

Acts and Consequences:
An Analysis of Modern Act-Consequentialism

Cameron Chandler-Haines

A Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the
MA degree in Philosophy

Philosophy

Arts

University of Ottawa

©Cameron Chandler-Haines, Ottawa, Canada, 2016

Abstract

In this thesis I look at the modern state of Act-Consequentialism and how it can defend against five of the more prominent objections found in recent literature on the topic. The five objections are as follows: AC's neglect of special personal relationships, its overdemandingness, its inability to handle expectation effects, its incompatibility with the claims of justice, and its suppression of our moral integrity. Although each of these objections differ significantly from one another, and this leads to separate treatment of each by chapter, all the objections deal on some level with the "derivativeness" of Act-Consequentialist moral thought when analyzing questions of a seemingly intrinsic quality, such as special personal relations or the dictates of justice.

The approach taken towards the objections is broadly similar. By carefully considering the nature of the well-being that is promoted under AC, and how to most efficiently and organically promote it, the most severe aspects of the objections are softened. In working through the objections it is hoped that a more plausible form of Act-Consequentialism, with a stress on organic and efficient well-being promotion, will emerge.

Table of Contents

Introduction - iv

Chapter 1: Personal Relations – page 1

Chapter 2: Overdemandingness – page 18

Chapter 3: Expectation Effects – page 37

Chapter 4: Justice – page 61

Chapter 5: Integrity – page 94

Conclusion – page 119

Introduction

In the following thesis I intend to examine five popular objections to the moral theory known as Act-Consequentialism (henceforth AC). The five objections are, in the order of which they will appear, AC's neglect of special relations, its overdemandingness, its inability to handle expectation effects, its incompatibility with the claims of justice, and its suppression of our moral integrity.

I take AC to be the teleological moral theory that says an act's rightness is determined by the overall goodness of its global consequences. This differs from Rule-Consequentialism in that AC evaluates only isolated acts, whereas Rule-Consequentialism, on one recent influential version, evaluates acts by whether or not they adhere to rules "whose communal acceptance would have the best consequences" (Hooker 1). AC differs also from the more specific Act-Utilitarianism in that the value that makes up the "good" to be promoted is left open, whereas for Act-Utilitarianism the value is specified. Hedonic Act-Utilitarianism, for example, is an Act-Utilitarian theory because it specifies that the value to be promoted is pleasure (and, by the same token, pain is to be avoided) (Sinnott-Armstrong para. 2).

For the purposes of this thesis I will refer to the good being maximized in AC simply as "well-being." It includes, beyond pleasure and happiness, human values such as living autonomously, having close friendships, achieving valuable goals, etc. (Hooker 39). This list by no means exhausts the component parts of well-being, and the full list will be left open. There is, admittedly, a potential danger in remaining "agnostic" about well-being in this way (Mulgan 23). Changes in the conception of

well-being might lead to different views about the functioning of the theory as a whole. However, I will largely leave this question off the table and will treat well-being as closely aligned with happiness.

My definition of AC differs from the common definition in that I see the theory as primarily concerned with well-being “promotion” as opposed to “maximization.” Typically the theory is interpreted as claiming that “an act is morally right if and only if that act maximizes the good, that is, if and only if the total amount of good for all minus the total amount of bad for all is greater than this net amount for any incompatible act available to the agent on that occasion” (Sinott-Armstrong para. 1). I do away with this focus on maximization because I think it leads to counter-intuitive moral evaluations that a focus on promotion does not. (I argue for this in the chapter on overdemandingness). I find the shift to well-being promotion does not severely alter the overall form of AC and so the theory should still be seen as working in much the same way.

Finally, the form of AC I will be dealing with focuses on expected consequences as opposed to actual consequences. (I argue for this in the chapter on expectation effects). As with the move to well-being promotion, I think this form of the theory is far more plausible because it does not lead to the same counter-intuitive ethical judgments. The use of expected consequences plays an important part in the defense against the five objections, particularly those concerned with expectation effects and justice. Within expected consequences there are many varieties (such as foreseeable, foreseen, intended, or likely consequences) (Sinott-Armstrong para. 5). Although I take foreseeable consequences to be the most

attractive variety, I do not argue for this in the thesis, but merely focus on expected as opposed to actual consequences.

Methodologically speaking, the treatment of the objections will focus primarily on the procedure of well-being promotion, as opposed to focusing on the actual make-up of the well-being that is being promoted. So, for example, in looking at the objection on personal relations, I look at how close relations promote well-being (such as how being aided by a loved one will promote greater well-being than being aided in the same way by a perfect stranger). (Other examples include the process of Diminishing Marginal Utility and the anticipation of expectation effects). I take this approach instead of, say, simply including close personal relations within well-being. As such, much of the argumentation used in the thesis could equally be applied to Act-Utilitarianism, since well-being will often take the same form as happiness. Although, because of other objections, such as G.E. Moore's, I do think well-being should take a much more complex form than happiness alone, this need not be the case for my approach to the five objections discussed in this thesis (Moore 59).

Another important methodological point concerns the use of intuitions in evaluating the objections. Although I discuss the question in more detail with reference to the works of Peter Unger and Shelly Kagan (who seek a "liberationist" or "extremist" approach to ethics that opposes the primacy of our moral intuitions) I should point out here that I take it that our moral intuitions have their own independent validity. One may argue against this that, on occasion, common folk intuitions about morality (even our considered moral convictions) are

demonstrably erroneous, as in the case of slaveholders in American and Canadian history (at least, in comparison with our present day intuitions) or in the modern day dismissal of far off third world suffering (Unger 7). Despite these mistakes, we should not, as a result, doubt our intuitions as a whole. As Frank Jackson puts it, “we must start from somewhere in current folk morality, otherwise we start from somewhere unintuitive, and that can hardly be a good place to start” (Jackson 135).

I should note also that, for the purposes of the thesis, intuitions should be taken as our considered moral convictions as opposed to mere unreflecting, instinctive or impulsive reactions. A valid intuition then will take in elements of what Rawls describes in his assessment of the competent moral judge (Rawls *Conception* 178). Such a person must be intelligent, though they need not be exceptionally so, for moral insight should not be the purview of only the elite. They should also be knowledgeable about the specific circumstances, reasonable in the sense of understanding and suppressing their own prejudices, and they should have a sympathetic understanding for the pain (or pleasures) involved in a given scenario. The scenario should not be one where the respondents face personal gain or loss (so for example, in the trolley case, legal repercussions of the decision do not play a part). Additionally, the scenario should not be wildly unrealistic and all relevant underlying facts should be known (Rawls *Conception* 178). In the thesis I try to adhere to these standards, or raise it as a concern when these standards do not apply.

As against Kagan and Unger, I follow Rawls in observing a degree of circularity in the liberationist approach. Therefore a valid moral intuition must “not

be determined by a systematic and conscious use of ethical principles. (...) It is clear that if we allowed these judgments to be determined by a conscious and systematic application of these principles, then the method is threatened with circularity. We cannot test a principle honestly by means of judgments wherein it has been consciously and systematically used to determine the decision” (Rawls *Conception* 182). “We wish to say of some principles for adjudicating interests” (e.g. AC) “that one ground for accepting them as reasonable principles is that competent judges seem to apply them intuitively to decide moral issues. Obviously if a competent judge were defined as one who applies those principles, this reasoning would be circular” (Rawls *Conception* 180).

I should note, however, that the validity of this method is based on avoiding the notion that we start “moral theorizing with some sort of moral theory in mind” already (Kagan 12).

Ultimately, the intuitional approach is supported by the consideration that we lack viable alternatives. To date no convincing method of ethics has been found that eschews our intuitions. Indeed, it is difficult to imagine how it could. If someone were to approach you with a newly crafted moral theory that is logically elegant and perfectly consistent, but happens to endorse the idea that we should arbitrarily kill innocent people, we would of course find fault with their theory, not our intuition that arbitrarily killing innocents is wrong. The language of common moral discourse (and the methodology of some of the most prominent ethicists of the last several decades, such as Rawls and Ross) takes our most grounded intuitions as valuable independent data.

In the end I will attempt to bolster the case of AC in regards to the five objections following the method, put forward by Sidgwick, that takes our intuitions as “valid alongside, and indeed as the foundation of, what he calls the alternative ‘methods’ of utilitarianism” (Moore 111).

One might object that if our intuitions can already deliver appropriate moral responses, then what need have we for the guidance of a moral theory?

We could, of course, simply assert that a moral theory like AC need not provide guidance, but only comprises an accurate assessment of the underlying rationale behind our several valid moral intuitions.

Beyond this, however, we could still uphold AC’s role as a moral guide. Imagine the analogy of building a shelter: We could ask, why would one build a shelter, if, without any shelter, one has managed to live well enough and have preserved enough strength to be able to build a shelter in the first place? The answer is that they build a shelter in preparation for when they one day require more rest or protection from the elements in order to preserve a sufficient level of strength and energy. Similarly, in the case of a moral theory, we build one up in order to prepare for cases where our intuitions do not deliver a clearly correct and unbiased response, as in the previously cited cases of slave holding or neglect of far off third world suffering.

In light of these considerations about the validity of our intuitions, I think it appropriate as a methodology to take my own intuitions in order to determine the judgments of common moral sense. I use this method as opposed to delving into extensive research of what most people would consider the considered moral

convictions to be. However, to avoid any mistakes that might arise out of this approach, I only make use of my intuitions when delivering (what I take to be) a fairly obvious moral response, and not in morally controversial scenarios. For example, I refrain from judgment in considering the competing moral claims of neglecting your closest loved ones in order to help with the more severe suffering of far off strangers. In cases like this I merely point out the controversy that seems to exist and competing claim that we could observe, i.e. the inhumanity of neglect of close personal relations vs. the selfishness of ignoring far off suffering.

As a final remark on methodology I would like to explain my choice of the five objections and how they are mutually related.

The five objections figure prominently in many popular overviews of AC (Hooker 150, Kymlicka 21, Shafer-Landau 32). So, inasmuch as these objections are taken to be prominent in the recent literature, and new (and I hope convincing) approaches can be taken in arguing against them, they are fit for inclusion in this thesis.

Furthermore, the five objections are included together because, to a large degree, the objections overlap in important ways. For example, Dale Dorsey finds that the integrity objection can be included as a derivative form of the overdemandingness objection. In many ways, the personal relations question can be seen as derivative of the integrity objection, which deals, in part, with one's personal commitments. However, I believe that by separating the five objections into distinct parts, we can better highlight different features of the interrelated objections, and in doing so I hope to avoid any repetition.

The objections also belong together because they can be defended against with a broadly similar approach. For all the objections, except integrity, we can argue against them by considering similar ideas. Firstly, we can make use of a refined conception of efficiency in well-being promotion, and what sort of elements to focus on in this promotion. By putting a “thumb in the scales” for important human concepts, such as close personal relations, promise-keeping, and fair treatment, we can offer these ideas their deserved position within impartial AC. The second method is to look carefully at the exact nature of the well-being being promoted. As I have already noted, this need not require a thorough list of the components of well-being, but instead it will do simply to understand that well-being is familiar element central to human life. So, in the case of personal relations for example, it is by considering this aspect of well-being that we can combat against the idea that AC dismisses these essential social practices. In seeing well-being this way, we can avoid the notion that AC is incompatible with the “warm and spontaneous expressions of emotion,” and this informs much of the defense of the objections (Kaplan 228). In this respect the integrity objection must be included as well, since it is here where I directly argue for the central quality of well-being in our lives, and that pursuing it will not lead us to dismiss our cherished commitments, since well-being is made up of, indeed defined by, precisely these cherished commitments. The third method of argumentation, broadly common to all the objections, is to allow for modifications to the overall form of AC. So, for example, in the case of the overdemandingness objection we can pursue a scalar

form of AC or, for the expectation effects objection we can focus on the “expected” form of AC.

The five objections also relate to AC in a similar way, namely they all have to do with AC’s nature as a teleological theory that seeks to promote a single end. We can see how all the objections are based on upholding a separate end. The personal relations and justice objections straightforwardly contend that well-being promotion neglects the promotion of special relations or fairness for their own sake. The expectation effects objection looks closely at treating certain moral practices, such as promise-keeping or telling the truth, as independent of well-being promotion. The integrity objection deals directly with the neglect of our ground projects in the name of well-being promotion. The overdemandingness objection examines not just the workability of maximizing a single standard, but also more generally with the validity of demanding that we neglect the intrinsic importance of our ground projects.

Ultimately, I seek to defend against the objections in a way that builds from one chapter to the next, with a broadly similar methodological approach (that focuses on refined elements involved in well-being promotion and the ultimate goals of a sophisticated AC theory). However, I will of course seek to avoid a dogmatic approach and will not shy away from any weaknesses in AC that these objections might expose. Additionally, this thesis does not put forward an exhaustive defense of AC, but only raises important considerations with respect to the five objections in question. It is hoped that by elucidating possible defenses of AC on these lines, a clearer picture of a more plausible form of AC will be presented.

Chapter 1: Personal Relations

The exact scope and limit of special relationships and their associated obligations raises a central question within ethics. We tend to criticize those who undervalue the moral importance of special relationships and the stringency of their associative duties. So, for example, a mother who leaves her family to help organize charity efforts in third world countries is said to be lacking in the essential human virtue of familial compassion. Conversely, we also criticize those who overvalue special relationships to the point of a “pernicious prejudice” (Scheffler *Relationships* 194). As an example, consider the injustice we perceive in acts of nepotism. Both examples are extreme, and seek to draw out very obvious moral judgments. However more difficult situations can be imagined. Consider the famous trolley example, but replace the one stranger who most would willingly sacrifice with one’s own brother. Many would change their decision in this new scenario (not just as to what they would do, but what objectively a person ought to do) on the grounds that killing your own flesh and blood to save five strangers evinces a cruel inhumanity. Still others would see in this sacrifice a difficult, yet ultimately laudable, moral choice. The diversity of opinions demonstrates the centrality of this issue within ethics.

In terms of the debate over the merits of AC the special obligation question is widely considered by critics to highlight the major weaknesses of the theory. Generally stated, AC has been considered guilty of undervaluing the importance of our special obligations in overemphasizing the moral necessity of impartiality.

In this chapter I will begin by discussing AC impartiality, and raise some of its strengths in relation to the personal relations objection. I then go into a discussion of what type of relationships should count as “close personal relations.” Next I present the method of argumentation against the objection, which will follow Sidgwick’s method of understanding these relations in such a way that they will have a strong utilitarian (and ACist) weight, even impartially considered. I divide Sidgwick’s arguments into four groups. First, the epistemic proximity of close relations will elevate their impartial worth. Second, both through internal satisfaction and increased motivation, AC will prize these relations. I call this the “agent reverberation” group of Sidgwick’s arguments. Third, an agent helped by a loved one will be especially motivated to go on and help others. I call this the “spillover effects” group of arguments. Fourth, the object of beneficence’s well-being will be magnified knowing this beneficence comes from a special relation. I call this group the “magnification” arguments. I analyze each of these groups of arguments, assessing their individual validity. I deduce that, in the end, these arguments taken together leave us with a difficult decision in terms of special relations, but this difficulty is not unlike the one our intuitions observe in these scenarios. Finally, I look at other forms of the objection against AC personal relations, such as Hooker’s argument via expectation effects, and show that Sidgwick’s form of arguments should militate against these objections as well.

The impartiality of AC can be seen as one of its strengths. Bentham’s dictum that “every man is to count for one and none for more than one” provides a pivotal moral insight. By stressing the impartial “point of view from the universe” AC does

well to recognize the significance of sympathy within a moral theory (Brink *Forms* 393). Indeed the most common immoral acts seem ultimately rooted in selfishness, an overvaluing of one's own desires and aversions over and above those of others. Immoral acts of selfishness extend to a partiality towards those with whom we are closest, that is, caring much more for the good of our close relations, as opposed to the good of perfect strangers. The immoral selfishness of partiality can be observed in the case of nepotism, racism, or indeed, preoccupying oneself with the comparatively trivial needs of one's close relations when a stranger's severe suffering could otherwise be relieved.

Despite the merit of AC's stress on impartiality, within the context of special obligations, many critics believe that the theory goes too far. Most would admit that our common sense observes something deeply unsettling in acts of, say, familial neglect, even if this neglect comes in the name of improving the lot of others. As Sidgwick noted, there is an "inhumanity" involved in such acts of neglect (Sidgwick 250). And, on the face of it, AC does seem committed to this inhumanity. The theory tells us to consider everyone impartially, and to maximize the good for everyone. As such, a clear corollary is that we should neglect our closest relations if the harm caused by this neglect is outweighed by the good done for others. Under modern circumstances of massive wealth disparity and world poverty, it certainly seems as if spending our energy alleviating the suffering of far off strangers would maximize the good. Faced with this difficulty, many authors choose to reject AC altogether (such as Hooker or Williams), or to advocate a hybrid theory that outlines a

restricted form of AC when matters of personal relations are considered (for example, Scheffler advocates this) (Scheffler *Rejection* 6).

Who should count among our special relations? Although Woodcock and Brink show this is a relevant question that is too often overlooked in this discussion, for our purposes it will do simply to consider a special relationship as the kind involved between close friends or members of a nuclear family (Woodcock 4).

In his lecture, *Self and Others*, Broad introduces the notion of self-referential altruism to explain our common moral sense of how one should treat one's special relations (Broad). Self-referential altruism deems that our moral attitude is altruistic, in that acts of self-sacrifice for the benefit of others, any others, even far off strangers, are morally praiseworthy. However this altruism is also said to be "limited in scope" because "each of us has specially urgent obligations to benefit certain individuals who stand in certain special relations to himself" (Broad). Brink objects to the terminology, and instead suggests that our altruism is unlimited in scope (that is, it includes all) but that it is variable in weight (there is a greater duty of altruism the closer one stands in relation to us) (Brink *Perfectionism* ch. 19). Brink suggests a concentric circle schematic to explain how to conceive of this idea (Brink *Impartiality* 159). Broad wonders whether this "common sense position can be circumvented by someone who found the (...) utilitarian principle self-evident" (Broad). In order to do so, a utilitarian would have to "show that all those special obligations which commonsense takes to be founded directly upon special relations of others to the agent, are derivable from one fundamental obligation to maximize the balance of good over evil in the lives' of everyone" (Broad).

I believe that it is possible to reconcile AC and our moral intuitions, and that it can be shown that AC does not advocate anything drastically contrary to our intuitions in the case of special relationships, or, in the words of Amartya Sen, special relations can be “caught in the impersonal metric of utility” (Sen *Introduction* 6). Sidgwick dedicated much of *The Method of Ethics* to showing how this reconciliation is possible when we employ sophisticated and subtle appraisals of utility within the utilitarian theory.¹ He remarks that, “the difficulties which we found in the way of determining by the Intuitional method the limits and the relative importance of these duties are reduced in the utilitarian system, to difficulties of hedonic comparison” (Sidgwick 439). If we take a sensitive enough approach to hedonic comparison, many of these difficulties will dissolve, revealing “the latent utilitarianism of common moral opinion” (Sidgwick 436).

In the case of special relationships, Sidgwick readily admitted the moral worth of committing oneself to those with whom we are closest. As he says, most deviations from this commitment evince a sort of “inhumanity” (Sidgwick 250). However, if we consider the matter carefully, staying true to our special relations is indeed often the most efficient use of our energy in maximizing well-being (or, for Sidgwick, utility) and as such would be advocated by AC or hedonic utilitarianism.

