope of Informed Altruism, when considered as a guide to intentional action, is therefore far less ambitious than the scope of the ultimate-outcome consequentialism that

Lenman dismantles so effectively. It is necessarily limited to the salient, or practically researchable, possibilities of consequence initiated by the informed intentional acts of existing agents –even though (as in the nuclear waste example) these can include some quite remote consequences that our collective knowledge can predict with a fair degree of confidence. (Historically, of course, most past human communities would not have had any evidential basis at all for thinking beyond a generation or two, beyond an almost tacit assumption that the conditions of human life would continue to be much the same.) We existing agents, therefore, have to act with the indicators we have. People in future generations will have the ability to take informed intentional altruistic decisions too, and we have to share the exercise of the moral faculty with them. We will have contributed to causing these people to exist, and to forming the conditions in which they will operate, but we cannot predetermine their decisions, because for them the horizon of salience will be different, exposing a different field of possible action. The most we can do to help them is to provide models of altruistic action by our own acts, and preserve a discourse of moral thought through education, literature and other media. We should not expect to be able to prove, at the end of our lives, that our own attempts to exercise the moral faculty will be endorsed by some final consequential reckoning.

Lenman himself notes the possibility of a consequentialism reined in, so to speak, to concern itself only with 'visible' consequences, but adds that "a consequentialist must understand this concern as motivated by the belief that maximizing value with respect to visible consequences is a reliable means to maximizing value with respect to overall consequences. And this belief does not appear at all secure" (2000, 365). Alastair Norcross, whose refined consequentialist position provokes this reservation, actually makes a slightly more cautious claim:

> It is rational for a consequentialist to accept that a choice with foreseeable good consequences may, for all she knows, have unforeseeable bad consequences which massively outweigh the good consequences, and yet to make the choice purely on the basis of foreseeable consequences. (1990, 256).

Norcross here disclaims the belief that maximizing visible consequences is a reliable guide to maximizing overall consequences, yet continues to affirm a consequentialist view. The rational decision, on this conception of consequentialism, is the one that best reflects the cost/benefit calculation we can apply to the thin, maybe very thin, sliver of foreseeable possible outcomes. Norcross compares this marginal choice to the choice between causing seven million deaths and causing seven million and one, where we can make a difference to the one and this difference has some moral importance. (If "seven million" was replaced by "an almost infinite number", the point would still hold good.) As Norcross writes, "the cure [to the potentially demoralizing realization of the unknowable future] is to focus on what is within one's control" (256). Norcross's modest version of FCU is (in this respect) close to Informed Altruism, in which the criterion for morally right action is to do the best within one's powers, after accessing and deploying whatever knowledge is realistically available, to help and not harm other people.

Norcross's later paper 'Good and Bad Actions' provides further arguments to trouble an impartial consequentialist. He imagines that an Agent-physician, overseeing palliative care for a terminally-ill Patient, decides to achieve a small extra alleviation of suffering for Patient at the cost of a slightly greater imposition of suffering on herself, in the form of exertion and stress. Norcross describes this as a "suboptimal act that is nevertheless good", adding that

> The only motivation I can see for insisting that her action is not good, on the grounds that she can do even better, is the determination to equate the notions of the right and the good as applied to actions. (1997, 11).

The moral notation using 'right' and 'good' here may cause momentary confusion, as some may prefer to say that Agent's action was *right* in its intention, if sub-optimal (less than ideally *good*) in its effects. What Norcross means is that the optimal act would have been ideally good, but Agent's actual act was still good, and that the example intimates the need for a scalar, and not an all-or-nothing, attribution of 'good'.

As Norcross notes, whereas Utilitarianism requires observance of impartiality between the agent and others, some consequentialists will allow an action to be good even when the agent falls *below* the level of impartiality. Samuel Scheffler's 'agent-centered prerogative' (1994, *passim*) allows an action to be good when it is sub-optimal in its consequences but only by reason of a certain permissible bias towards the agent's self. Norcross's Agent/Patient's example shows the reverse, where an action is good when it is sub-optimal in its consequences but only by reason of an increment of altruism. Just as Scheffler postulates a limit to permissible bias towards oneself, Informed Altruism suggests a limit to altruistic bias towards the other person: that limit being the threshold at which the risk of weakening one's future capacity for effective altruism outweighs, on a reasonable assessment by the agent, the benefit that would be achieved in this present case by a further effort of altruistic action.