Sidgwick outlines three broad reasons for why utilitarianism need not clash with our considered moral convictions on the issue of increased concern for special relations. Firstly, he shows that our increased familiarity with our closest relations

¹ Sidgwick’s arguments support a simple hedonic utilitarianism, not the more sophisticated AC theory. However, if a reconciliation between utilitarianism and common sense can be shown, then I believe the same argument could easily be extended to AC.

will mean we are more effectively placed to aid than we are for perfect strangers. This means a utilitarian analysis will demand greater concern for our loved ones, all else being equal. However, especially in our modern circumstances of tremendous wealth disparity, all else is not equal. As such more is required to show an AC justification of devotion to our closest relations. Sidgwick then adds a second “thumb in the scales” for a utilitarian reconciliation with common sense. He shows that our increased concern for others is internally satisfying. This leads to two distinct sub-justifications. The first is that, as helping a loved one gives us more pleasure than helping a perfect stranger, this pleasure on the agent’s part must be counted within the overall utilitarian calculus. The second distinct sub-justification relates to Sidgwick’s “government house” utilitarianism. Because helping a loved one will be pleasurable (more pleasurable than helping a perfect stranger), and because we are naturally inclined to do that which we find pleasurable, acts of this kind have a built in motivator. Along these lines, nothing is superior in our benevolence for loved ones as opposed to our benevolence for strangers, and yet because of the built in motivator, more good will be done if agents pursue the former kind of benevolence. It is a familiar argument in modern capitalism and the motto, “greed is good” or, more accurately “greed leads to good” fits the exact same mold.² Sidgwick then offers a third “thumb in scales” argument. He says that acts of benevolence to our close relations will involve a sort of “spillover effect” (my term). When, for example, a child is cared for and nurtured by their parent, even when (in a more direct sense) more good could be done if the parent took his efforts elsewhere,

² In AC, as in capitalism, it is ideal to harness the selfishness of individuals to achieve a greater, communal end.

the difference is more than made up by the child's future moral decision making. Such a child, nurtured and well-adjusted, will live a life of benevolence and goodwill to the end of his days. A life of benevolence will far exceed the loss involved in his parent's increased concern for him.

Sidgwick offers a fourth reason, though only briefly and in passing. It is closely related to the second and third reasons. He says that "the development of such affections is of the highest importance to human happiness" (Sidgwick 439). Conversely, he explains that most members of society expect this affection from their loved ones and that "the disappointment of such expectations is inevitably painful" (Sidgwick 439). The basic idea is that, it matters for a person not just that someone is benevolent to them, but that that person is someone with whom they share a special relationship. It might be said that the same act of benevolence is worth more to a person in direct proportion to how close one's relationship is with the benefactor. It is a phenomenon of which almost everyone is deeply aware, and examples abound. Consider the common notion that a child's parent is considered, in most cases, to be their best possible guardian. Of course this is not because one's parent is always the most well-trained or the most intelligent guardian, but because the child already shares a special relationship with their parent, and the parent is uniquely situated to reinforce this bond as a guardian. Consider another example: the case of home cooking. A guardian provides much more than only food in this case, and most people recognize a profound value in a home cooked meal, even when compared with a meal provided by a much more capable cook. Because the benefactor in this case is someone with whom we share a special relationship, it

offers far more than a full belly, but a sense of community, self-worth, and feeling loved. Such pleasures would surely rank quite high on Mill's scale. Indeed, it is a principle of which Mill seemed fully aware, evidenced by his claim in *Utilitarianism* that, "few hurts which human beings can sustain are greater, and none wound more, than when that on which they habitually and with full assurance relied fails them in the hour of need" (Mill *Utilitarianism* ch.5).³ As such it should be considered the most impactful of the "thumb in the scales" arguments, one that truly catches the sense of special relationships within "the impersonal net of utility" (Sen *Introduction* 6).

Considering, also, that this is a widely recognized phenomenon, it is highly surprising that it has not been more widely discussed in the literature. Apart from Sidgwick's sidelong glance, another mention of this argument can be found in Mill's *Utilitarianism* (quoted last paragraph).⁴ I have also found very brief mention of it in Broad's "Self and Others" when he says "each of us is an object of interest, affection, and *natural expectation* only to a limited class of his fellow-men" (Broad). Among modern scholars it is hardly mentioned. Woodcock, however, comes close when he says there is a "profound value of friendship for human agents" but he does not stress the key value this has for the *object* of friendly acts (Woodcock 1).

Are Sidgwick's "thumb in the scales" arguments convincing? Does he adequately elevate the Consequentialist concern of special relationships to a point where it conforms to our common sense?

³ Mill uses the term "that" on which we rely, instead of saying "those" or "they" on which we rely. This way of talking truly stresses the impersonal nature of a utilitarian calculus.

⁴ It is interesting that in all cases this phenomenon is only referred to in the negative sense: that is, not in the pleasure associated with it, but with the pain associated with its loss.

The first argument from “efficiency,” as some have dubbed it, can be argued against in various ways (Kapur 488). As Hooker notes, the idea that we are better positioned to help our closest relations as opposed to suffering strangers is far less convincing today (Hooker 153). (This might explain why arguments that rely on the special well-being one receives from a loved one’s affection are so notably absent in modern scholarship. There may be a conflation with that argument and the dismissed “efficiency” argument currently under discussion.) Resources can easily be transferred and the disparity of wealth is so great in the modern world, that even highly inefficient resource reallocation will bring about far more good than would be lost (Hooker 153).⁵ The second form of the argument, that we are better epistemically situated to help our closest relations, seems to fall short. Admittedly, humans can have highly different psychological makeups, and our intimate relations can better serve our more subtle desires. However, except in rare cases, the pain of hunger and disease will affect us all in much the same way, and no sophisticated psychological analysis is needed to understand that pain (Brink *Forms* 402). Therefore, the efficiency argument, whether in its causal or epistemic form will be largely invalidated when circumstances suggest that the clear and present pain of a stranger can easily be alleviated. Such circumstances hold today, when billions find themselves subject to the clear and present pain of hunger and disease. However, in Sidgwick’s defense, the efficiency argument certainly does put a “thumb in the scales,” but, under current circumstances, goes only a very small way towards a total Consequentialist justification of our proclivity to nurture our special relations.

⁵ By resources, I mean not just wealth, but also time and effort, which, for the purposes of charitable donations, can be transformed into wealth.

Sidgwick's first form of (what I will call) the "agent reverberation argument" also falls well short of justifying our natural proclivities. This argument explains that the agent will derive extra pleasure from helping a special relation. Certainly there is a pleasure we derive from doing good for our most cherished relations. This is precisely what it means for them to be our cherished relations: we cherish their well-being and revile their pain as if it were our own. Indeed, we can imagine cases where another's well-being or pain matters more to us than our own direct well-being or pain. As such, there is greater overall well-being (impartially considered) when we do a service for our close relations as opposed to when we do the same service for a perfect stranger. Again, however, the need of the perfect stranger will often be far greater, and the reverberating effect of a service for a loved one will not outweigh the much greater good that could be done for a stranger. Furthermore, as Brink notes, a perfect stranger will have their own special associates, and, by the same token these associates also gain a reverberating boost to well-being from any service we do for their special relation (Brink *Forms* 400). However, when the agent himself enjoys this reverberating well-being, it might also be reinforced by the sense of pride involved in knowing that he was the cause of it.

The second form of the "agent reverberation argument" raises interesting considerations. As I have previously noted it comes under the heading of Sidgwick's style of "government house" utilitarianism, which holds that often utility would be maximized if agents do not attempt to maximize utility. In the case of special obligations, it might be true that we could do more good by shifting our benevolence towards the suffering stranger, but few have a strong motivation to do so. As such,

utilitarianism should not fight against the motivations of agents on this issue, but should take advantage of their predisposition towards acts of benevolence, albeit of a limited nature. Government-house utilitarianism relinquishes the typical utilitarian role as a decision-making procedure, and operates exclusively as a criterion of right action. This raises charges of elitism. In addition, “non-Consequentialists characteristically argue that there is something irrational or otherwise objectionable about a principle of right action which, from its own point of view, may not provide an acceptable source of motives or a permissible basis for decision-making” (Scheffler *Rejection* 45). I will avoid these questions for now, since I will contend that the rest of Sidgwick’s justifications, taken together, adequately lift the utilitarian concern for special relations to a point where it closely conforms to common sense.

Sidgwick’s third argument is that of “spillover” effects. It seems, at first glance, hardly controversial that someone who is on the receiving end of special acts of affection will become a more stable member of society, more readily prepared to be charitable towards others and perpetuate the process by bestowing acts of affection on their loved ones. In addition, Sidgwick also notes the role of gratitude in such relationships so that many acts of affectionate kindness will be reciprocated, adding once again to the overall pool of well-being (Sidgwick 433). Indeed some authors have observed in special relationships a sort of “tacit contract” that is essential for both members (Scheffler *Relationships* 190). One could argue against these spillover effects in various ways. Firstly, while it may be true that there are spillover effects involved in acts of affection for our close relations it remains to be

seen that they are not matched (or even exceeded) by the spillover effects involved in acts of charity for perfect strangers. Someone in a position of dire need might be tremendously motivated by the charity of a far off stranger, and this could lead them to do much more good for others than they otherwise would have. Additionally, while Sidgwick aptly observes the beneficial effect acts of affection for a loved one might have on someone's character, there are many counter examples of oversaturation of these kinds of acts, as in many cases of children being "spoiled rotten" by their parents. In the end, it seems the spillover argument has some validity, but could also work in favour of charitable acts towards strangers.

The last of Sidgwick's arguments is that of the increased benefit that a beneficiary receives from an act of goodwill, when this act comes from someone with whom they share a special relationship. This is much like the "reverberation argument," but focuses on the beneficiary of a kind act, as opposed to the acting agent. For brevity's sake I will call this phenomenon "the magnification effect for the beneficiary." As I have noted, the idea does not receive a lot of attention from Sidgwick. For the most part, he does not distinguish it from the reverberation effects for acting agents, or he lumps it in with the question of spillover effects. He does, however, acknowledge the phenomenon in a negative sense when he says, "we expect affection through tradition, and so disappointment is painful" (Sidgwick 439). Notice that the "magnification effect for the beneficiary" side of the question escapes Brink's criticism against "the reverberation effect for the agent." While those close to a beneficiary might take part in their pleasure, a far off stranger will

not realize any of the crucial magnification benefits of close affection that a loved one would.⁶

If we take Sidgwick's justifications together, stressing the "magnification effect for the beneficiary," we begin to see that AC need not advocate the neglect of all our special relations in the name of promoting well-being. Admittedly, an AC agent will be faced with scenarios where neglecting special relations would be the optimal course. However, if we consider carefully how to weigh the ill effects of neglecting these relations, such a decision will not be taken lightly. We are still left with difficult questions of when to neglect our relations but, as noted in the introduction, these questions occur to every sensitive moral agent. It is to the credit of AC that it explains the difficulties involved and provides us with an underlying rationale for why one course would be preferable to another (Brink *Impartiality* 164).

Throughout the twentieth century, Sidgwick's arguments were met with disapproval. As things stand now, it is generally considered that AC does advocate an inhuman, over-demanding neglect of our special relations. However, if we look at some of these theorists, often we can see an implicit understanding of the disutility involved in acts of neglect, and therefore a tacit acknowledgment of AC's disapproval of such acts. As a sample case, I will look at Brad Hooker's investigation of the special relations question in terms of generalization.

Hooker says that a single person neglecting their special relationships to aid with the intense suffering of far off strangers will, by AC accounts, increase the

⁶ It might be argued, however, that a person, alone and in dire straits, receives a tremendous boon from a stranger's helping hand, lifting them up and restoring their faith in humanity.

overall level of well-being in the world (Hooker 154). However, if *everyone* were to do this, the world would lack any sense of meaningful connection between people, leaving in us in an anodyne, faceless world community (albeit one with far less intense suffering). Hooker maintains that even AC would admit this sterile state of affairs would be worse than the state of affairs that holds today. As he says, “if the cost of everyone’s being motivated by an equal concern for all is the elimination of strong affections and deep attachments, greater good will result if everyone keeps their greater concern for themselves and their family and friends” (Hooker 154).⁷ If this is true, then AC is left in an incoherent position: if everyone were to follow an AC theory the consequences would be worse (on the very grounds of AC) than if everyone just ignored the theory and followed their own common moral sense.⁸ This is a common argument against AC and finds most traction in the question of promise keeping. (AC says I should break my promises if more good could be done thereby. If everyone did so, promises would lose all force, and the world would lose an invaluable institution, making the world worse off overall.)⁹

But arguably, Hooker’s argument loses much of its force when we consider the full significance of Sidgwick’s “thumb in the scales” arguments. The generalization argument depends upon the notion of non-linearity of the

⁷ Hooker does seem to neglect the tradeoff of a world where intolerable suffering is tremendously reduced, and it is not so obvious that AC would object to this generalized AC world.

⁸ Perhaps government house utilitarianism is a solution, but as I have already suggested this raises other issues. I look at further drawback to this form of the theory in Chapter 3, on expectation effects.

⁹ I look at these questions in full in Chapter 3, on expectation effects.

aggregation of acts (Lyons *Forms* 65).¹⁰ Accordingly, an act entails a certain disutility (i.e. the pain breaking my promise will have on the promisee), but when the same act is done “n” times it has a different disutility than the act multiplied by “n” (i.e. if promises are broken a million times in a society, the entire institution of promising will crumble, resulting in far more disutility than just the pain involved in breaking a single promise, multiplied by a million). Lyons maintains that aggregation of acts of a non-linear kind admit of a threshold effect (Lyons *Forms* 69). As in the case of promising, there will be a threshold point we pass where promises no longer carry any weight, and the disutility curve becomes non-linear.

However, should we consider generalization in the same way when looking at special relations? The threshold point in promise breaking is fairly easy to identify, namely the point at which trust in promises within a society begins to break down. In the case of special relations, it is hard to imagine what such a threshold point could entail. In the case of promises, the promise itself no longer holds any weight, but in the case of neglect for our special relations, it is doubtful that our instinctive yearning for special relations would cease to exist if neglect became sufficiently pervasive. Even if it did the disutility involved in the disappointment of expected affection (as noted by Mill, Sidgwick and Broad) would no longer play a part. If there is no distinctive threshold point then the aggregation of acts of neglect for our special relations should not be treated as non-linear. And it seems overwhelmingly plausible that the sterile society that Hooker dreads contains

¹⁰ I look more closely at Lyons’ concept in chapter 3, on expectation effects.

the exact same pain of neglect present in a single case of neglect, only directly multiplied (in a linear sense) by the number of members within a community.

In fact, the regrettable sterility of Hooker's imagined AC generalized society points to the very kind of disutility that so many critics fail to observe in a single case of neglect. If, as Hooker suggests, the AC agent should want to avoid this neglect when it pervades a society, then they should also want to avoid it in a single case, assuming the tradeoff is not overall beneficial. Ultimately, the tradeoff might be considered overall beneficial, but an incredible amount of suffering will have to be alleviated in order for acts of neglect to be justified.

The preceding arguments seek to show that a sensitive appraisal of AC values does lift the importance of close relations to a level consistent with our common sense. However, some critics see a problem with the way AC treats special relations, even if the well-being value associated with them can be "caught in the impersonal net of utility." Bernard Williams, for example, objects that AC does not respect the "seperateness of persons" when it puts stress on the impartial view (Williams *Luck* 3). I will look at this side of the question again in chapter 5, on the integrity objections.

Ultimately, we can see that much of the criticism of AC on the issue of special relations is misplaced. If, following Sidgwick, we take a sophisticated approach to assessing the relevant factors within the decision to fulfill our obligations to our loved ones then AC need not counter-intuitively advocate a cold and mechanical neglect. Indeed many of the criticisms against AC's approach to special relations, such as Hooker's generalization objection, imply an understanding of the severe loss

in well-being that occurs in a single case of neglect. If this loss in well-being is taken seriously, then we should not think that an AC agent would take lightly the choice to neglect a loved one in favour of helping others in more dire need. Of course, if more well-being results from helping far off strangers than from spending our energy in caring for our loved ones, then AC will advocate doing so. Given the desperate state of the world's worst off, this very well might be what a correct assessment of AC advocates. However this need not be seen as an AC weakness in callously disregarding our loved ones, but its strength in compassionately caring for the suffering of all humans. Faced with such a choice in the modern world, our common moral sense does not deliver any easy answers either, and in this way AC is in tune with our moral sense. Perhaps this is overdemanding, but this is another criticism leveled against AC that requires a separate chapter to fully investigate.

Chapter 2: Overdemandingness

Many critics argue that AC demands more of agents than any plausible moral theory should. This criticism, commonly known as the “overdemandingness objection,” usually comes in two forms. In the first sense, critics argue that AC is intuitively overdemanding, by requiring beneficent acts that initially appear to be supererogatory, above and beyond the call of duty. In the second sense, it is argued that AC demands so much of agents (even to the extent that one should neglect their “lesser” benevolent interests in favour of optimal ones) that this results in incoherence. In this latter regard, AC “fails by its own lights,” advocating acts that would ultimately diminish, not promote, overall well-being (Parfit *Reasons* 25). In this chapter I will evaluate the first form of the objection, while leaving the second form to my discussion of expectation effects in a following chapter.

Intuitively speaking, many see AC as requiring too much of moral agents. If AC deems the right action only that which optimizes overall well-being then, on any plausible reading, doing the right thing will entail tremendous sacrifice on the part of an AC agent. In order to act rightly, an agent would have to give away most of his possessions if that would help those in greater need. This would also require him to work tremendously hard to accumulate more resources and influence with which to help others. He could only rest or enjoy himself when this served the purpose of boosting his future ability to work to help others. He would not be permitted to choose who receives the benefits of his labours, but must give to those in direst need, even if this means overlooking the suffering of his closest friend (Hooker 150).

A life lived by the standards of AC would likely be one of “hardship, self-denial and austerity” (Kagan 360).

To begin, I will look at Mill’s arguments for why utilitarianism does not demand so much of agents, except in exceptional scenarios. I attempt to show that his reasons do not hold as much weight under modern conditions as they would have in his time.

I then consider another possible response to the overdemandingness objection, which is that we should just bite the bullet and accept that a plausible moral theory will be extremely demanding. In this respect, I look at Shelly Kagan’s arguments against the validity of our intuitions in observing overdemandingness on the part of AC.

Continuing with biting the bullet defenses, I look at David Sobel’s arguments for seeing non-AC theories as overdemanding on the object of aid, since they lose out on the aid that AC says they deserve. This argument relies on a breaking down of the distinction between causing and allowing. I then consider Matthew Tedesco’s reasons for why we should not break down this distinction, at least for Sobel’s purposes.

Next I look at another possible response to the objection, based on a remodeling of AC from an optimific to a satisficing form. I analyze Michael Slote’s thought experiments in support of this, and then Phillip Pettit’s arguments for the irrationality of this concept. I also raise added concerns about Satisficing Consequentialism’s inability to account for the praiseworthiness of extreme moral sacrifice.

In lieu of these objections I support another remodeling of AC, also supported by Slote, which is known as Scalar Consequentialism. This form of the theory is not as rigid and makes only judgments of better or worse between different acts. I argue for this form over Optimific Consequentialism for reasons of Optimific Consequentialism's counterintuitive grouping of suboptimal moral acts.

Finally I look at challenges to the scalar form. I attempt to argue against the idea that this form relinquishes the objectivity of AC, as well as its role as a decision-making procedure, and, lastly, that it is incoherent.

The early utilitarians understood the importance of the overdemandingness objection. John Stuart Mill defended the theory against overdemandingness by employing a sensitive interpretation of how to promote the well-being of everyone. We should not just throw money at a perfect stranger because we have a deeper understanding of how to fulfill our own desires and the desires of those whom we care deeply about. Furthermore we are often ideally situated (in terms of proximity, relational status, and familiarity) to help those closest to us and also ourselves. Because these advantageous conditions may not hold when it comes to helping an intensely suffering far off stranger, we may be able to provide greater benefit to ourselves and those closest to us at a lesser cost. This leads Mill to conclude that, "the occasions on which any person (except one in a thousand) has it in his power to do this on an extended scale, in other words to be a public benefactor, are but exceptional; and on these occasions alone is he called on to consider public utility; in every other case, private utility, the interest or happiness of some few persons, is all he has to attend to" (Mill *Utilitarianism* ch. 2).

In this defense Mill correctly identifies a common misinterpretation of utilitarianism. If we consider carefully the effect our position in relation to others, then efficiency will dictate a concern for our loved ones and ourselves.¹¹

However, much of Mill's defense crumbles when we take into account present day conditions of first world affluence and massive wealth disparity (Hooker 153). The assertion that few will have to take time away from their direct interests seems more plausible in Mill's day, when a large segment of the population endured very real hardships related to poverty, hunger and disease. Today, extreme poverty is relatively rare in many first world countries. It seems plausible that we could more efficiently add to well-being by alleviating the hunger of a far off stranger than by attending to some other needs of our own or those close to us. Furthermore, in the modern world, we are provided with accurate knowledge of far off disasters, and can donate more efficiently to far off places (Hooker 153).

AC allows one to decide what is the right action, regardless of what others are doing. In the case of donating to those suffering in far off countries, AC's demands on an affluent agent might be lessened if other affluent agents cooperated in aiding the poor. Since that is not the case, AC will remain very demanding for an individual agent.

A possible obvious response to the demandingness objection is that morality should be very demanding. The plausibility of AC is only fortified by considerations of its demandingness. After all, it should come as little surprise that a correct moral theory asks us to alleviate the intense suffering of strangers, even if this asks us to

¹¹ This argument is pursued in greater detail in the previous chapter on Special Relations.

(in some way) take on their suffering as our own. If one is better off than others should they not be asked to rectify this imbalance? Is it not plainly selfish for one to insist otherwise? Indeed, any plausible moral theory, not just AC, should demand very much of an individual agent (Mulgan 9).

It seems, however, that while most people would find such behavior admirable and good, they would not find it morally obligatory. Indeed we can perceive this in the fact that so few (if any) choose to pursue such a course of complete self-sacrifice. We feel the need to act in a morally acceptable way, but (judging by our actions) we draw the line of acceptability well short of complete devotion to the common good.

Two major factors involved in this opinion seem to be the proximity of the suffering stranger and the selfishness of our special concerns. Shelley Kagan provided the example, introduced by Peter Singer, of deciding whether to (easily) save a nearby drowning child or to avoid exerting the physical energy required (Kagan 3). Clearly moral intuition tells us that we must save the drowning child. Why then do we not feel we must save a far off suffering stranger?¹² It seems proximity plays an important part in the distinction. Perhaps as evidence of Hume's conception of sympathy, we feel more morally bound to alleviate the suffering of someone with whom we are in close contact.

A second factor in the intuitive permissibility of a sub-optimal moral action is the degree of selfishness involved in the special concern. If someone asks us why we

¹² It might be argued, however, that the inefficiencies of international aid make it highly unlikely that saving far off strangers is so easy, or indeed possible. I leave this question open, but take it as a given that, if we cannot save lives through international aid, we can at least greatly improve their quality of life.

do not donate a hundred dollars to famine relief, our intuition tells us that an example of a morally acceptable reply would be that we are the guardian of a small child and need to spend that money to insure they eat healthily. An example of a less morally acceptable reply would be that we would prefer to spend the money in treating our self to a luxurious dinner. Indeed, many might follow such a course of action, but would probably feel ashamed to openly admit their reasoning.

In these ways, our everyday intuitions distinguish between morally permissible and impermissible decisions. Our intuitions draw the line of acceptability well short of the criterion of right action offered by optimizing AC (Kagan 2).

Shelly Kagan responds to this form of the objection simply by maintaining that our intuitions about morality are not wholly reliable. He objects to the validity of “reflexive equilibrium,” which states that a moral theory is heavily informed by our moral intuitions, and vice versa (Rawls *Theory* 20). Such a model of moral theorizing treats our intuitions as basic “data,” from which we attempt to formulate cohesive moral principles (Kagan 11). However, as Kagan notes, “it is doubtful that we have any pre-theoretical intuitions. (...) It would be more correct to say that we begin moral philosophy already possessing some moral theory” (Kagan 12). If this is so, then intuitions cannot serve as data independent of the theory.

Kagan prefers to evaluate moral theories based on whether or not they are independently rational. He rejects an intuition-based approach that simply yields an axiomatic set of maxims. A good moral theory, rather, should have power, simplicity, consistency and explanatory force (Kagan 13).

Explanatory force he takes to be the most crucial element of a moral theory. Without it, an intuition-based approach will contain “dangling distinctions.” He provides the example of an antebellum American slaveholder. Such a person might hold on to the belief that it is wrong to coerce a person into servitude if they are white, but permissible to do so when that person is black. However, the distinction between the two cases is left unexplained, or “dangling” (Kagan 12). If the slaveholder had to attempt an explanation, they would have to revise their stance or offer a justification for this (seemingly arbitrary) distinction. Following Wittgenstein, Kagan understands that explanations must come to an end somewhere, “but not at an arbitrary level” (Kagan 14).

The difficulty involved in an intuition based methodology leads Kagan to call for an “extreme” approach to morality that takes a theory’s internal rationality as paramount (Kagan 390). This contrasts with the a “moderate” approach to morality that trusts our intuitive belief that we have morally acceptable “options” to fall short of the theoretical goal, when doing so is, say, overdemanding. In the context of AC then, Kagan would not see the intuitive disapproval of overdemandingness as a valid objection.

A full critical examination of Kagan’s downgrading of intuitions would throw our present argument far off course. However, in support of Kagan, we can tentatively demonstrate some important cases of irrational intuitions. As already noted, commonly held irrational intuitions held force in (among many other eras) the antebellum American south. The arbitrary distinctions involved in those intuitions cannot be criticized unless we do so from the perspective of a cohesive

moral theory, or we simply dismiss those intuitions as unworthy. Yet on what authority can we take the latter course?

David Sobel takes up Kagan's arguments to form a counter-objection to the charge of overdemandingness. He maintains that the objection begs the question against AC. It does so by pre-supposing the falsity of the AC lack of distinction between what he calls "causing and allowing" (Sobel 3).

Sobel explains by way of an example. Imagine a situation arises where a person, Joe, can donate a kidney for a transplant in order to save another person, Sally, from imminent death (Sobel 4). Losing a kidney might cost Joe severe medical difficulties in the future, but would not likely cause his imminent death. Receiving a kidney would save Sally from imminent death. By an impartial Consequentialist calculation, Joe should donate the kidney. Intuitively we would find this Ned Flanders-type of action (giving away a kidney to a perfect stranger on a "first come first serve" basis) ridiculous.

However, Sobel explains that this opinion is based on the intuitive distinction between causing and allowing. We see Joe as allowing Sally's death to occur, but not as causing that death. If we eliminate this distinction then, Sobel asks, why not see our common anti-Consequentialist morality (which says one may refuse to donate the kidney) as overdemanding on Sally? As he puts it, the demandingness objection is "already in the grip of the thought that a moral theory that requires x to sacrifice for y is demanding on x but a moral view that permits y to suffer rather than insist that x help is not similarly demanding on y" (Sobel 6).

In shifting the focus from the costs imposed on the “aider” to one who would lose out on aid, Sobel provides a clever counter to the standard overdemandingness objection. He finds the objection impotent – it adds nothing to the debate but hinges on whether or not we accept Consequentialism in the first place. It remains to be seen to what degree AC is committed to the lack of distinction between causing and allowing, but on standard readings it can be plausibly taken as breaking down this distinction (Williams *Critique* 99). I will then take it as given that the causing/allowing distinction on which the demandingness objection lies does indeed presuppose an anti-Consequentialist way of thinking.

Matthew Tedesco attempts to counter Sobel’s objection by claiming that Sobel, in turn, begs the question against intuitions (Tedesco 99). Sobel presupposes that our intuitions, which observe a relevant distinction between causing and allowing, are not independently reliable. He follows Kagan in claiming that our intuitions stand in need of a rational defense.

We might then refer back to our brief discussion of Kagan in order to dispel any fears of question begging on Sobel’s part. The belief that the intuitive causing/allowing distinction is unfounded gains traction if we have good reason to believe that many of our intuitions are unfounded. This was shown in considering “dangling distinctions” and arbitrary intuitions. Tedesco objects, however, that the causing/allowing distinction is “a distinction whose intuitive force is independently credible and so immune from the kind of scrutiny that Kagan recommends in his proviso against dangling distinctions” (Tedesco 100). So Sobel begs the question by

assuming “that a methodology that finds intuitions to be independently credible is mistaken.” (Tedesco 100).

A complete discussion of the internal rationality of the causing/allowing distinction might take us too far off course here. I only wish to show the difficulty in assigning infallibility to some of our moral intuitions, and the implications this has on the overdemanding objection.

Another way of treating the overdemandingness objection is by accepting the claim that our intuitions rightly deem an overdemanding moral theory implausible, but that AC need not be taken as “demanding” so much of an agent.

This can be done by treating the standard view of AC as only one version of the theory, namely the “optimific” version. In contrast to the optimific version there may be other, more plausible forms of the theory.

Slote advocates a form called “Satisficing Consequentialism.” This form of Consequentialism treats the criterion of right action as drawing a line between “good enough” and everything else that falls short (Pettit 149). Where that line is drawn will probably come down to societal standards (Slote 78).

Slote believes AC can accommodate such a transformation because “the idea that the rightness of an act depends solely on its consequences, i.e., on how (impersonally) good its consequences are, is separable from the idea that the rightness of an act depends on its having the best consequences (producible in the circumstances)” (Slote 36). Slote sees AC as essentially related to the idea of promoting good consequences, and this need not mean that optimizing Consequentialism is the only possible variant of the theory.

Slote uses examples to demonstrate what is at play in a Satisficing Consequentialist theory. One example is that of the lunchtime snack. A person who has just finished their lunch might be offered a snack, and although they are still somewhat hungry and think they would enjoy the snack, they refuse anyways on the grounds that they are “fine as they are” (Slote 39). Another, example, this time in the moral realm, is that of the innkeeper (Slote 41). An innkeeper encounters a family stranded by a car wreck and decides to put them up for the night. The innkeeper puts them in the first room she sees, when there were better rooms available. Such an agent would not be blamed, because they would have brought about a sufficient level of well-being under the circumstances.

Philip Pettit takes issue with this form of Consequentialism. He objects that Satisficing Consequentialism does not explain our intuitions. He finds the behavior of the agent in the lunchtime snack example to be irrational. If the agent wants more food then he should take the snack. If he does not it is either because he is not really hungry or there are some competing interests, such as shyness, that prevent him from taking the snack (Pettit 169). Moving to the innkeeper example, the behavior is again irrational if the agent is attempting to promote well-being. Either the agent is not aware of the better rooms, has some competing interests or is not concerned with promoting well-being.

Any rational reason for such agents to fall short of the optimal decision must be motivated by some Consequentialist concern (as in there must be more pleasure to be derived from not following the seemingly optimal course, or taking an optimizing stance takes time and energy away from other Consequentialist

pursuits) or it is motivated by some other, non-Consequentialist concern (such as pure selfishness or a regard for the autonomy of others) (Pettit 169). Slote hints at the (Consequentialist) benefits involved in the computational ease of Satisficing Consequentialism, but as Pettit notes, this threatens a regress back to an optimizing form (Pettit 170).

Furthermore, Satisficing Consequentialism seems unable to explain the moral pull of giving up one's time to a far off deserving charity. While it is seen as overly demanding to require such extreme devotion to charity, we do consider it morally laudable. Satisficing Consequentialism seems unable to accommodate this common intuition. In this way Satisficing Consequentialism appears to throw the baby out with the bath water.

Slote has a simple response to these objections. In *Common Sense Morality and Consequentialism*, he outlines a form of Scalar Consequentialism "that makes only comparative judgments of better or worse" between actions (Slote 77). This scalar criterion of rightness judges every act as better or worse. He sees it as a combination of the optimific and satisficing forms of Consequentialism (Slote 77). Like optimific AC, it deems the best action that which maximizes well-being; and, like satisficing AC, it allows for acts which fall short of total maximization to still be considered right, on the whole. It is an interesting move on Slote's part, one that shifts the emphasis of the basic Consequentialist criterion, but does so in a way that does not immediately appear to deprive Consequentialism of any of its force or coherence.

Indeed it is worth considering how deep a departure his move represents from the “standard” form of Consequentialism. Some authors describe Consequentialism as “demanding” the optimal decision from us while others see it as only “asking” for this optimal decision. Tedesco, for example, takes for granted that the standard form of Consequentialism is the optimizing form (Tedesco 97).

Mill seems to advocate something along the lines of a scalar form. In keeping with Satisficing Consequentialism he finds that “an act is morally wrong only when it fails to optimize utility *and* its agent is liable for punishment for the failure” (Sinott-Armstrong para. 66).

Clearly Mill understood the importance of the overdemanding objection when he said, “it is exacting too much to require that people shall always act from the inducement of promoting the general interests of society” (Mill *Utilitarianism* ch. 2). We have already seen how he weathers these fears through a sophisticated treatment of how to efficiently add to overall utility.¹³

Bentham can also be seen as supporting a scalar form. He puts forward the principle of utility as that which “approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question” (Bentham *Principles* ch.1). D. H. Hodgson believes this phrasing is perfectly consistent with a scalar criterion of right action (Hodgson 10). Indeed Slote’s analysis of revisions of Bentham’s text shows that the original contained notes that softened any optimistic

¹³ Page 20-21.

interpretation. In addition terms like “best” and “greatest good” did not appear in the original (Pettit 153).

It seems then that the scalar form of Consequentialism is not such a severe mutation of the original form of the theory intended by early utilitarians. Indeed, such a scalar criterion seems fitting for a Consequentialist theory, with its quantitative evaluations, as opposed to more binary Deontological theories.

In an important way the scalar form avoids absurdities contained within the optimific form, beyond the unpalatable and counter-intuitive demands it imposes on an agent.

Optimific Consequentialism unreasonably groups together all deviations from the optimal standard. Imagine three possible courses of action. One could give to an efficient charity that provides clean drinking water to those in need of it. This action would bring about the greatest good possible. Alternatively, one could give to a slightly less efficient charity that provided help with disease control in poor areas. As a third option, one could go on a murder spree. Optimizing Consequentialism maintains that the first choice is the right one, but what about the other two? Are both simply wrong and therefore we should be indifferent between the two? Clearly giving to the world’s second best charity and going on a murder spree are of a different moral quality. However, Optimific Consequentialism, in its binary rigidity, treats the two the same, lumping them together equally as wrong actions.

The scalar form of the theory has the flexibility to treat these last two acts differently. It can say that the murder spree was a morally reprehensible decision but it can also say that giving to the second best charity was a very worthy decision,

though still strictly worse than giving to the best charity. It can say that giving to the hundredth best charity is worse than giving to the second best. It can also say that doing nothing is worse than giving to the hundredth best charity but much better than going on a murder spree.

Where is the line to be drawn between a right and wrong action? If a line has to be drawn, Slote thinks it should be done along the lines of societal opinion. He compares this with baldness. As we pluck hairs away from a full head of hair that head becomes more and more bald, but where we decide that that head is now bald will have to be determined by popular opinion (Slote 78). In essence such a move does away with a binary criterion of right action and replaces it with a binary criterion of praise- and blame-worthiness.

Scalar Consequentialism appears to answer Pettit's objections. Because an agent is no longer maximizing, any charges of irrationality in this regard are thrown out. Also we can now explain the intuition that devoting one's self to charity is admirable but supererogatory and not required.

The scalar form of Consequentialism solves problems of counter-intuitive overdemandingness and overly rigid grouping of wrong doers. It does so without, at first blush, curbing AC of any of its force or validity. The supremacy of well-being promotion in consequences is maintained, while doing away with counter-intuitive side effects.

However it may curb the force of AC in two ways, on grounds of objectivism and its validity as a decision-making procedure.

If, as Slote suggested, we draw the line of rightness and wrongness on societal lines then it appears that AC loses force as an objective principle. The principle of well-being promotion is said to hold universally under all circumstances. If we take a societal standard of rightness and wrongness then does this not subordinate this principle to the subjective whims of a given society? What about the aforementioned antebellum South? Are we simply to say that this society drew the line of rightness at a very low level of well-being (or high level of suffering)?¹⁴

For these reasons, it seems wrong to derive the criterion of rightness from societal acceptance. We might then treat the criterion as determined simply by the reasonable expectation of a sensitive moral agent.

This leads into another objection against Scalar Consequentialism. Typically the AC criterion of rightness serves to guide our actions (although some forms, such as government-house utilitarianism, separate these two aspects). When someone wants to know what to do in a given moral scenario, Optimific AC can point to the optimal action and say “that and that alone is what you should do.” Scalar Consequentialism, on the other hand, seems unable to point to a single course of action in this way (Slote 86).

The force of this objection depends on how we take the agent’s question. If we take her to be asking, “what is the best course of action?” then Scalar Consequentialism provides a definite response, the same answer provided by

¹⁴ In this case it is more probable that slaveholders had no commitment to the impartiality of Consequentialism (or any other moral theory) or that they arbitrarily distinguished between whites and blacks

Optimific Consequentialism. But if we take her to be asking, “what is the acceptable course of action?” then Scalar Consequentialism does not provide the same definite response as Optimific Consequentialism. However, as we have already seen, a definite response entails a counter-intuitive grouping of all suboptimal behavior, and, of course, invites charges of implausible overdemandingness. The advantage of a definite response to questions of acceptability is not worth these drawbacks.

Furthermore, one wonders whether a definite line of acceptability should be considered an advantage at all. Whereas Optimific Consequentialism provides an agent with two options – one right and one wrong, Scalar Consequentialism provides an agent with several possible courses of varying degrees of rightness or wrongness. It offers the agent freedom to decide for themselves what is right, but also provides them with a definite guide for what makes an act right, namely well-being promotion.

One could also object to Scalar Consequentialism that it paints an incoherent picture. One may ask, what motivates any deviation from the optimal standard action? A rigid Consequentialism will not allow for competing considerations of rightness (i.e. a sense of duty to one’s family, or respect for another’s autonomy, etc.) to motivate this deviation. Such a Consequentialist must maintain that one deviates from the standard purely out of wicked anti-Consequentialist considerations. Therefore a Scalar Consequentialist must count wicked considerations as good when the resulting act surpasses the line of acceptability. It seems, at first blush, incoherent to say that an action can be both right and wrong in this way.