Parfit also suggests that we might sometimes have sufficient reason to decide to give, not merely equal, but somewhat greater weight to helping another. If I gave up my life to save that of an objectively-similar stranger, he remarks,

> I am inclined to think this act *might* be fully rational. This stranger's well-being matters just as much as mine. And if I gave up my life to save this stranger, this act would be generous and fine. (2011, I, 139).

This example may perhaps be dismissed as too quixotic to be generalized into a moral theory. But Norcross gives a more mundane example. If I get out of bed at night to calm my crying child, even though there would be slightly more utility in leaving the task to my wife who "has a slightly less burdensome day ahead", this is not "pointless masochism" (13). My action is non-optimal, but it is still, Norcross suggests, intuitively good. For Informed Altruism, getting out of bed is the morally right action, unless (allowing here for the need to be vigilant against a self-serving excuse) a more compelling altruistic project outweighs it. For example, suppose that my wife and I are both surgeons, but in the day ahead I need to be extra-alert because I will be performing an exceptionally complex operation on a gravely ill patient, while my wife will be attending a routine committee meeting. I would then be justified in leaving the task to my wife, and she would be doing what is morally right by taking it on. We do not need to look for corroboration from ultimate consequences to see that this is the moral faculty operating normally.

V

In this final section, I will review some of the principal differences between Informed Altruism and Utilitarianism as social theories, and in doing so, summarise some key points from the earlier discussion.

Informed Altruism assigns positive moral value to the right or commendable individual action (*my* sufficiently informed intentional action to help and not harm others). Collective actions, institutions, moral codes and the like equally derive their moral value from sufficiently informed collective and reciprocal endeavours to help and not harm others. The creation of morally right or commendable actions, institutions, codes, policies, practices, etc. just is the exercise of the moral faculty by human beings, who, along with the other remarkable faculties they have developed throughout their life-history (medicine, technology, music, etc.), have this one, morality, whereby individually and collectively they intentionally take the interests of others as their aim.

It would be possible, while remaining within a naturalistic meta-ethical framework, to accept that there exists a human faculty properly named 'morality', yet to argue that it is quite different from the one I have described. However, to be a credible competitor, any alternative characterisation of the faculty would need to have something of the same simplicity and the same capability of complex and diachronic development. Morality, like technology, must be the sort of faculty that most human beings can understand, at least in its rudimentary form, and be capable of using intentionally and habitually.

It would also be possible to deny that any such natural faculty as morality exists, or to appeal, as falsifying examples, to the existence of codes of behaviour that clearly flout the altruistic principle. On a possible version of this alternative view, the true natural faculty would be that arising from our cognitive capability of forming and being guided by codes: 'morality' would then either have to be identified with this code-following faculty, or denied the status of a natural faculty at all. But it seems more illuminating to characterise these exceptions (e.g. the

action-guiding codes of the Aztecs, or the Nazis, in their relations to other human groups) as non-moral, or in their predominant tendency immoral, codes. We can then retain a useful substantive conception of morality.

Since Informed Altruism is an intention-focused theory, it begins by adopting the perspective of a single human being: it addresses each agent in the process of forming, and carrying through into action, her intentions. To be an altruist (or rather, if *I* am an altruist) I should favour your and others' interests over my own, except in so far as preserving my life, health and capabilities of action is necessary to ensure that, over the long term, I can continue to be an effective altruist. If you are I are *both* altruists and find ourselves in an active relation to one another, we can sometimes benefit from one another's altruism by reciprocal acts of kindness, but will sometimes have to negotiate a mutually satisfactory compromise: we cannot both take the smaller of two slices of cake, or carry the heavier of two burdens. But then, that is also true if we are both egoists, and need or want to live together harmoniously. The negotiation between altruists –imagine them as two castaways establishing a shared life on a desert island– is not likely to be more difficult than the negotiation between egoists in the same situation. We can extend this thought to wider communities. A social system in so far as it is informed by the moral faculty (understood as mutual informed altruistic intention) is an eminently plausible project. Since each person within the system is an object as well as a subject of unconditional altruism, it can hope to sustain the pressures of egoistic need and desire within an orderly scheme of co-operation.

The Utilitarian project requires, for optimal effectiveness, a system of social organisation, law and education whereby the diverse actual motives of individuals may be channelled to produce the best consequences. Individual altruistic intention is at best supererogatory within

this system, much as the need for any personal moral commitment to equality and care for the disadvantaged is bypassed by the 'scientific' historical narrative of Marxism.[27] Informed Altruism, in contrast, recognises the independent existence of the capability for altruism, which is activated as we imitate and reciprocate the care shown by those closest to us. We are able to perceive analogies between the goods and harms of these loved ones and those of a progressively wider conspectus of human beings –the friends, parents, siblings, children of strangers. Conscious practice of the moral faculty, aided (usually) by the moral code or codes in which we are educated, can give these altruistic intentions more focus and order. It is true that egoistic reflection on our own goods and harms can also generate, by analogy, concern for others.[28] But, equally, our self-love can generate an ever wider conspectus of people to be regarded with malice or indifference –rivals, competitors, people who get in the way, or are too much trouble to bother about.

A further difference between Informed Altruism and Utilitarianism lies in the extent to which each view could strategically implement the other's recommendations. A utilitarian might come to believe that the practice of favouring others' interests should be fostered and commended, because such a practice across some relevant population tends to produce the best consequences: but it would be possible to think the opposite and still be a utilitarian. The position for an altruist, however, is not symmetrical with this. An altruist might come to believe that the practice of favouring one's own interests, or of weighing one's own interests equally with others, would tend across some relevant population to produce the best consequences. It would be consistent with her altruism to co-operate with that strategy, but only if the unconditional character of the altruistic intention were guaranteed in one of two

[27] cf. Cohen 2001, 101-115.

[28] cf. Bentham 1834, II, 36-37.

ways. Either (1) *all* persons within the scope of the relevant situation must benefit, so far as she can reasonably judge, from her self-interested actions, or (2) her self-interested actions must represent the resolution of the moral conflict that arises when *not all* persons can benefit. An example of the latter would be a doctor's saving her own life rather than another's during some emergency, when her survival would make it possible to save additional lives. Naturally, a doctor might prefer to save her own life simply because she wishes to live. But if her own wishing to live were the determining reason, then in that action she would not be acting morally: rather, she would be acting according to a natural and powerful non-moral imperative. On the other hand, an agent who urges *another* to act self-interestedly might well be acting morally, because altruistically. In encouraging a guest to eat a second helping of strawberries, one is altruistically offering him a self-satisfying pleasure. The guest's action in accepting the offer may also not be wholly egoistic: it may have an altruistic tinge in so far as he knows he is giving pleasure to you by taking pleasure himself.

How significant are these differences between Informed Altruism and Utilitarianism when we move from private acts to public policies and systems? The two theories are quite close in their political implications. Indeed, it may seem that Informed Altruism effectively turns into Utilitarianism as soon as it moves from individual acts of altruism to an institutionalised social practice. According to Informed Altruism, we should take *every* person affected by such a practice as a possible object of altruistic concern. Since this creates problems of priority, how, it might be asked, are these to be resolved, other than by treating each person as counting equally, and weighing utilities? However, there are differences between the two theories even at this level of application. I conclude by mentioning two of these, both of

which are derived from the unconditionality of altruism and the consequent frequency of moral conflicts.[29]

First, Informed Altruism draws attention to the continuing moral significance of the harm suffered when a moral conflict is (as is commonly the case) resolved in an imperfect way. Utilitarianism concedes that acting rightly, by maximising utility, can lead to regrettable harm for some; Informed Altruism identifies such causing or allowing harm not merely as regrettable, but as a moral failure. The distinction may seem a merely rhetorical one, but the rhetoric of human interactions can be important: Informed Altruism avoids the suspicion to which Utilitarianism is exposed, that of perceiving the suffering of the harmed person as part of a situation which, morally speaking, has been satisfactorily resolved. When, for example, the harmed person is a criminal punished for the benefit of others, Informed Altruism thus offers an especially emphatic repudiation of the idea that it is good in itself that the criminal be caused to suffer. Utilitarianism recognises the criminal's suffering as bad, and awards it a negative 'weight'; Informed Altruism additionally notes that the morally required intention to help and not harm a person has failed to be enacted. (A theory is not discredited when it runs up against an irremovable irreconcilability of ends. The best it can do is acknowledge and explain that irreconcilability.)

Secondly, there may be actions that express Informed Altruism best, but do not correspond to a utilitarian judgement. Practices which lead to the greatest net happiness might, as Scanlon argues, be wrong if they require an especially glaring failure of consideration towards an individual. (2000, 153-155 *et passim*). (This is not the same thing as attributing an indefeasible right to that person –that is to say, Informed Altruism does not dissolve into a

[29] See 13 above.

Kantian rights-based theory. Possession of a right is not a prerequisite for altruistic concern: altruistic concern is unconditional.) At the collective level, Informed Altruism requires that a society formulate policies and practices based on the principle that altruistic concern is to be applied to every person. This is not the same as applying the principle that each person counts one when a weighing of utilities is carried out. Informed Altruism might justify, for example, a minimum level of welfare for every person, even if, as a result of this policy, aggregate utility was less than could be achieved by abandoning that minimum.