This objection begs the question against Scalar Consequentialism. It already assumes that all actions must be either wholly right or wholly wrong, without varying degrees. Although Scalar Consequentialism may allow for *some* (wicked or selfish) anti-Consequentialist concerns in a right action, it will remain strictly opposed to such concerns, maintaining that any wrong action will occur when one is *overly* motivated by such selfish concerns.

It seems the confusion arises out of the terms “right,” and “wrong.” How could an act be more or less “right?” This seems akin to saying a person is “a little pregnant.” It might then be more apt to use the terms “good” and “bad.” Surely an act can be more or less good, or more or less bad, as evidenced by the two charities example.

The overdemandingness objection raises important considerations for AC. Although overdemandingness will lead to issues for any plausible moral theory, it is especially pertinent for AC, with its stress on perfect impartiality. Attempts to justify special concerns via reflections on the efficiency of aid cannot fully dispel the idea that the optimal ACist course will entail tremendous self-sacrifice. The inadequacy of this defense is only amplified under modern conditions of severe wealth disparity and third world suffering. Another possible response to the objection, following authors like Kagan and Singer, is that morality should be very demanding and AC does well to capture this insight. Our intuitions might tell us that AC demands too much, but our intuitions might not be independently reliable and stand in need of a rational defense. On the other hand, if one wishes to preserve the importance of our moral intuitions then AC can respond to the overdemanding objection by supporting

a scalar criterion of rightness, instead of the commonly understood optimizing one. This also eliminates the unreasonable grouping of “wrong” actions entailed in an optimizing model. Through these considerations it can be seen that overdemandingness will be no more of a problem for AC than for any other plausible moral theory.

Chapter 3: Expectation Effects

Objectors to AC claim that when agents attempt to act according to the dictates of AC (that is, when they attempt to bring about the most well-being) they do no such thing, and indeed reduce the level of well-being that would have obtained had the agents just followed their common sense morality (Hooker 142). In this way, AC is said to be self-defeating, and to “fail in its own terms” (Parfit *Reasons* 28). In this chapter I will consider to what degree this is true of AC, and if it is indeed true, to what degree this delegitimizes AC as an accurate moral theory.

The problem of self-defeat rears its ugly head under circumstances of universalization of the AC disposition throughout a given society. Under such conditions, the “mercenary”¹⁵ actions of an individual ACist agent are said to be blind to the harmful expectation effects that occur as a result of large scale disregard for highly revered human institutions.

Critics have suggested this possibility using examples of promise-keeping, truth telling, and increased concern for loved ones (Hooker 156). In the case of promise-keeping, it is said that if everyone follows AC then everyone will respect promises only as much as their “utility,” in an individual case, is worth. Therefore, whenever more well-being could be brought about by breaking a promise, any agent whose goal is to promote well-being will be morally encouraged to break that promise. If this disposition persists within a society, promises will lose their binding force and the society will lose out on a valuable system necessary for social order.

¹⁵ I use this term to describe the AC agent’s ability to work alone, regardless of what others are doing.

This may occur because enough AC agents will break promises to the point where promises lose all legitimacy, or it may occur simply because all agents hold such a disposition and will implicitly understand the lack of respect their fellow AC agents have for the rule of promise-keeping. Either interpretation has important implications. It is simply worth noting for now how several agents attempting to follow the AC principle might actually bring about a result that AC itself would reject. Generally, the same process would occur with truth telling as with promise-keeping. In either case the knowledge that others act according to AC will result in harmful expectation effects.

In addition, authors have suggested a way in which a universalized AC disposition might result in worse consequences in the case of increased concern for loved ones. An AC agent will have to spend their beneficent efforts wherever it will have the greatest impact. As a result it is said that AC agents will have to do good for far off strangers as opposed to the loved ones they are most naturally disposed to aid.¹⁶ When this behaviour is universalized, we are left with a soulless, anodyne society devoid of meaningful relationships between loved ones.

The collectively self-defeating nature of individually rational acts is a commonly cited problem, not just in moral domains. Even a purely self-interested rationale leads to self-defeat when coordination must be countenanced, as demonstrated by the well-known prisoner's dilemma. Parfit finds that AC is only collectively self-defeating, not individually (Parfit *Prudence* 555).

¹⁶ In chapter 1, p. 9-16, I argue that AC might not lead to this conclusion

In cases of voting or paying taxes, it has been demonstrated how individually rational acts become irrational under conditions of collectivization. As examples, a single person might not receive any tangible benefits from voting in a large democracy, but if everyone follows this course then no one will vote, resulting in disaster. Similarly a rational self-interested person would not pay taxes (if this would not result in punishment) but if no one pays taxes, again the result is disastrous (Lyons *Rights* 45).

Such problems of “collectivization” have in part motivated the shift from AC to Rule Consequentialism (RC). Under RC, certain rule violations, such as breaking a promise or telling a lie, are prohibited as a way to solve problems of coordination. Interestingly, RC is criticized for running into an opposing dilemma. If AC cannot serve as a functioning moral theory under conditions of universal assent, then we might argue that RC cannot function for an individual moral agent when others do not follow the same moral code. RC is said to suffer from problems of utopianism, since it only delivers the proper rules to follow when those rules are followed by all or most of society (Arneson 236). The severity of either the utopian or coordination problems depends on how we treat a moral theory, either as instructing an individual “mercenary” moral agent how to act under whatever conditions hold in the wider society, or as instructing us how an entire society should act.

Regardless of how the theory is seen, AC writers have attempted to deflect charges that, for certain time honoured institutions, their theory is collectively self-defeating.

In this chapter, I will begin by analyzing David Lyons' concept of non-linearity in moral acts, which is meant to explain harmful expectation effects. Although I find some problems with this concept I think it provides a useful explanation of this process, but I argue for shifting the focus towards the "invisibility" of these effects in each act.

I next look at examples of when expectation effects are said to pose a threat, and argue for promise-keeping as a more plausible scenario, than, say, neglect of special relations.

I then argue for the gradual nature of the harmful effects of breaking promises, which would suggest that non-linearity, or case-by-case "invisibility" does not apply. Still I find certain collateral effects involved in promise breaking do lead to legitimate fears over expectation effects.

Next I look at Sidgwick's arguments for raising the utilitarian concern for promise breaking, impartially considered, but, in following reasons put forward by W.D. Ross, I find that Sidgwick's arguments do not convince.

I then look at Lyons' solution to the problem, which is to maintain that AC agents will be able to anticipate the danger to an overall institution, and can therefore avoid reaching the threshold point where this institution starts to crumble. In this context, I consider Mill's arguments for the utilitarian value of upholding valuable institutions. In the end I think a serious consideration of Expected Consequentialism, which judges acts according to the possibilities of its effects, does provide a suitable counter to the expectation effects problem. However, since this approach might be very difficult for every agent, I argue for the

incorporation of rules of thumb into AC. I also consider Hodgson's objection to this type of solution, which maintains that, in a universalized AC society we cannot trust others, even before a single promise is broken. Still I maintain that Expected Consequentialism fortifies against Hodgson's objection.

Following this I consider another possible solution to the expectation effects problem, namely a Sidgwickian style government-house utilitarianism. In relation to this, I look at R.M. Hare's version, which avoids calling for elitist division within a society.

Finally, I argue for the inadequacy of this solution based on the problems of self-suppression within a moral theory, which will lead to counter-intuitive judgments.

In order to help explain how harmful expectation effects occur, David Lyons introduced the idea of causal non-linearity (Lyons *Forms* 69). An act is said to be causally non-linear if the total effect of many of those acts results in a greater or lesser effect than just the sum of each of those acts, treated independently. For acts of this non-linear type, there is a threshold where another effect comes into play, for example, the point where expectation effects start to play a decisive role (Lyons *Forms* 73).

It may not be the case that all expectation effects follow such a model of non-linearity. It seems more likely that, from the start, breaking a promise, for example, will have harmful effects on other people's trust in the reliability of promises. As such, a threshold point may not arise in the way Lyons suggests. However, even if expectation effects play a role from the start, it would still plausibly be the case that

as more and more (sticking with the example) promises are broken, this will negatively affect a community's trust in promises to a greater and greater degree.

While Lyons' model may not be strictly accurate it does elucidate why self-defeating expectation effects pose such a problem for AC. Even if expectation effects play a role from the start, it still could be the case that they are so miniscule in the early phases of universalized AC behaviour that they are practically *invisible* for a normal AC agent. The *invisibility* of these subtle effects raises concern over AC's ability to avoid self-defeat under conditions of universalized adherence.

The invisibility factor that non-linearity seeks to explain is said to arise because of various violations against common sense morality.

As I have already mentioned, the case of neglect of special relations is said to lead to such self-defeating universalized effects. As Hooker says, "the world might well be a poorer place if *all people* lost their strong affections. But only I and the comparatively tiny circle of people with whom I am connected would lose if *only my* strong affections were eliminated" (Hooker 154). However it is unclear how non-linearity (or any other process of case-by-case "invisibility") could apply in this case. It seems that the disutility of an entire society neglecting their loved ones is simply a summing up of the disutility in individual cases of neglect. In order for the universalization of this conduct to pose a distinct problem it must be demonstrated how an entire society neglecting their loved ones would affect a member of that society differently than just his own special relations neglecting him. Indeed, the latter case might represent a greater loss in well-being, as the subject suffers not

only from a loss of familial love, but is made more acutely aware of this loss through the familial love that others enjoy.

Furthermore, it appears more likely that the cumulative effect of many cases of neglect would be non-linear in the reverse sense. As more and more agents devote time to those who are worst off as opposed to their loved ones, the situation of the worst off will, as a direct result, become better and better. As this process continues, the needs of the worst off will become less and less dire and the call for agents to neglect their special concerns will become less and less pressing. Once total universalization holds, it seems plausible that the increased well-being one receives from acts of kindness from a loved one (as Sidgwick argued for, and I do in the chapter on personal relations) would play a decisive role in determining the target of an agent's beneficence. In this respect, the dilemma of special relationships is best demonstrated in cases of an AC agent acting alone, and so Hooker's characterization of the problem as an example of universalized expectation effects appears misplaced.

In the case of promise keeping and telling the truth, Lyons' non-linearity (and case-by-case invisibility) presents itself far more obviously.¹⁷ As more and more agents break a promise in the name of increased well-being, more and more people will expect promises to be broken and wide spread reliance on the "bindingness" of promises will be lost. This distrust would arise in the betrayed promisee, and to a lesser extent, in those who are witness to any broken promises. This negative effect

¹⁷ Much of the what follows applies to both promise-keeping and telling the truth so I will simply describe the case of promise-keeping.

would be negligible in the case of a single broken promise but would become exponentially more significant as more promises are broken.

How different is this from the case of close personal bonds? In that case we explained that one instance contains proportionally the same amount of disutility as many cases would, or at least that there was no good reason to think otherwise. Whereas, in the case of promise-keeping it seems there is a higher chance of cumulative effects. In breaking a promise we affect the promisee's future trust in promises. But a cumulative effect might occur in the case of neglecting our personal relations as well, namely in terms of that person's ability to form and nurture future close personal bonds, which are also based on trust.

Perhaps then, from the start, each additional broken promise contributes to the downfall of the institution at a gradual, uniform rate. This would contradict the view that AC's treatment of promises leads to self-defeat, whereby the institution is said to abruptly "fall of a cliff," without any warning for diligent AC agents. If, on the hand, a gradual degradation of the institution occurs, then the effects cannot be said to be "invisible" for an AC agent, but are in plain sight.

If an AC agent breaks a promise, even once, then the betrayed promisee will lose some faith in *that* person's promises. If several promises are broken within an AC society, it may simply be the case that the overall degradation of promises is a direct result of this process playing out several times. This effect is plainly visible from act-to-act and proceeds at a uniform rate, like the case of personal relations.

However, we can also assume that under such conditions promises in general will not be trusted, regardless of the promisor's reputation. This is different from an

AC agent degrading the force of his own promises via repeated betrayals. It is this collateral effect, (comparatively) less “foreseeable,” that leads to the objection of self-defeating expectation effects.

Sidgwick suggested a possible response to this problem (Sidgwick 442). He attempted to show the disutility of breaking a promise in a single case. If this disutility were great enough then the breaking of promises by an AC agent in the name of greater utility would be so rare that the harmful expectation effects would, plausibly speaking, never have a chance to present themselves in a fully ACist society. He shows that breaking a promise is an act of infidelity, a betrayal. He uses the example of someone who says they will stop drinking. If this were simply a statement, then we would not accuse him of lying if it turns out to not be true. However, if we take it as a vow, then his not doing so would be taken as a kind of lie (Sidgwick 145). So inasmuch as we are betrayed by his breaking of a promise, there is a loss of well-being on the promisee’s part. A true AC agent will “hold any disappointment of expectation to be *pro tanto* an evil, but a greater evil in proportion to the previous security of the expectant individual, from the greater shock thus given to his reliance on the conduct of his fellow-men generally: and many times greater in proportion as the expectation is generally recognized as normal and reasonable, as in this case the shock extends to all who are in any way cognizant of his disappointment” (Sidgwick 443). This sense of betrayal and disappointment illustrates a loss of well-being in promise breaking beyond the mere inconvenience the promisee may endure as a result. An AC agent will still break promises, but perhaps only when any other good person, following their intuitions,

would do the same. Therefore, the force of promises is no more endangered in a universalized AC society than in our present day society.

However, despite the importance an AC agent would place on promises (in this way), some object that the problem persists because the value placed on promises is purely instrumental. W.D. Ross shows that an AC agent will break a promise whenever more well-being (even the least degree) could be gained thereby (Ross *Right* 34). As such, Sidgwick's explanation does not do enough to avoid fears of harmful expectation effects in the case of promises, and they cannot be relied upon in a universalized AC society

Ultimately, AC might have to concede that, despite Sidgwick's defense, promises will be broken more easily than would be allowed under common sense morality. Still this need not mean that AC admits its own incoherence in light of harmful expectation effects.

Lyons proposes an alternative to Sidgwick's solution of giving a strong utilitarian weight to individual promises. He suggests that an AC agent can use their "mercenary" moral status as a way to guard against more and more severe expectation effects. Lyons insists our acts are not considered "in vacuo" but that circumstances and the behaviour of others are taken into account (Lyons *Forms* 73). To a degree, we can ascertain when the threshold point of the expectation effect will be reached, and decide the right course of action based on that knowledge. So, as in the case of promising, an AC agent will take stock of all the positive and negative effects of breaking a promise. As such, an AC agent will have every reason to take into account the negative expectation effects associated with more and more broken

promises. As noted, a single violation might carry with it little to no negative expectation effects and so an AC agent will not consider this effect in their moral analysis. However, after enough violations, an AC agent will be fully sensitive to the danger posed to the overall force of promises and as a result will assign greater and greater importance to the keeping of one's promises. Such sensitivity to the prevailing circumstances would be required of an AC agent as it would be of any worthy moral agent. As Lyons puts it, "such difficulties are no different in principle from those one faces in reckoning any causal circumstance of an act, especially circumstances involving others' behaviour" (Lyons *Forms* 82). John Rawls also understood this prospect. He said that someone who has made a promise will weigh "not only the effects of breaking his promise on the particular case, but also the effect which his breaking his promise will have on the practice itself" (Rawls *Theory* 232).

As we saw earlier, when an AC agent breaks a promise, his promises will obviously carry less weight for the one betrayed. This consequence is so predictable that it does not admit of the "invisibility" required to explain expectation effects. The collateral effects on promises as a whole are comparatively harder to recognize, but should not be considered wholly invisible. Inasmuch as we are able to voice a concern over these effects in the first place, a subtle and sensitive AC agent should be able to take them into account when acting.

In "On Liberty," Mill discusses the importance of preserving valuable institutions in this way, in order to promote overall utility. When a person makes a promise "a new series of moral obligations arises on his part towards that person,

which may possibly be overruled, but cannot be ignored (...) and if he does not allow proper weight to those interests, he is morally responsible for the wrong” (Mill *Liberty* ch. 5). Part of these “moral considerations” can entail the chance that breaking a promise will lead to the degradation of the practice as a whole. This allows Mill to analyze, from a utilitarian perspective, the nature and validity of promises. An invalid promise (such as a promise to enslave oneself) does not include the same subtle moral considerations, but can be broken on the grounds of a more direct utilitarian analysis. The validity of a promise matters for Mill because when a valid promise is broken, it carries with it potential harm for the practice as a whole, and includes considerations beyond the more direct utilitarian analysis.

Such a maneuver, in dealing with expectation effects, fits with the overall view of Consequentialism in terms of Expected vs. Actual Consequentialists (Sinott-Armstrong para. 37). An Actual Consequentialist bases the rightness of an action on the consequences that actually come about as a result. An Expected Consequentialist bases rightness on the various probabilities related to the different value outcomes that could come about as a result of an act. Most commentators agree that Expected Consequentialism is the more plausible form of the theory, because Actual Consequentialism might require that we deem certain acts of wicked intention to be morally right and vice versa (Sinott-Armstrong para. 45 Hooker 2)¹⁸

Expected Consequentialism is able to deal with expectation effects far more easily. A sensitive Expected Consequentialist could simply include in his calculations

¹⁸ I will discuss the distinction between the two theories in greater depth at the end of this chapter p. 58-60.

the chance that breaking a promise will lead to the degradation of the practice as a whole and then count this disutility against the violation of a promise.

This might not completely solve the problem for AC however. An AC agent might be required to take into account all the foreseeable effects of her actions, including expectation effects, but that simply might not be possible. "It is sometimes claimed (and must be admitted) that we hardly ever have all the information required for the application of (AC)" (Lyons *Forms* 144). Lyons proposed solution is to allow for general "rules of thumb." Smart and Moore, among many others, have suggested the same (Smart 42). This concession need not represent a total shift from AC to RC since the rules will only be guiding, not overriding, and their purpose is judged according to AC (Lyons *Forms* 120).

D.H. Hodgson presents a possible counter to Lyons'/Expected Consequentialism's solution. He shows that, in a universalized AC society, the problem of detrimental expectation effects may not be done away with so easily. If all members of a society behave as AC agents *and* understand that others behave this way as well, then before a single promise is broken all agents will intuitively understand that promises lack intrinsic importance and are binding only inasmuch as they serve overall utility and are therefore not independently binding at all (Hodgson 40).

Indeed, going beyond Lyons'/Expected Consequentialism's proposed solution, this dilemma extends to Sidgwick's solution as well. He attempted to show that the promisee's trust is betrayed in cases where a promise is broken, resulting in a significant loss of well-being. However, Hodgson believes that, in a universal AC

society, a promisee cannot plausibly have this trust and so no such disappointment need result from a broken promise (Hodgson 42). As a result, promises are severely endangered in a universalized AC society.

Hodgson's problem seems somewhat more plausible in reference to Lyons' solution than it is against the solution presented by Expected Consequentialism. In Lyons' solution a greater and greater emphasis is placed on promises as they approach a threshold point where breaking a promise does more harm to the overall practice than any benefit of breaking the promise is worth. In the case of Expected Consequentialism, from the start this possibility is recognized and promises are given an added weight as a result. (Perhaps under Expected Consequentialism, following Lyons' approach to a degree, the negative expectations are increased as more and more promises are broken.) Regardless, from the start, a sensitive Expected Consequentialist should understand the possibility of a broken promise leading to harmful expectation effects. Under this model, it appears that Hodgson's problem can never get off the ground. There will be no widespread distrust in promise keeping from the start, because capable AC agents understand the great value of this practice and recognize the need to uphold it. They will perceive this in themselves and will therefore expect it in fellow, likeminded AC agents.

Following this line of reasoning, we can expect great disappointment and shock when a promisee is betrayed in this universalized AC society. Therefore, Sidgwick's defense is upheld as well. We should include this disappointment in the overall Consequentialist calculation and so promises will be further protected.

Peter Singer counters Hodgson's claim in another way by highlighting the importance of impartial beneficence in a universal AC society. In such a society, one would not need to tell a lie since a truly impartial and beneficent person would not need to be deceived in order to accept a decision that would lead to increasing overall well-being. Similarly a promise would only be broken in cases where the promisee would desire that that promise be broken (that is, in cases where it raises the overall well-being). For example, in the "desert island deathbed" scenario, the AC agent's obligation is to lie to a person who makes a deathbed wish that would be contrary to well-being promotion. Since (in a universalized AC society) no person would make such a wish that was contrary to well-being promotion, this obligation to lie would never arise. The need for an AC agent to lie or break a promise only comes about in a society where others are not motivated by AC, and are instead motivated by such things as "self-interest, malevolence, pride, and so on" (Singer 97). Therefore, Hodgson's claim about the disposition to lie or break a promise in a universalized AC society fails to get a foothold. "His argument is based not on the existence of a reason for lying or breaking a promise, but on the absence of a sufficient reason for telling the truth or keeping a promise" (Singer 97).

Singer illustrates by way of an example. Imagine that in this hypothetical universalized AC society you are approached by a stranger who says, "There is a very good film on at the local cinema this week." Following Hodgson's objection you could not accept this information as true. It may be that the film is quite bad and the stranger only told you this so that you would pay for a ticket and help save the cinema from bankruptcy. However, Singer shows that if that were indeed the case,

and both people were truly motivated by impartial beneficence *and* knew this about each other then the stranger would just come right out and say it (Singer 100). The reason to lie in the first place would not appear in a universalized AC society.

However, Singer makes certain assumptions that might not be true of a truly universalized AC society. He assumes that in this society, because people have accepted the validity of AC, they will then agree on the same courses of action. However, simply because agents have all agreed on act-based well-being promotion as the ideal moral code, it does not follow that all agents would agree on how to best promote well-being. Returning to the cinema example, one agent might think that the cinema should be supported, while another might think otherwise. Perhaps this disagreement arises over differing opinions of how to promote the same values or indeed over different conceptions of which values are ranked highest within well-being. In this case an AC agent might be induced to lie for moral reasons in order to get the other known AC agent to help bring about his own vision of well-being promotion.

In a related sense, just because AC agents have accepted the validity of AC does not preclude the possibility that they will act contrary to AC for self-interested reasons. In our society, scores of people commonly act in a way they know to be contrary to the moral theory they take to be correct. Both people may have accepted AC but one might lie to the other so that she will go to the cinema he owns, purely for self-interested reasons. Alternatively, they might think their reasons beneficent, but are only falling prey to “cooking up” bogus moral reasons that coincide with

their self-interested reasons. This common behaviour further emphasizes the importance of “rules of thumb” for AC.

It seems then that Singer’s utopian vision of an AC society does not meet Hodgson’s objection. Instead, Expected Consequentialism can consistently safeguard against the degradation of valued practices Hodgson imagines in a universalized AC society.

Another possible way for AC to confront the problem of expectation effects would be to emphasize the significance of the theory as a criterion of right action as opposed to as a guide to moral decision-making. Sidgwick endorsed the idea that AC could, with full accuracy, describe what made an action right while advocating that people do not follow AC, but instead their common sense morality. Bentham and Mill appear to have had a similar outlook. (Sinott-Armstrong para. 38) As Sidgwick puts it, “it is not necessary that the end which gives the criterion of rightness should always be the end at which we consciously aim: and if experience shows that the general happiness will be more satisfactorily attained if men frequently act from other motives than pure universal philanthropy, it is obvious that these other motives are reasonably to be preferred on utilitarian principles” (Sidgwick 413). In other words, it may be that expectation effects thrust AC into a self-defeating position when it comes to guiding actions, and yet it would remain unscathed as a correct moral theory. AC will then be self-effacing, but not self-defeating (Parfit *Reasons* 40).

Opponents attacked hedonic utilitarianism “on the ground that the general happiness will be best attained by inequality in the distribution of each one’s

services. But so far as it is clear that it will be best attained in this way, utilitarianism will necessarily prescribe this way of aiming at it” (Sidgwick 432). One can imagine the analogy of a cannon. In this case one rarely points directly at their target but has to make alterations for gravity, wind conditions, internal deficiencies, etc. in order to achieve the goal of hitting the target. In this way utilitarianism can adapt to prevailing societal conditions and even internal deficiencies of character in order to achieve the highest possible overall utility. One can alter one’s motives to make beneficence more psychologically tolerable. An example of this might be to give special regard to one’s loved ones, instead of facing discouraging difficulties in trying to aid the world’s worst off, evincing the proverb “charity begins at home.”

Sidgwick advocated for utilitarianism as an esoteric morality that could be used to directly motivate only those “persons defined by exceptional qualities of intellect, temperament, or character” (Sidgwick 489). Such people should be careful to keep their theory shrouded in “secrecy” and “concealment” and should ensure that “the vulgar should keep aloof from his system as a whole, in so far as the inevitable indefiniteness and complexity of its calculations render it likely to lead to bad results in their hands” (Sidgwick 490).

Naturally such language opens the door to charges of elitism. Bernard Williams derisively called such esteem for esotericism “government-house utilitarianism,” a method uniquely suited to Sidgwick’s era of oppressive colonialism (Sen *Introduction* 16).

Taken one way, Sidgwick’s theory does unfairly advocate for an elitist system. It seems to call for the dividing up of society along lines of those who are

capable and those who are incapable, and then asks the capable ones to lie to, or at least keep secrets from, the morally incapable. Taken at its worst it might assert an essential link between education and one's capacity to be moral, and therefore judges that all uneducated people simply cannot be good on their own, but must be fooled into it.

It might however be possible to do away with the more objectionable elements of Sidgwick's suggestion, while preserving its ability to solve problems of AC self-defeat. Certainly, in the century and a half since Sidgwick made his suggestion, history has shown us the repugnance of forming strong divisions within a society, and indeed the immense loss of well-being to which this ultimately leads. R.M. Hare, then, takes up Sidgwick's suggestion, but alters the focus from divisions within society to divisions within each individual. He explains that each of us has an "archangel" and "prole" within us (Hare *Method* 45). At times our moral thinking can be perfectly rational, and we can correctly calculate all the relevant moral features in order to determine the right action. In this "archangel mode" an AC agent could fully take account of future harmful expectation effects, as well as avoid "cooking up" an analysis that serves his own interests, and any other difficulty involved in performing an AC analysis. At other times, we think like the "prole." Here we fall prey to the many difficulties involved in moral decision making, such as underestimating expectation effects or cooking up self-interested reasons. When in this frame of mind, we should act according to rules of thumb, as Lyons and Sidgwick suggested, or as Hare prefers to call them, "prima facie principles."

Hare's internal division, similar to Aristotle's "intellect and character," still raises some problems. How do we know when we can think with full rationality and when we cannot? There is no guarantee that in a "cool hour" we will not fall prey to cooking up reasons or other subtle mistakes. Similarly, McNaughton doubts that our moral thinking can be "compartmentalized" in this way (McNaughton 180).

In any case, both Sidgwick and Hare believe that, at times, some can rationalize in a full and sensitive way that anticipates harmful expectation effects along with many other subtle factors. When we cannot do this, AC itself can simply propose that we forego AC analysis and operate according to "rules of thumb."

However some authors believe that such self-suppression undermines a moral theory at a basic level (Scheffler *Rejection* 48). A moral theory that gives up ground in this way fails to guide our actions, and a moral theory that fails to guide our action can be said to fail in some way as a moral theory.

In non-moral enterprises such self-suppression is often called for and need not result in incoherence. Take the example of a student taking an exam (Scheffler *Rejection* 46). It would be beneficial for the student to be calm during the exam. However, certain people, when actively trying to be calm, actually become more agitated. So in such a case in trying to make the student calm one should not advise them to be calm.

Simon Blackburn raises the example of a referee (Blackburn 40). Presumably a referee participates in an activity because they value what the sport can bring about, such as a sense of suspense, excitement or camaraderie. Does this mean if a referee could make an incorrect call that would undoubtedly stimulate greater

suspense, that he should then do so? Presumably not, but it is not so obvious why. It is a very similar case to that which the AC agent finds himself in. Blackburn finds that the referee need not embark down this incoherent path, but can rely on his “participatory role” (as opposed to a “reflective role”) in order to justify his decision to make a correct call. In a similar sense a soldier, as a participant, follows orders they might disagree with (Blackburn 40). In both cases, their “overall rationality” tells them to forego their case-by-case judgments in order to better achieve their goals, and this need not entail a contradiction.

In the case of self-effacing Consequentialism, Blackburn finds that the theory can only serve as an explanation while giving up ground as a justification that guides our actions. Some deontological theories, on the other hand, can serve as justification without explanation (Blackburn 43). By this he refers to weak forms of Deontology that fall back on rule worshipping. The rules provide a full justification for an action, but these rules cannot be explained via their connection to outcomes. It should be noted that more plausible forms of Deontology do care about consequences and therefore provide an explanation that supports which rules are selected.

As already noted, many believe that a moral theory that cannot serve as a justification that guides our action is no moral theory at all. Samuel Scheffler and Derek Parfit disagree by pointing out that there is a difference between whether something is true and whether it should be believed. Scheffler uses the analogy of a scientific theory (Scheffler *Rejection* 52). It might have been better had no one known about nuclear physics in the lead up to the Manhattan Project, but that is

very different from saying that what nuclear physics reveals is untrue. It seems more than likely that a moral agent can know how to act rightly without understanding the finer points of what makes an act right, just as a golfer can play well without knowing the finer points of the physics of ball movement. (Sinott Armstong para. 37).

However, scientific theories are of different nature than moral theories. Moral theory is essentially linked to human attitudes and intentions. In this respect, can AC give up its role a guide to action and exist only as a criterion of rightness? It seems doing so leads to counterintuitive results.

To bring out these intuitions I will look again at the distinction between Actual and Expected Consequentialism. Indeed, these two forms are sometimes described as Objective Consequentialism and Subjective Consequentialism (Sinott-Armstrong).

The former (Objective or Actual Consequentialism) treats the theory as delivering only a criterion of right action that can assess any moral action. Yet these assessments lead to many difficulties, evincing a problem with this style of moral theory. Take the following example. A man intends to violently rob someone for his own personal gain. In doing so he kills the victim who, it turns out, was going to cause a horrific genocide. Because of the robber's action, a large group of people have been saved. His action (unwittingly) promoted well-being. Unwittingly or not, Actual (Objective) Consequentialism must judge his action to be best. But this flies in the face our intuition, which holds that the robber expected to do great harm to another only for his own, comparatively small, personal advantage. His intentions

are critical to our moral assessment, but Actual Consequentialism cannot take them into account. Actual Consequentialism then fails to deliver a criterion of rightness with which our intuitions agree.

The theory has other intuitive problems. In the last case we saw that it judged the robber's actions best because it prevented genocide. But can this assessment be made at this point? It seems the theory cannot decisively judge an act right or wrong until the end of time. Perhaps in billions of years the robber's act leads to a much worse catastrophe than the genocide it prevented. Granted, this is highly unlikely, but because it has not happened yet we are forced to use the language of likelihood. Expected Consequentialism can dismiss highly unlikely outcomes. Actual Consequentialism cannot.

Expected Consequentialism can run into its own intuitive problems though. In one sense this theory excuses negligence on the part of a moral agent. If an Expected Consequentialist agent does not foresee the terrible consequences of an act and if it maximizes utility in his faulty assessment, it is the act he is encouraged to perform. Also, his right action would not be the same as that of someone else in the same scenario who did properly consider the expected outcomes. How can one agent act in a way that is said to be right, while another in the exact same scenario performs the same act and it is said to be wrong? Intuitively we do not think this can happen because we find fault in the negligent person.

In order to solve this problem, proponents of Expected Consequentialism shift the focus to that of "reasonably foreseeable" consequences (Sinnott-

Armstrong). However, what is “reasonably foreseeable” will depend on an objective viewpoint. In this sense the theory must go beyond the intentions of the agent.

It seems then that on an intuitive level neither of these forms of the theory, when overly isolated from the other, will deliver an acceptable analysis. In order to remain intuitively plausible, AC must retain its role as both a criterion of right action and as a guide to moral action. When either element is dismissed, the theory delivers unintuitive results.

If the theory is to maintain its role as a guide to moral action, it must then look at expected consequences, and take the possibility of harmful expectation effects into account for this analysis. However, it will often be exceedingly difficult to precisely take into account this possibility. As a result AC can institute “rules of thumb” as many authors have suggested. These rules of thumb will not entail abandoning the AC theory because their institution will be justified according to AC and they can be overridden when doing so clearly results in an act of which AC would approve.

Chapter 4: Justice

Critics object to AC on the grounds that it is incompatible with justice. They claim that AC fails when it attempts to deliver answers on questions of tradeoffs and the dismissal of basic moral rights. Additionally AC is said to be unable to deliver a fair distribution of well-being, and it also disregards the importance of just deserts (whether in terms of punishment or reward).

In the case of tradeoffs and rights, the worry is that rights will be considered only derivatively. If denying the rights of a person will lead to greater well-being overall, then an ACist must take this course of action. In so doing, the ACist denies the inviolability that is said to be essential for rights, and so denies the validity of rights as a whole. Indeed, Bentham straightforwardly demonstrated the utilitarian lack of respect for rights when he described the idea of natural rights as “nonsense,” and imprescriptible natural rights as “nonsense on stilts” (Bentham *Anarchical* 2).

It is also said that AC will call for an unjust distribution of resources. In advocating total impartiality, AC cannot give priority to those who most desperately need it. In light of this shortcoming, some authors have suggested a modified form of AC that gives priority to the well-being of those who are worst off (Hooker 59).

Finally, some object that AC cannot deliver a conception of desert that accords with our intuition. We perceive an injustice when a person commits a crime and remains unpunished, or when a person acts in a praiseworthy way but their good deeds are never acknowledged or rewarded. AC, however, is predisposed to be against punishment per se since punishment is, on some level, the deliberate

infliction of pain (or negative well-being) on someone simply for the pain's sake. Even if all the benefits of punishment, such as deterrence or rehabilitation, could be achieved without actually inflicting pain, this would not satisfy a retributive conception of punishment. On some level the pain must be felt. AC, however, will advocate for punishment only when it somehow promotes overall well-being, either by serving as a deterrent to others or by preventing the perpetrator from causing further harm. As for rewards, AC will be predisposed to support any form of rewards, because they are intended to contribute to another's well-being. However, AC will support a just reward no more than an undeserved reward, except in cases where the reward serves as further incentive.

In this chapter I will look at these three related issues and whether they should lead to the conclusion that AC cannot account for justice within its moral framework.

To begin, I will look at John Stuart Mill's argument for a utilitarian explanation of justice, as outlined in his *Utilitarianism*. Although I find his arguments persuasive in many respects, in other ways they fall short. Utilitarianism will give justice a very strong weight as it relates to our very important sense of security. Also, Mill shows that our reactions to acts of injustice and just rewards suggest a strong relation between justice and utility. Still, Mill's arguments cannot fully explain our sense of retributive justice, and while they do show a close connection between utility and justice they do not prove this connection to be a foundational one.

Next I look at John Rawls and Robert Nozick's arguments against an AC backed form of justice in regards to respecting rights. These arguments undermine much of Mill's defense because they support the intrinsic inviolability of rights, whereas Mill's arguments support rights only inasmuch as they instrumentally promote utility. In this context I go into a discussion about rights and question their inviolability.

I then look at Ross' idea that, in certain circumstances, we can violate another's right, but AC can be blamed for doing so whenever even the least amount of well-being is thereby promoted. In relation to this I consider the arguments of Frances Kamm and Judith Jarvis Thomson that seek to deny a well-being based foundation of rights. I argue that some of their illustrative thought experiments do not deny this connection but, indeed, can be shown to support it.

Next I will discuss Rawls' objection that AC is unjustly insensitive to the distribution of well-being. In order to counter this objection, I look at the possibility of invalidating sadistic pleasures, and discuss certain problems with this approach. Another possibility then might be to alter the components of the AC maximand,¹⁹ but this might lead to problems of inconsistency. Rawls' objection also gains traction without allowing for sadistic pleasures, and so I look at these types of cases and ask whether they are too far-fetched to deliver a plausible intuitive judgment.

I attempt to counter the objection of unjust tradeoffs in AC through use of the concept of "diminishing marginal utility," (DMU) which explains that preventing

¹⁹ I use the term maximand insofar as the conception is a maximising one; otherwise one can understand the term as denoting for present purposes the overall amount of value to be promoted

pain, as opposed to bestowing pleasure, will be the main focus of AC. I examine the validity of this concept and argue that it will call for an organic prioritizing of the worst off. Although this still might allow for tradeoffs between pain and pleasure at some level, it occurs at such a drastic level that our intuition fails to deliver a clear response in these cases.

Finally I look at Amartya Sen's idea of incorporating rights within a Consequentialist maximand and Nozick's arguments against this.

In many respects the justice problem, particularly the tradeoffs and rights aspect of the justice problem, relates to the expectation effects problem discussed in the last chapter. Some commentators see rights and rules as an effective solution to problems of coordination that are said to lead to self-defeating expectation effects (Scanlon 83). In the last chapter I attempted to show that an AC agent could incorporate rules of thumb into their moral analysis to prevent this issue. It is worth noting that Rule-Consequentialism (RC) has a much simpler task in accommodating such rules in its moral analysis.

The justice problem differs from the expectation effects problem in that the matter no longer revolves around self-defeat. The strategy in the last chapter relied on AC's self-corrective ability when universalized AC morality resulted in paradox. In the case of tradeoffs, well-being may be maximized without paradox when rights are allowed to be broken (even when this disposition is universalized) but our moral intuition still finds something wrong in such acts of injustice. However, a sensitive treatment of well-being promotion might make room for our sense of justice, and may even show that well-being promotion coincides with justice.

In the fifth chapter of *Utilitarianism*, John Stuart Mill attempted to show the high value of justice within a utilitarian framework.

He argues that utilitarianism is highly committed to justice and, indeed, with its commitment to impartiality, makes a higher claim to justice than other moral theories that could allow for unfair bias.

On the issue of tradeoffs and rights he brings up their relation to our security. If society did not make a strong commitment to uphold our individual rights then this would have a detrimental effect on our sense of security. As such a theory focused on well-being will be greatly concerned with rights, since “security no human being can possibly do without. On it we depend for all our immunity from evil, and for the whole value of all and every good” (Mill *Utilitarianism* ch. 5). In this way security plays a foundational part in any attempt to promote well-being. Because rights provide this security, an AC agent will give them precedence, since any attempt to build on another’s well-being without this foundation will be futile. In light of this insight, AC will not undermine rights for some trivial boost to well-being, anymore than it would advocate providing cake to someone dying of thirst.