Utilitarianism is a consequentialist, and an impartialist, theory. There might not seem to be much hope for the Utilitarian tradition if both these foundations of the theory are undermined. But it has other strengths, untouched by the criticisms developed in the present paper: and principal among these is its clear-headed focus upon the common-sense objectives of aiding human happiness and mitigating suffering. Though Bentham marginalised altruistic intention within his classical statement of the theory, and his successors have largely followed him in this respect, moral philosophy would, I believe, benefit by calling that marginalisation into question.

REFERENCES

Adams, Robert Merrihew. 1976. Motive Utilitarianism. *The Journal of Philosophy* 73: 14 467-481.

Bentham, Jeremy. 1834. *Deontology*, edited by J. Bowring. London: Longman, Rees, Orme, Green, Browne and Longman.

Bradley, Ben. 2006. Against Satisficing Consequentialism. *Utilitas* 18: 2 (2006) 97-108.

Cohen, G. A. 2011. *If you're an Egalitarian, How Come you're so Rich*? Cambridge, Mass.: Harvard University Press.

Dewey, John, and Tufts, John Hayden. [1932.] 1985. *Ethics*, edited by Jo Ann Boydston. *John Dewey: The Later Works*, Volume 7. Carbondale, Il.: Southern Illinois University Press.

Ellis, Brian. 1981. Retrospective and Prospective Utilitarianism. *Noûs* 15: 3 (1981), 325-339.

Feldman, Fred. 2006. Actual Utility, the Objection from Impracticality, and the Move to Expected Utility. *Philosophical Studies* 129 (2006), 49-79.

Frazier, Robert L. 1994. Act Utilitarianism and Decision Procedures. *Utilitas* 6: 1, 45-53.

Gruzalski, Bart. 1981. Foreseeable Consequence Utilitarianism. *Australasian Journal of Philosophy* 59: 2 (1981), 163-176.

Hill, Thomas E., Jr. 1991. *Autonomy and Self-Respect*. Cambridge: Cambridge University Press, 1991.

Howard-Snyder, Frances. 1997. The Rejection of Objective Consequentialism. *Utilitas* 9: 2, 241-248.

Jackson, Frank. 1991. Decision-Theoretic Consequentialism and the Nearest and Dearest Objection. *Ethics* 101: 3, 461-482.

Kagan, Shelly. 1998. *Normative Ethics.* Boulder: Westview Press.

Lewis, C. I. 1969. *Values and Imperatives*. Stanford: Stanford University Press.

Mason, Elinor. 2019. *Ways to be Blameworthy: Rightness, Wrongness and Responsibility*. Oxford: Oxford University Press.

Miller, Dale E. 2003. Actual-Consequence Utilitarianism and the Best Possible Humans. *Ratio (new series)* XVI, 49-62.

Moore, G. E. 1907. *Ethics*. New York: Henry Holt & Co.

Norcross, Alastair. 1990. Consequentialism and the Unforeseeable Future. *Analysis* 50, 253-256.

Norcross, Alastair. 1994. Good and Bad Actions. *Philosophical Review* 106: 1, 1-34.

Parfit, Derek. 2011. *On What Matters.* Oxford: Oxford University Press.

Parfit, Derek. 1984. *Reasons and Persons.* Oxford: Oxford University Press.

Scanlon, T. M. 2000. *What We Owe To Each Other*. Cambridge, Mass.: Harvard University Press.

Scheffler, Samuel. 1994. *The Rejection of Consequentialism*. Oxford: Oxford University Press.

Searle, John R. 1969. Deriving 'Ought' from 'Is'. In Searle, *Speech Acts*, 175-198. Cambridge: Cambridge University Press.

Sidgwick, Henry. 1874. *The Methods of Ethics*. London: Macmillan.

Singer, Marcus G. 1977. Actual Consequence Utilitarianism. *Mind* 86 (341), 67-77.

Stark, Cynthia A. 1997. Decision Procedures, Standards of Rightness and Impartiality. *Noûs* 31: 4, 478-495.

Strasser, Mark. 1989. Actual Versus Probable Utilitarianism. *Southern Journal of Philosophy*, 27: 4, 585-597.

Temkin, Jack. 1978. Actual Consequence Utilitarianism: a reply to Professor Singer. *Mind* 87 (347), 412-414.

Warnock, G. J. 1971. *The Object of Morality*. London: Methuen.

Wong, David B. 2006. *Natural Moralities: A Defence of Pluralistic Relativism*. New York: Oxford University Press.