Mill demonstrates that throughout human history we have been greatly concerned with the preservation of our society’s security. He believes that the revenge instinct has served as a mechanism to protect against the abuse of others, since they will understand the danger involved in this abuse. It is worth noting, in this regard, that *Utilitarianism* was published two years after *On the Origin of Species*. Mill was perhaps influenced by Darwin’s idea that certain species traits are adaptations which are crucial for the survival of the species. Mill might have seen

the revenge instinct as such a crucial adaptation, thus giving it an extremely strong utilitarian weight.

Mill attempts to show the underlying utilitarian rationalization for justice by pointing out the confusion involved when people try to speak about justice as divorced from utility. "If justice be totally independent of utility (...) it is hard to understand why that internal oracle is so ambiguous, and why so many things appear either just or unjust, according to the light in which they are regarded" (Mill *Utilitarianism* ch. 5). Of course, disagreements over justice cannot alone show its dependence on utility. Indeed many evaluations of utility, independently considered, will lead to their own disagreements.

However he does show that the few agreements we do have in regards to justice seem inextricably linked to utility. Our outrage over matters of justice seems directly proportional to the utility involved. "The most marked cases of injustice (...) are acts of wrongful aggression" (Mill *Utilitarianism* ch. 5). These also cause the most pain or greatest loss of well-being. Next comes "wrongfully withholding from someone something which is his due," and after will come the failure to reward "good for good" (Mill *Utilitarianism* ch. 5). It does seem clear from these rankings that the severity of an injustice relates directly to the utility involved.

In the case of rights we can see a parallel. The ambiguity of some rights seems only resolvable by appeal to the utility of its consequences. As Scanlon notes, we only know that restricting speech constitutes a violation of the right to free speech by examining the consequences of the restrictions (Scanlon 74).

For just punishment, Mill continues to observe the important role of utility. When a punishment is excessive we find it unjust because we object to “any amount of suffering beyond the least that will suffice to prevent him from repeating, and others imitating, his misconduct” (Mill *Utilitarianism* ch. 5). We can see here the utilitarian emphasis on deterrence, protection, and rehabilitation in punishment.

This conception of justice might leave out an important aspect that our intuition observes. It seems in many cases where deterrence and rehabilitation play no part we still see a reason to punish an offender. Take for example, the continued search for those involved in war atrocities from long ago. For the most part, we feel it just to punish these people well past the date of their crimes, when there is no danger of them committing the crime again. As a deterrent, many “argue that the very nature of the crimes prosecuted by the International Criminal Court – war crimes, crimes against humanity and genocide – make them resistant to deterrence through prosecution, and that the record so far suggests that not only do international prosecutions offer little hope of preventing future atrocities, they in fact risk prolonging conflicts” (Grono para. 5).

A utilitarian analysis still might explain the motivation for such punishments. As already noted, Mill showed the importance of a vengeance instinct in early man. This need for vengeance persists even in rare cases where the action can serve no purpose as a deterrent or prevent further wrongdoing. However the disposition to always seek vengeance in this way would serve as a deterrent in the long run, since offenders would fear the instinct in others that they observe in themselves. Common knowledge of this disposition throughout society, a disposition present even when

direct appeals to utility are absent, would lead to a far more secure society, and so this disposition would find utilitarian support.

However, to rely on a disposition in this way, indeed a disposition that will run contrary to utility promotion in specific cases, leads back to the problem of government-house utilitarianism.²⁰ To avoid this problem, AC might have to bite the bullet and classify our “thirst for revenge” in such cases as a vicious, atavistic desire, one which modern day civility should eschew. To bolster the cause of this bullet-biting, we should note that such cases tend to raise plenty of debate, and our collective intuitions on this issue are far from unanimously in favour of punishment without deterrent. However, most of society (myself included) does see a moral need for the punishment of people who have committed such awful acts, without regards to deterrence. It may be that this moral cause is simply a bygone viciousness in disguise, but I think it more likely that this is an area in which AC is in error, unable to account for the moral worth of retributive punishment without deterrent value.

In the case of rewards Mill sees a strong connection with utility. There may be disagreement between what is just that the individual should receive and what is just that the community should give. Mill thinks we cannot resolve this disagreement without reference to utility. Any answer “on grounds of justice, must be perfectly arbitrary. Social utility alone can decide the preference” (Mill *Utilitarianism* ch. 5). In other words, just rewards are determined by their effectiveness in providing proper incentives.

²⁰ I look at the problems with such a form of utilitarianism in the chapter on expectation effects p. 53-59.

Mill's arguments do show a strong connection between justice and utility. Furthermore, they show that many questions about justice are left ambiguous until utility is considered. However his arguments only show a connection, and do not prove that justice is motivated by utility promotion, or indeed that the promotion of utility will not prove itself incompatible with justice.

Rawls objects to Mill's style of defending justice on utilitarian grounds. "From a utilitarian standpoint the explanation of these precepts and of their seemingly stringent character is that they are those precepts which experience shows should be strictly respected and departed from only under exceptional circumstances if the sum of advantages is to be maximized" (Rawls *Theory* 23). To treat the precepts of justice in this way is to treat justice only instrumentally. The only way to respect someone's right is to not violate that right, and not by merely including them in an overall aggregative calculation. By resorting to this latter approach, utilitarianism fails to account for how well-being is distributed and will therefore not take "seriously the distinction between persons" (Rawls *Theory* 24).

As Nozick puts it, justice requires that we respect "side constraints" which prevent us from acting in a certain way (such as violating a person's rights) even when the goal of doing so is to increase overall well-being (Nozick 30). To not respect these side constraints would be to treat rights as instrumental or derivative of well-being promotion. In so doing we cannot respect rights at all since an essential characteristic of rights is their inviolability.

These arguments undercut Mill's defense. His defense, along with the expectation effects solution provided in the previous chapter, seek to uphold certain

rules by asserting the immense Consequentialist value of having them upheld throughout a society. Mill also showed the connection between utility and rights but, as already noted, this connection does not prove a utilitarian basis for justice.

Mill's response does nothing for Nozick and Rawls, who see any instrumental treatment of rights as opposed to their nature. As Lyons puts it, the utilitarian model can account for "legal rights" and the immense utility they provide, but will not be able to account for "moral rights" since this places an intrinsic value on upholding rights, regardless of the effect on overall well-being (Lyons *Forms* 150).

Consequentialist theories "do not generate any obligation to adhere to the rules that they regard as justified. And they cannot do so unless they are restricted for just such a purpose" (Lyons *Forms* 167). H. J. McCloskey identifies a similar dichotomy between, on the one hand, moral rights, and on the other, what he calls legal and social rights. "We may believe that there is no moral right to propagate immoral, false beliefs, nor to publish or read the obscene and pornographic. Yet we may rightly believe that there are good moral reasons for according the legal and social rights to do these things" (McCloskey 123).

At this point it is worth questioning how we regard the obligation to respect moral rights, and how essential inviolability is to this conception. I believe it can be shown that our intuitive view of rights does not allow for complete inviolability, and that this view can be made compatible with the ACist treatment of rights.

The inviolability of rights is difficult to prove. The famous trolley example illustrates that many of us would willingly kill a person, violating their right, if it would lead to saving five others. Modifications of the example lead to different

responses but almost any respondent agrees that if the drawbacks are great enough (if not five deaths then five million) a “rights abuse” can be justified. It seems then that rights have to be regarded within a wider moral framework. So in treating rights derivatively AC only follows the way of our common moral insight.

Still one might object that the problem with the AC conception of rights is not that it allows rights to be broken, but treats rights purely as cogs within a machine, and calls for rights to be broken whenever the least increase in well-being can be brought about thereby.

As W.D. Ross explains, it is true that our duty to uphold rights (say, the duty to keep a promise) can be mitigated under extreme circumstances. We should surely break a trivial promise if this was the only way to prevent a disaster (Ross ch. 1). However, AC errs by explaining that our moral duty to keep a promise lasts only as long as we cannot bring about more well-being by breaking the promise, *even the least degree of well-being* (Ross ch. 1). By putting the focus solely on well-being output, AC mechanizes our moral thought and reduces important relations to merely “possible beneficiaries by my action” (Ross ch. 1).

AC can answer Ross by explaining that their commitment to promises, or any rights, is co-extensive with Ross’ commitment. When a good moral agent (the kind Ross has in mind) decides to neglect a duty in favour of some greater duty this is only motivated by promoting overall well-being, even the least degree. The callousness that Ross observes in the AC agent comes about as a result of his underestimating the great value that an AC agent places on upholding rights, as seen in Mill’s argument about the social utility of rights and the argument via expectation

effects in the previous chapter. If one understands the great value of these rights, then an AC agent will only pursue a different course under extreme circumstances. These extreme circumstances may mean a right violation will lead to only an infinitesimal increase in well-being, but it is still a circumstance in which, conceivably, Ross would think it right to do the same. When two actions differ by only the least degree of well-being, it will (and should) be difficult for an AC agent to decide which course to take, since absolute precision in well-being assessments will be extremely difficult. Ross, however, seems to treat the AC agent as a robotic well-being calculator that can decide the right course in an instant. In doing so he oversimplifies the ACist aim.

This response to Ross' criticism relies heavily on Mill's idea that rights are grounded in social utility. However, many authors object to the idea that the stringency of a right is explained solely by the social utility involved (Kamm 260, Thomson 123). Kamm, for example, believes that rights stem from our nature as people and that the "strength of a right is greater than the weight of the interests it protects, because the right implicates the worth of the person, not only his interests" (Kamm 260).

On the other side of the debate authors like Joseph Raz have put forward an "Interest Theory" of rights, whereby "there is a right if and only if some interest (i.e., aspect of well-being) of some entity capable of being a right holder is sufficient to ground a duty to care for and promote the interest in a significant way" (Raz 183).

I believe an examination of cases will show that Raz' Interest Theory better explains our notion of rights.

When we are asked to adjudicate between rights, the key factor that dictates which right carries the day appears to be well-being. We have already seen how this plays out in regards to promises, where one's right to have a promise kept is justifiably withheld when this is the only way to save another (and preserve their right to life). In such a case it would be cruel and immoral to stay true to a trivial promise at the expense of another's life. Clearly the more stringent right (to life) involves a far greater well-being value than the less stringent right (to have a promise kept).

It seems that the stringency of a right depends entirely on the use that right has in protecting our well-being. As already noted, free speech is defined as a right only when it results in some type of protection of our wider social well-being. This explains Oliver Wendell Holmes' famous saying that a right to free speech does not allow one to shout "fire" (expecting to be believed) in a crowded theater (that is not on fire). The right does not apply in this case for two reasons, both related to well-being. In the first place, allowing this person to act might result in a panic, causing injury and perhaps death to the theatergoers. This loss of well-being will militate against protecting the "right" to act in this way. Secondly, society will gain no added security, essential to well-being, by protecting such false utterances about fire, as they would by protecting a reporter's right to reveal information to the public, or for somebody publicly airing their unpopular beliefs. It is the latter cases, in which we are able to "speak truth (real or imagined) to power" to which we ascribe the right of free speech, because they will ultimately add to the overall protection of the well-being of a society. When free speech provides the latter sort of benefit and results in

the former sort of harm (for example, the recent “wikileaks” scandal; or a case where a racist speech incites violence) we question whether or not the right should apply. Ultimately the answer seems to come with reference to the degree of harm or benefit that society gains as a result of setting the precedent. This shows how discussion of rights is couched in questions of well-being promotion.

Critics have tried to use thought experiments to highlight cases where rights cannot be explained solely by the well-being involved.

Frances Kamm reevaluates the Trolley and Bystander cases. She compares the standard Trolley case with a modified Bystander case, in which we could stop the trolley “by pushing another person in the way of the trolley (saving five lives) when we know that this will (only) paralyze him” (Kamm 264). She then determines that it “may be permissible to do the first (standard trolley case) but not the latter (modified bystander case)” (Kamm 264). From this she concludes that well-being, or interest, does not play a decisive role in the moral action.

It seems highly questionable in this case that our intuition would find permissible the killing of one to save five, but would find impermissible the injuring of one to save five. Indeed one should morally prefer the second course, precisely because it sacrifices less well-being. If respondents do in fact find the killing more permissible (which Kamm does not show) this may well be because we cannot imagine a case where we know with absolute certainty that the bystander will only be injured, and not killed.

Judith Jarvis Thomson believes that one may violate a right “if sufficiently more good comes about;” however, like Ross, she thinks this sufficient amount is far greater than what AC calls for (Thomson 123).

She provides the following thought experiment to show this. If two people are in danger of dying, and we could only save one, it would be permissible to flip a coin to decide which one to save. However, she contends, an interest theory of rights like AC will not find this permissible, but must decide on the well-being values of each person’s life and save the person with greater value. Therefore, AC will find impermissible acts that our intuitive notion of rights finds permissible (Thomson 163).

I find three potential mistakes in this conclusion. For one, a scalar form of Consequentialism²¹ might find that flipping a coin, while not ideal, will spare the agent from making an extremely difficult, overdemanding choice, and so it is permissible. Secondly, flipping a coin will eliminate any possibility of bias on the part of the agent, or the harmful appearance of bias. Thirdly, such a situation will not involve obvious well-being calculations. Especially when we take into account a sophisticated maximand that allows for more subtle values, we should find it very difficult to compare the values of separate lives. However, when we can take into account some more obvious features, the permissibility of flipping a coin comes into question. Imagine a case where we could save either a young child (who will presumably live for much longer and have more time to experience subtle values) or a dying, violent person (who will not only have little time left to experience value,

²¹ I discuss this in the chapter on the overdemandingness objection

but has a good chance of diminishing the value of others). In such a case, with more obvious value considerations, it could be argued that deciding to flip a coin would amount to immoral cowardice and negligence.

It seems neither of these thought experiments can be used to further the notion that rights are not couched in public well-being. Indeed a critical examination of the thought experiments seems to show the importance of well-being in interpreting rights. If we follow Mill's argument in assigning a very high social value to upholding rights, then AC can avoid the Rossian criticism that we should not violate rights when only slightly more well-being is at stake, since even slightly more well-being than is at stake with rights will entail a tremendous amount.

We can perhaps refine Rawls' objection at this juncture. AC does not support the complete inviolability of rights, but nor does our intuition. However, one can still object against the AC treatment of rights by its indifference to how well-being is distributed. There are many examples that illustrate this. In ancient Roman gladiator spectacles, some would be forced to die yet several thousands would be immensely entertained by their deaths. A lynching might confer pleasure on a large lynch mob, and this pleasure is only moderated in an AC analysis by the comparatively lesser amount of pain endured by the person lynched. AC is said to support these actions since overall well-being is increased, regardless of the distribution of this well-being. Now, surely, our intuition disagrees with trading off the immense pain of a single person for the relatively trivial pleasure of several others.

One initial defense that AC might make here is to continue to uphold the Consequentialist value of preserving rights. If we use a sensitive analysis we can see that allowing such actions to occur could lead to severe loss of security throughout a society.²² Furthermore, indulging the sadistic pleasure of the participants may lead them to gain an appetite for such acts, increasing the likelihood of those acts being repeated. Such a style of argumentation has been made against (typically victimless) sadistic pleasures, such as violent pornography and video games.

Another more profound argument AC can make is to simply disqualify the type of pleasure enjoyed by the lynch mob or the gladiator spectators. Along these lines, AC would surely oppose the death of the gladiators, but it would also see the sadistic joy of the spectators as a negative thing in itself, not as a benefit that counters the pain of the gladiator. Just as virtue is its own reward, for the gladiator spectators, evil is its own punishment.

It is a bit easier to make this justification with AC as opposed to purely hedonic utilitarianism, since we could include within the maximand of well-being something like “compassion.” However, perhaps even a purely hedonic utilitarianism could envision sadistic joy as actually a form of self-induced pain.

However, restricting pleasure in this way leads to problems. It seems to run contrary to the impartial, democratizing ideal of utilitarianism to determine which sort of pleasures count and which do not. Shouldn't we take someone at their word when they say something is pleasurable for them? To do anything else seems to

²² However, many scenarios will not involve this loss in a sense of security. For instance, in the Roman gladiatorial example, free Roman citizens would usually not have to worry about becoming gladiators themselves.

constitute a profound insult. Certainly Mill had no problem with qualifying pleasures, but having “objective list” criteria for pleasure (or well-being) can be seen as elitist and authoritarian (Hooker 41). Additionally, to rely too much on disqualifying certain kinds of “pleasures” can seem like an ad hoc begging of the question by AC to avoid these types of problems.

On perfectionist grounds we can easily assign negative value to a sadistic pleasure, but a hedonic utilitarianism seems committed to relying on the self-reporting of individuals as to how much pleasure or pain they experience. As Mill puts it, “What means are there of determining which is the acutest of two pains, or the intensest of two pleasurable sensations, except the general suffrage of those who are familiar with both?” (Mill *Utilitarianism* ch.2) It seems that in order to count sadistic pleasures as a negative for AC, we must make our maximand more sophisticated and include compassion and harmony as positive values.

However, to allow for this solution might lead to other problems. If we allow for compassion and harmony as independent values within the maximand of well-being, perhaps we should just allow for fairness itself to be included. This certainly seems the simplest solution to the justice objection. But to do so may undermine the distinct aspect of a Consequentialist project, since the theory would no longer appear to distinguish itself from deontology. Such a maneuver could be seen as “an implicit appeal to other norms, which (Consequentialists) covertly use to define (well-being)” (MacIntyre 238).

In order to avoid this criticism, I am trying to examine to what degree our sense of a just distribution can be arrived at by following only a hedonic utilitarian

approach, without including other values besides pleasure within the maximand. In the case of sadistic pleasures we seem to reach a difficult obstacle, although I will attempt to deflect some of these fears by use of the concept of diminishing marginal utility later on in this chapter. In the end, a more sophisticated maximand might have to be allowed for and this will lead into MacIntyre's objection. However, since a full examination of the components of a more sophisticated maximand is beyond the scope of this thesis, I will also leave MacIntyre's objection (as well as objections about the incommensurability of independent components of well-being) unanswered.

Still, inasmuch as fairness and compassion seem instrumental towards the promotion of utilitarian happiness, there are grounds on which to include them as desirable features within a hedonic utilitarian framework. For example, Mill, who wanted to avoid the conclusion that utilitarianism prioritized the "lower pleasures" (and sadism could count as one) saw "dignity" as "so essential a part of happiness of those in whom it is strong, that nothing which conflicts with it could be, otherwise than momentarily, an object of desire to them" (Mill *Utilitarianism* ch. 2). Fairness could be construed as an equally essential part of happiness, which would oppose sadism in almost all its manifestations. However, it remains a lasting question, to what degree seeing these apparently distinct values as instrumental towards happiness counts as "an implicit appeal to other norms" (MacIntyre 238).

In light of these problems, the inclusion of sadistic pleasures remains a lasting problem in the context of harmful tradeoffs.

Furthermore, there are other examples of harmful tradeoffs that do not involve sadistic pleasures at all. One popular example is the case of a surgeon who can save the lives of five very useful people by murdering a healthy, innocent and comparably useless person in order to harvest his organs (Thomson 145). On any formulation the saving of these five people's lives will constitute a legitimate positive pleasure for them. Another example was given in the film *Soylent Green* in which the elderly and infirm were killed and used for food to save other healthy members of society from starvation. Again, avoiding starvation must be considered as a legitimate increase in well-being.

It is worth noting that, although our intuition tends to censure these types of actions, we find in them far more justification than in the case of lynch mobs or gladiatorial shows, and they lead to far more disagreement about what is right to do. However, generally it does seem that many cases of legitimate positive pleasures cannot justify a single case of causing great harm. No amount of ice cream licks can justify torture (Raz).

One strategy AC can employ to dispel the fear that it will condone such actions is to raise concerns over the plausibility of the scenario. Following Mill, if this murder were found out, it would lead to a strong loss of security (and therefore well-being) throughout the entire society. Furthermore, how could the doctor know that the "organ recipients will emerge healthy, the source of the organs will remain secret, the doctor won't be caught or punished for cutting up the "donor"²³, and the doctor knows all of this to a high degree of probability (despite the fact that many

²³ However this would probably no longer be a concern if the AC justification were accepted and made into law.

others will help in the operation)” (Sinnott-Armstrong)? It can be argued that it is these practical concerns alone that our intuition latches onto in condemning such an act.

Even if we were able to contemplate this scenario without these practical complications, many Consequentialists feel comfortable in biting the bullet and saying that (under these extremely rare circumstances) the doctor should kill the one to save the five (Sprigge 264). Of course these authors are not saying that under normal circumstances a doctor can be allowed to act in this way.

John Harris outlines a variation of the transplant case in his *Survival Lottery*. In this example two patients (Y and Z) need a new heart and new lung, respectively. They maintain that a third person (A) should be forced to give up their organs (and their life) to save Y and Z. We object to this procedure because it is seen as killing A. However, should we not see the reverse as killing Y and Z? After all, for doctors not to operate on Y and Z when healthy transplant organs are readily available would be seen as neglect equivalent to killing (Harris 81). In a sense, A’s organs are readily available. To not kill to save Y and Z is “a decision to prefer the lives of the fortunate to those of the unfortunate” (Harris 81). As we have already seen, the effects on our communal sense of security would be jeopardized when A is selected arbitrarily, so if this action were to be made public we would rightly dismiss the idea. However, what if everyone were given a number in a lottery, to be chosen (based on criteria like age optimization) by a computer program? This would no longer undermine security in the same way because “the truth is that lives might well be more secure under such a scheme” (Harris 84). Harris contends that such a scheme would be

“rational on prudential grounds” and “mandatory on utilitarian grounds” (Harris 87).

Why, then, does our intuition seem to disagree with this proposal? One might argue that to admit the possibility of killing a random innocent in this way reveals a corrupt mind. But, as already noted, why should we see Y and Z as not worth saving in the same way as A simply because they were unlucky? Would failing to consider this not be the “product of a closed mind” (Harris 86)?

Perhaps a better reason for why we disagree with this idea is, again, because of the implausible conditions that our intuition fails to account for. We think it implausible that this system would not be manipulated to save the friends and family of the powerful. After all, humans are corruptible, but diseases are not. So we prefer to have our security threatened in a way (by disease) that cannot be manipulated by the powerful. In addition, we see practical problems in discerning when a disease is the result of bad fortune and not recklessness on the part of the sick person. So Harris, like our intuition, is opposed to the scheme – but “whether we could retain this confidence in our intuitions if it were to be confronted by a society in which the survival lottery operated (without these practical difficulties), was accepted by all, and was seen to save many lives that would otherwise have been lost, it would be interesting to know” (Harris 87).

So, the far-fetched conditions of the transplant scenarios do seem to answer our intuitive disapproval. When we try to allow for some of the more far-fetched conditions (such as accepting as possible an incorruptible computer program) as Harris does, the intuitive answer becomes far less obvious.

R.M. Hare objects to the use of far-fetched scenarios because it confuses our intuition. He believes that AC does not conflict with our every day intuition. The conflict only arises in bizarre corner cases, highly uncommon scenarios to which our intuitions are not suited (Hare *Thinking* 49). Using these far-fetched scenarios to provide any type of objection to utilitarianism is the anti-utilitarians' "commonest trick" (Hare *Theory* 123).

Sen, on the other hand, thinks it a methodological error to entirely dismiss our intuition in bizarre scenarios. In common cases, "intuitions based on quite different principles tend to run in the same direction" (Sen *Rights* 197). Indeed it is only because error will not arise in common scenarios that a fallacious theory is able to gain any acceptance in the first place. The only way to "discriminate" between theories is to envision cases where they will prescribe opposing actions (Sen *Rights* 198).

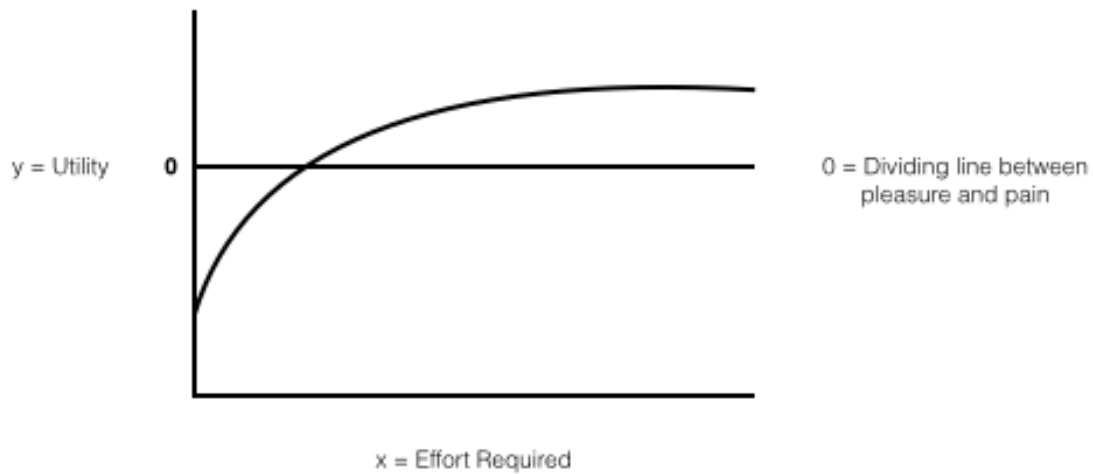
On the other hand, it must be admitted that, in envisioning these scenarios, one cannot completely dismiss their past experiences. For example, if it were specified that punching someone did not hurt but actually gave a sense of pleasure, many people would still say they would not punch their mother in the face. This is understandable. One cannot divorce themselves entirely from the real world where heavy impacts hurt and doctors cannot know with full assurance that their surgeries will be successful and their disturbing methods will not be discovered.

In arguing these matters it seems clear that we can easily fall prey to making use of intuitions when they help our cause (as in the trolley example) but disavowing their validity when they do not (as in the transplant example).

What the preceding examples of gladiator spectacles and killer surgeons do show is that comparing well-being values of different outcomes is very difficult. How should we compare one instance of tasting ice cream to another instance of being tortured?

It does seem very common for authors describing the tradeoffs problem to provide some wildly unrealistic quantifications of pleasure or pain. For example, Sen described a case in which a storeowner, Ali, is in danger of being beaten by a group of racists, called Bashers. If the bashing occurs each Basher will receive an extra “point” of utility whereas Ali will lose five “points” of utility (Sen *Rights* 192). Because there are ten Bashers, it is said that AC will approve of the action. Even disregarding any notion of the illegitimate pleasure of the Bashers or the loss of overall security involved, it seems highly implausible that the act of bashing a person (which taken extremely charitably, might provide a sense of camaraderie, release of aggression and exercise, as well as sadistic joy) when denied results in a fifth of the disutility of being bashed (which will lead to not just extreme pain, but a sense of lack of security and the hatred of your fellow man).

Consequentialists have attempted to explain well-being values through the use of “diminishing marginal utility.” This concept explains that well-being follows a square root function so that more and more effort is required to add the same amount to a person’s utility (Brandt 312). Here is a graphical representation:



This idea is one borrowed from economics and follows much the same pattern. If a person has no money then giving him a dollar will add greatly to his well-being, perhaps allowing him to buy some bread and stave off death by starvation. If a person is a billionaire then giving him a dollar will add practically nothing to his well-being. In AC the x-axis will indicate effort expended, whereas in economics this is usually simplified as dollars spent, but the principle is the same.

Diminishing marginal utility (DMU) serves as an explanatory model that reveals important aspects of the process of AC thought.

Firstly, it explains how to quantify a person's well-being. The early parts of a person's "well-being graph" will be the most essential, and they will include things like, in order of importance, freedom from being tortured, then killed, then injured, etc. After a point, additions to one's well-being will be unessential and practically negligible (Shaw 239). For instance, once someone has a sense of personal security, meaningful personal relations, etc. then the addition of, say, a fifth suit will have a

near negligible impact on that person's well-being (relative to freedom from torture, etc.). In this way, diminishing marginal utility suits our intuition about what we would want for ourselves.

Diminishing marginal utility then gives Consequentialist precedence to the prevention of pain over the addition of pleasure. We can imagine a zero point on the graph representing this separation.²⁴ Most of a person's well-being will come before this point.

This explains how an AC agent should act. Often an agent can best proceed simply by not causing harm, and leaving "well enough alone." An AC agent should also try with all their effort to prevent the torture, death, etc. of others, even if this means giving up on their own pleasure, or the pleasure of those they care most for. This lends force to the overdemandingness, neglect of personal relations and integrity objections.

We can see how this concept does a lot of work for AC in denying the permissibility of harmful tradeoffs. Not torturing someone occurs near the beginning of the well-being function and provides tremendous additions to a person's well-being. Giving someone ice cream occurs near the end of the well-being curve, and will only be enjoyed after many more necessary aspects of well-being have been satisfied first.

²⁴ Admittedly, what constitutes freedom from pain as opposed to an addition of pleasure is up to a sophisticated debate. Should a child not receiving ice cream be considered a pain or a loss of pleasure? What if all the other children receive ice cream but not her? Actual cases will quickly become very complex for this model. The complications only increase when we implement a more sophisticated maximand. The model suffices only to show on a general level the priorities an AC agent should have.

In this way AC will put precedence on acts that prevent pain over acts that confer pleasure. This will lead to a natural respect for rights of others since these involve freedom from having harm done to them and harm takes precedence in terms of utility. Karl Popper advocated for a form of utilitarianism that only focused on the prevention of pain (Smart 28). We can see now, taking diminishing marginal utility into account, that the natural form of AC will not be far off from Popper's proposal. We saw that Mill naturally ascribed greater importance (and greater utility) to matters of prevention of "wrongful aggression" than to "rewarding good for good" (Mill *Utilitarianism* ch. 5). Pain is given precedence over the pleasure involved in rewards.

Rawls proposed a form of distribution sensitive utility bestowal that gave precedence to the member of society with the lowest level of utility. This is called the maximin, where the best distribution is one where the lowest member of society had more utility than in any other distribution, regardless of the utility values of the other members (Rawls *Theory* 133). So he would prefer, or at least find more just, a three-person society with utility values (8, 10, 10) to a society with utility values (7, 10, 30). When we consider diminishing marginal utility it seems impossible that one person (30) could attain such a high level of well-being since after a while the utility curve will turn into a flat line. It seems that Rawls is mistaken in conflating the idea of utility with money (where one extra dollar will always bring one extra "point" of utility). We do not want to see a rich person receive enough extra money to get a fifth suit (30) when this same amount of money could be used to prevent someone else from going hungry for a while (7). But these two occurrences are only worth the

same amount of money, not utility, and this is precisely why we are opposed to it. If the difference between a week's hunger is the difference between values 7 and 8, then it seems nearly unimaginable that someone could ever reach a utility value of 30.

Many Consequentialist authors support the idea of "prioritizing" the worst off (Hooker 26). However, taking stock of diminishing marginal utility, we can see that the worst off is quite often the one who can most easily be helped, and a direct form of AC which seeks to allocate well-being as efficiently as possible will organically give precedence to the worst off in a society. Indeed this natural priority given to the worst off is one of the major reasons for the objection that AC cannot give priority to our special relations.

Diminishing marginal utility appears to do much to allay fears of harmful tradeoffs within AC, but is it an accurate concept? It certainly seems as if there are examples where giving a dollar to a billionaire provides her with more well-being than giving a dollar to a homeless person. Often extremely rich people will fervently seek out small deals where they can save a dollar, and be very happy for it, whereas a homeless person spends their money recklessly when they do have it, and gets little joy out of saving a dollar on everyday items.

There are many ways to explain this difference. It seems in both cases the subjects just do not understand the relation of money to what money can buy, albeit in opposite ways. The rich person often fetishizes money, seeking it for its own sake, without any concept of its relation to practical well-being. Indeed, this attitude often goes a long way in explaining how they became rich. In the case of the poor person,

he suffers from the opposing viewpoint. He underestimates the importance of money, and wastes it frivolously. This again can explain his destitute position.

The concept of diminishing marginal utility remains unscathed. It is still true that if you provided the poor person with a dollar's *worth* of needed resources it would give him far greater pleasure than the rich person would receive with a dollar's *worth* of resources.

Diminishing marginal utility cannot fully dispel fears about harmful tradeoffs. Although it explains how trivial pleasures matter far less than serious harms, the two still occupy the same scale and can be traded one for the other. Diminishing marginal utility does however demonstrate how drastically separated the utility quantities would be between a serious harm and a trivial pleasure. I prefer a conception like that of Popper's where benefits and harms cannot be compared, but still a relatively trivial harm, a mere bother such as being spat on, must be on the same scale as a very serious harm, like torture.

Given the concept of "diminishing marginal utility," what would such a tradeoff look like? I think we can take the concept to the degree where such a tradeoff would never practically occur in the real world. Say we assess the disutility of torture at negative one trillion and the disutility of being spat on at negative one. This vast difference seems to capture the vast difference our intuition perceives in such harms.

However, unlike our intuition, AC will advocate for a tradeoff between a single instance of torture and one trillion and one instances of different people being spat on (isolating each instance so there are no compounding effects). Of course

such a tradeoff can never actually occur since there are not that many people in the world. Still it seems cruel of AC to entertain condoning torture, even if the tradeoff could never realistically take place.

At this point I think we can see this as an example of such a bizarre case that our intuition can no longer be taken as a reliable guide. Perhaps the idea of a tradeoff here seems cruel only because our minds can entertain a maximum of a hundred people being spat on, so it seems like we are trading only a hundred for a case of torture.

Although we have already seen the difficulty in allowing certain intuitions while disregarding others, I believe it is not an arbitrary judgment to say our minds cannot fathom an offense to a million (let alone a million million) people. Our intuitions evolved among tribes of a few dozen and it is those circumstances to which they are suited.

Therefore, if we use diminishing marginal utility to stipulate such a large difference between these offenses, we can see that objectors to AC on the lines of tradeoffs only use intuition in examples where our intuition truly can no longer serve us. It is a good example of what Hare calls the “commonest trick” (Hare *Theory* 123).

It is perhaps worth noting that objectors to tradeoffs in AC might complain that I tip the scales in favour of AC by stipulating such a large quantitative difference between torture and being spat on. However, objectors who find such a tradeoff impermissible because one is so much worse than the other, cannot complain that AC should not treat one as so much worse than the other in quantitative terms.

A last strategy AC could use to dispel fears of harmful tradeoffs would be to simply include respect for rights within the maximand. Sen puts forward such an idea with his “goal rights system” (Sen *Rights* 187). In this scheme we seek to minimize rights violations. However, as Nozick points out, this does nothing to solve the problem that rights are always derivative within an impersonal maximizing system. What matters for my morality is not that rights are not violated but that *I* do not violate anyone’s rights (Nozick 30). Under Sen’s “goal rights system” we could be called on to violate someone’s rights if this meant the rights of others were saved from violation by others.

Nozick opposes this possibility because it would infringe on the autonomy of those whose rights we violate. Although this violation will lead to others’ rights being upheld, it still treats a person as a mere means. He advocates following Kant’s underlying principle, which states “that individuals are ends and not merely means; they may not be sacrificed or used for the achieving of other ends without their consent” (Nozick 31). However, as Nozick readily admits, it is not entirely clear what exactly it means to treat another merely as a means. “In getting pleasure from seeing an attractive person go by, does one use the other solely as a means?” (Nozick 32). He provides no answer but simply specifies that he is primarily concerned with preventing physical aggression against others. In this context, it is unclear why we should give autonomy total precedence over well-being, to the point where they cannot be compared or traded off.

Furthermore Sen provides some advantages to his rights-based Consequentialism, as opposed to a strict deontology that espouses side constraints.

Firstly, in Sen's system we allow for the possibility that rights can be overridden by other benefits if they are sufficiently strong. This would allow us to switch the tracks in the Trolley example and save the five, which most respondents feel is right.

Secondly, it imposes a duty on any "third parties" who can prevent somebody from unjustly violating the right of another (Sen *Rights* 222). To see this as a duty also suits our intuition.

Whatever the merits of Sen's conception I think we can show that if we take even a purely hedonic utilitarian analysis seriously, keeping in mind diminishing marginal utility, we can see that this will very rarely call for the violation of rights and it does so only when our common sense would agree or cannot frame the situation effectively, and so the disagreement with intuition is not a crucial weakness.

To conclude, AC has many strategies to deal with the objection that it lacks respect for justice, in regards to tradeoffs, just deserts, and fair distribution. The argument from expectation effects and Mill's stress on the importance of security show that justice will be given a very high value within a Consequentialist system. This covers punishment, rewards and respect of rights. Indeed it seems that all talk of justice in regards to these issues leans heavily on the idea of utility. Still Rawls and Nozick, among others, object that giving a high value to just aims is not enough for us to see AC as respecting justice. It must not treat rights derivatively, and so must respect their inviolability. However, it seems any plausible moral theory will not take rights to be completely inviolable. If we take into account diminishing marginal utility, then it can be shown that AC will only call for trading off rights in

very extreme scenarios, where either our common sense would agree with the tradeoff or our common sense cannot arrive an accurate judgment because of the oddity of the scenario. Diminishing marginal utility might also dictate a fair distribution of utility, one that prioritizes the worst off, since the most efficient way of distributing utility will naturally prioritize those who are worst off.

Chapter 5: Integrity

The integrity objection claims that Consequentialism will undermine our integrity as moral agents by asking us to abandon our personal commitments, or at least to treat them as one set of commitments among many, without any special significance. As many authors note, the precise content of this objection “is a matter of some interpretive dispute” (Dorsey 7). However most agree that the objection is most explicitly put by Bernard Williams in his rebuttal against J.C.C. Smart in *Utilitarianism For and Against* (Williams 77).

For the purposes of this chapter I will treat the objection as making three interrelated claims. Firstly, AC does not properly explain how we treat others as integral moral patients. AC, it is said, suppresses our feeling that we should care more for the persons who are closest to us. Because AC demands impartial treatment we cannot express this valuable human sentiment.²⁵ Secondly, AC betrays our sense that certain moral features are taken at a fundamental level. Instead of say, expressing the fundamental wrongness of killing, AC only condemns killing via its overall effect on well-being. In this way, AC is said to treat our most fundamental moral concerns as merely instrumental or derivative (Brink *Forms* 413). Thirdly, it is argued that an AC agent will have “one thought too many” in contemplating moral questions (Williams *Luck* 18). Because AC seeks to explain its moral decision-making by reference to well-being promotion, it is said that AC distorts and overcomplicates the processes involved in moral thought.

²⁵ I discuss this side of the integrity objection in the chapter on Personal Relations.

In order to look at this objection I will closely examine the two examples Williams goes through in his essay in *Utilitarianism For and Against*. I hope to show that for several reasons these examples do not (as he maintains) yield counter-intuitive results when put through an AC analysis. Firstly, I highlight the real world complexities of these examples in an attempt to contradict the “obviousness” of AC’s verdict. Then I look at the psychological harm the agent endures in acting against his integrity and how this relates to the overall AC calculus. Next, and in a related sense, I consider how an AC agent should regard his own emotions in cases where he has to perform a detestable task in the name of the greater good. Then I look at the idea that each of us is especially responsible for what he and he alone does, and what bearing this has on the examples. Finally, for these examples, I examine the role of the prisoner’s dilemma and demonstration effects in complicating the AC verdict. From all these considerations, I believe it can be shown that the integrity side of the question about these examples does not gain much traction.

Next I plan to argue for the circularity of the integrity objection and how it cannot independently demonstrate the weakness of AC since it already assumes this weakness. In this respect, the integrity objection could be as convincingly leveled against competing moral theories (such as Deontology) in much the same way.

In this context, then, I will examine why the objection seems especially pertinent in regards to AC for five (mistaken) reasons: the impartiality of AC, Sidgwick’s government-house utilitarianism, the overdemandingness of optimific AC, the pursuit of well-being as an abstract, foreign ideal; and the over-complexity of AC moral thought.

Williams' two examples run as follows (Williams *Critique* 97). In the first example, George, a chemist, badly needs a job to properly care for his family. He has just been offered employment with a chemical weapons firm. If he does not accept the job he knows that someone else, indeed someone with overabundant zeal, will take the job instead. George objects to working for a chemical weapons firm but AC morality seems to demand that he should take the job. In this way, Williams says, AC asks us to abandon our firmly held moral commitments.

In the second example, Jim is in a South American town where he is detained by a military captain, Pedro. Pedro has rounded up twenty Indian protesters. He is about to kill them all but offers Jim the chance to kill one instead, and the other nineteen will be spared. Jim, like the rest of us, is opposed to killing but AC demands that he does the least harm, and kill the one to save the nineteen.

Williams objects to how AC sees the choice as "obvious" in either case. In George's case we might think taking the job is in fact the wrong course of action. In Jim's case, we might feel compelled to kill one to save nineteen, but we should expect it to be a difficult thing to overcome our moral compunction against killing, and not as clear cut as AC indicates. Furthermore, Williams objects to the way that AC arrives at its answer. We should not see ourselves as merely operating global "causal levers," which would undermine our integrity, but as individual agents with internally competing personal commitments, i.e. caring for our family vs. refusing employment in a harmful industry, or refusing to kill vs. allowing others to die (Williams *Critique* 82)

Does AC actually require an agent to act in the way Williams thinks in these cases, and with such ease and simple deliberation?

One possible response AC can make is to highlight the real world complexities of such scenarios. In the case of George, for example, it might be possible that another job comes along soon after, which would not cause him such emotional stress. Many other possible effects could be highlighted as well.

Williams objects to this response because, as he says, “the certainty that attaches to these hypotheses about possible effects is usually pretty low; in some cases, indeed, the hypothesis invoked is so implausible that it would scarcely pass if it were not being used to deliver the respectable moral answer” (Williams *Critique* 101). The objection is well put. AC justifications can easily fall into this question-begging trap, where the so-called justifying calculation is manipulated in such a way as to suit the moral answer, which has already been arrived at through our intuition.

However, if we take seriously “expected AC,” where difficult probabilities must be assessed and compared, then it must be admitted that many remoter effects will have to be countenanced in moral decision-making. Of course these probabilities cannot be skewed to fit the moral answer, but they still must be included in the overall calculation.

A second way AC can respond is by highlighting the harm done to the agent in acting against his own commitments. For George, taking the job will result in a guilty feeling that counts against his well-being. If this negative feeling is strong enough, taking the job will result in an overall loss of well-being, and so AC will advocate against it.

Williams objects to this maneuver for two reasons. Firstly, as with the previous response, he notices a tendency for AC apologists to overestimate the intensity of this feeling, so that an AC calculation will deliver the already apprehended correct moral response. Such maneuvers can only amount to a “finger in the dike” solution to the integrity objection. However, as Williams notes, in the case of Jim this effect of guilt will have little weight compared to the great loss of well-being at stake, while in the case of George the guilty feeling might tip the scales. This requires no overestimate on the part of AC, which already sees Jim’s case as a clear moral choice, but sees George’s as a more difficult one.

Williams has a second objection against giving status to our own guilty feeling in cases like those of Jim or George. Since the AC calculus showed that one should undertake the action, it is unclear why either of the men should feel guilty about acting in the way AC deems is right. “If a course of action is, before taking these sorts of feelings into account, utilitarianly preferable, then bad feelings about that kind of action will be from a utilitarian point of view irrational” (Williams *Critique* 104). If this guilt does exist, and AC is a correct moral theory, then this guilt must be based on some sort of “confusion,” as Williams puts it (Williams *Critique* 101).

However, whether or not the agent’s feeling is perfectly rational, what matters for AC is simply that it exists. It is often considered a pragmatic strength of AC, as opposed to theories like RC, that it is able take things as they are.²⁶ Williams recognizes this but still contends “that the utilitarian could reasonably say that such

²⁶ I discuss the negative side of this aspect of AC in the justice chapter p. 76-79.

feelings should not be encouraged, even if we accept their existence, and that to give them weight is to encourage them” (Williams *Critique* 106). But the question of encouragement is really besides the point here. AC takes all emotions, vicious or virtuous, as they are and adds them to an overall calculation. In so doing, AC has the advantage of avoiding any unfair bias in favour of certain kinds of desires and emotions, particularly those desires shared by the acting agent.²⁷ On the other hand, the inclusion of sadistic desires has led to a deep criticism of AC, one Williams makes in the same essay (105). If AC must accept this criticism then it should not also have to accept any criticism based on the idea that it cannot include, and therefore promote, such anti-utilitarian feelings or desires.

Williams also objects to how the AC agent “should regard his feeling just as unpleasant experiences of his, and he cannot, by doing that, answer the question they pose when they are precisely not so regarded, but are regarded as indications of what he thinks is right and wrong” (Williams *Critique* 103). In response to this, AC can say that in the majority of cases our moral feelings do have Consequentialist underpinnings, and so they can indeed be taken as indications of what is right and wrong, from a Consequentialist standard.²⁸ For the most part, doing the right thing will be associated with a pleasurable feeling at having done the right thing, reinforcing its Consequentialist worth. However, in the highly uncommon scenarios presented to Jim or George, our moral emotions run contrary to the dictates of AC.

²⁷ For example, if an agent receives no joy from music, he will unfairly downgrade the joy others receive from music, possibly to their detriment. If, on the other hand, he must accept as legitimate the pleasure of a joy he does not share or even understand, he cannot dismiss this value and cause them to suffer for it.

²⁸ see chapters on personal relations, justice

Considering the many intricacies often involved in moral thinking, it should come as no surprise that in certain corner cases, our moral emotion might find difficulties with what actually ought to be done. As mentioned in the previous chapter, Hare calls this type of argumentation the “commonest trick,” where anti-AC critics illustrate extremely far-fetched scenarios, and then criticize AC based on its disagreement with our intuition in this scenario (Hare *Theory* 123). As an analogy, notice how our sight is often a useful guide for telling us whether or not we are moving. However, on certain rare occasions, such as when we are reading in a car, our sight tells us, wrongly, that we are sitting motionless. Simply because our sight informs us incorrectly on this one rare occasion, should not totally undermine the usefulness of our sight for this purpose.

It is, then, somewhat misleading how Williams labels such difficulties as a “confusion” on the part of an AC agent, to have these feelings in cases where they run contrary to AC. They are very valuable feelings to have on the whole, and will often correctly guide our moral thinking, but could still be mistaken. One might ask whether it would be a “confusion” on the part of a thirsty man to crave ocean water. He is only exercising an important instinct, an instinct that will usually serve him well but which his reason should suppress or ignore in certain rare scenarios. In a certain sense this is a “confusion” on his part, but not in the sense that we should suppress and act against this instinct in all cases (as Williams seems to be saying about AC’s treatment of our moral commitments).

It can be considered a strength of AC that it allows moral reasoning to intervene against our emotions. In giving supremacy to the integrity of our

emotions, Williams must struggle to explain how and when to suppress certain vicious emotions, albeit ones which are felt with integrity. Should we support an agent in attacking someone who annoys him, simply because he feels this desire with perfect integrity? Of course, Williams might think a good moral agent would not have these feelings, or would have competing claims that equally admit of integrity (Williams *Luck* 13). However, whereas AC will approve of these “valid” claims, and explain why they are valid, “integrity” may equally reinforce either morally “valid” or “invalid” claims.

To a degree Williams anticipates this response. Particularly in the case of Jim, if he let his feeling to not kill take precedence, and thus allowing the same plus nineteen to die, he could be blamed for “a kind of self-indulgent squeamishness” (Williams *Critique* 103). He admits this is “a familiar, and powerful, weapon of utilitarianism” (Williams *Critique* 103). However, he maintains that this accusation of squeamishness “will not itself answer the (moral) question, or even help answer it,” because dismissing this emotion only serves to “alienate one from one’s moral feelings” (Williams *Critique* 104). I hope, though, that it has already been shown how in certain bizarre scenarios it is indeed morally best to suppress one’s moral feelings.

The idea of squeamishness gets to the heart of what is going on in the Jim case. I think we would feel somewhat odd in blaming someone in Jim’s scenario when he chooses not to kill. After all, Jim would be expressing a caring and humane feeling in refusing to kill, and it should be extremely difficult for him to go against these feelings. Still, given the large number of people that would be saved, it seems

to me, despite Williams' reservations, that Jim ought to kill one to save nineteen. And, as we increase the number of lives saved in the scenario, it seems that other moral theories, even the most tough-minded deontology, will have to at some point face a serious dilemma, either to allow that in some cases killing must be permitted or to embrace total implausibility and insist that we should not kill, even to save seven billion.

It seems that Williams' objection mistakes the obviousness of a decision with the ease of actually going through with it. The obviously correct choice might not be easily undertaken, and this is what we observe in the Jim example. If we feel that refusing to kill is not exactly deserving of blame, then AC, in its scalar form, could accommodate this belief. As outlined in the chapter on overdemandingness, AC could plausibly accept a scalar criterion where we judge acts as better or worse, according to their well-being promotion. Acting optimally can be so difficult that we can understand why someone might fail to meet this standard. In understanding this shortcoming, we would not see their actions as totally blameworthy. Often we cannot convincingly declare we would have acted differently. So, in the case of Jim, we could see his decision not to kill as an understandable moral shortcoming.²⁹

The problem of "dirty hands" is a familiar one in moral and political philosophy. The term was introduced by Michael Walzer in his *Political Action: The Problem of Dirty Hands*. It refers to the (supposed) need, particularly in political action, to commit a wicked (or, at least, reprehensible) act in order to prevent

²⁹ Incidentally, the numbers in this case are so large that I think we would find this decision an example of blameworthy "squeamishness," but as the numbers saved are lowered, his moral shortcoming becomes more and more one we can understand and allow for.

greater harm. Although his position changes on the matter, in the end Walzer felt that “dirty hands” were only right in circumstances of “supreme emergency,” such as civilian bombing in the context of the Second World War (Coady). Max Weber, on the other hand, followed broadly AC lines in finding such acts necessary whenever it prevented more harm (Coady). In all cases though, an apparent paradox arises since an act will appear both morally wrong and morally necessary. In relation to Williams’ examples, the feeling associated with an act, in rare circumstances, must be divorced from the considered judgment of that act. As Kai Nelson puts it, “feeling guilty is not to be confused with being guilty” (Nielson 140).

Williams makes a last complaint against AC in these cases in that it undermines our integrity by cutting out “the consideration (...) that each of us is specially responsible for what he does, rather than for what other people do” (Williams *Critique* 99). However AC does not see us as responsible for the actions of others, but only responsible for intervening when this is on the whole beneficial. And this only suits our common sense, since we do observe a moral pull of intervening in the detrimental actions of others. As the Jim case readily shows, we feel morally compelled to curb Pedro’s awful deeds, because we are in a special position to do so. Any system that entirely neglects this moral pull could be accused of squeamishness, or cowardice, verging on selfish immorality.³⁰

³⁰ Jim’s refusal need not be motivated by cowardice, however. Refusing to indulge Pedro’s sick wish, and so defying a murder, could be motivated by brave defiance. Still, in considering the results of this defiance (twenty deaths, or, if we modify the example, many more) we must consider it an immoral act, if not motivated by cowardice, then perhaps stubbornness or callousness.

However, we must consider the value of integrity when the stakes are much lower. For example, how should integrity come into play when we must decide between breaking a trivial promise and causing another to break five equally trivial promises? Although the harm levels are relatively minor, AC still seems to indicate we should break our own promise to save five others. But it seems as if in this case there is a greater pull to maintain our integrity and keep the promise we have made. Does this suggest integrity has value outside of well-being promotion?

To avoid this conclusion one could accept that we should indeed keep our own promise, but maintain that the reasons are still ACist. There is, perhaps, ACist value in holding true to our integrity, and seeing it as an independent value, even if this is not the case. We might all take our promises to be weightier if we see them as our special concerns. The promises belong to us (they are not, so to speak, rented) and we should take special care of them. However, by the lights of AC, we must regard this as a commitment to an illusion, a self-suppression that raises its own worries. Perhaps then, we can fall back on Scalar Consequentialism once again to explain our approval of the promise-keeper. To keep the promise, given the low levels of harm involved, is not seen as below the level of wrongness, in fact we might admire the pride of the promise-keeper, but this admiration is for extra-moral reasons, and we can still believe that it would be morally better if he did break the promise to save five others, thus sacrificing his own pride for the greater good.

In the end, AC seems vindicated for Jim's case. AC provides the correct answer, and the fact that it makes this correct answer "obvious" (but will undertake the difficult task with a heavy heart) should not count against it.

However, in the case of George, it can be argued that AC delivers the wrong answer. In this case, involving a difficult moral choice, Williams appears partially justified in criticizing the “obviousness” of the AC decision. Still, as already noted this obviousness is slightly moderated by the negative psychological effect (rational and justified or not) on George, as well as subtle considerations of practical factors (like the chance of another job opportunity arising for George in the near future). In fact the right answer is thrown into confusion when we consider it as a prisoner’s dilemma type example. George will be in the free-rider position if he takes the job. He will receive the personal benefits of the job, while taking advantage of the reduced competition offered by others who will not take the job on moral grounds (since it is in an industry that might cause great harm). As we have seen in the chapter on expectation effects, AC is thought to suffer from problems of coordination, where a choice that on an individual level leads to well-being promotion, is actually detrimental to well-being when universalized. AC is not the only moral system that has difficulty dealing with prisoner’s dilemma type paradoxes. We tend not to blame someone for wishing to avoid the “sucker” position, where they are cooperating with non-cooperators.

AC would also object to George’s decision to take the job because of the harmful demonstration effects it causes. If George took the job other scientists might see it as an endorsement of the industry, causing other scientists to take similar offers, and allowing the industry to flourish. Conversely, his refusal would also entail such exponential effects. This assumes that George’s decision would be made public.

George would have to count these harmful consequences against his decision to take the job.

As I attempted to show in the chapter on expectation effects, an act's contribution to many people acting in this way should be approximated at in the act of a single person, and the chance of this should count for or against this individual act.³¹ Williams criticizes AC for making moves like this because it is very difficult to assess what values to assign to remote probabilities (Williams *Critique* 100). Now, as previously noted, this criticism is valid in cases where AC supporters make use of this confusion in order to push through an ACist decision that conforms to our common sense. However, just because this complexity can be manipulated, does not mean we should pretend these complex factors do not exist. Dishonest accounting need not occur in these cases and the fact that AC legitimately encounters cases involving very difficult analysis should not count against it per se. Indeed, we should expect nothing else in cases that involve complex prisoner's dilemma type scenarios as well as demonstration effects, as we do with George.

If we now reexamine the George scenario in this light, it can be shown that the situation is not as clear or "obvious" for AC as Williams would have us believe. Against the needs of his family, George must also count his contribution to a possible situation wherein chemists enter any industry, regardless of the harm that industry causes. Furthermore, the immediate needs of his family are softened by the chances of another, less detestable, form of employment arising in the near future. If we

³¹ Page 47-49.

combine all these factors, it can be seen that George's choice is not so obvious by the lights of AC, and it may indeed advocate that he should not take the job.

In examining both these scenarios, the George and the Jim case, by a more refined version of AC, it might not be the case that they shed doubt on AC. In Jim's case, AC does indeed advocate the same decision as our common sense morality, and does so in a way that understands our every day aversion to such acts (such as killing). In George's case, it is difficult to determine what AC would require. This suits our intuition, which finds it equally difficult to arrive at a decision under these difficult circumstances. We might blame George for selfishly neglecting the wider society if he decides to take the job, or we might blame him for inhumanly neglecting the needs of his family if he refuses the job. The problems AC encounters in George's case are not at all unlike the problems our intuition encounters. As such, Williams' use of these examples does not shine a light on the weaknesses of AC. In fact, his appeals to the importance of integrity, could yield counter-intuitive results, as by allowing George to refuse the job without taking into account the pain of his family, or by allowing "squeamish" and cowardly Jim to walk away, allowing nineteen people to be unnecessarily murdered. We would observe negative aspects in these actions (particularly Jim's), the same negative aspects that AC observes. These negative aspects persist, whether or not the agent acts with integrity. So, in the end, these cases do not reveal any weakness in AC but do show how taking integrity as paramount could result in problems.

I have looked at these examples to show how Williams' analysis unfairly characterizes the AC suggestion as wildly opposed to our intuitions. The appeal to integrity adds little to the idea that AC is counterintuitive.

Dale Dorsey, along these lines, suggests that the integrity argument against Consequentialism is wildly question begging (Dorsey 26). Dorsey defines the integrity objection as the idea that "Consequentialism requires agents to perform actions, that, plausibly speaking, agents lack decisive practical reasons to perform" (Dorsey 1). However, the appeal to integrity is not just an appeal to selfishness. It should be presumed that the moral action will not (at least some of the time) be the most personally gratifying action available. As I have already noted in discussing Williams' examples, an agent will include among his cherished commitments the commitment to act morally (most clearly seen in the George situation). Indeed authors who have taken up the integrity objection put forward the idea that morality ought to be an overriding personal commitment (Stroud 185).

Herein lies the circularity of the integrity objection. For if morality ought to be an overriding personal commitment then the demands of AC, given that it is a true and accurate moral theory, ought to override our other personal goals. A good moral agent will treat these demands as one of his own personal commitments. Of course, Williams and other objectors do not accept that AC is a true and accurate moral theory (as the thought experiments, particularly George's, are meant to illustrate). However if the integrity objection against AC only gains traction once it is assumed that AC is not a correct moral theory, then the objection itself does nothing

to demonstrate the falsity of AC, but is subsumed under all the other objections against AC.

In fact, it appears there is nothing especially anti-AC about the integrity objection, and it can be leveled against other theories. If, for example, we have already assumed the falsity of a strict deontology that prohibits us from killing, then it could be seen as an attack on Jim's integrity to ask him to consider this rule that, as a good moral agent, he does not include among his personal commitments. But the assertion that it damages his integrity does not add anything to the argument. It simply repeats the assertion that this is not a true moral rule, and a good moral agent should not count it among his commitments.

Why then does the integrity objection seem especially relevant for AC?

Firstly, because AC stresses impartiality it is thought that it undermines our relation to our own values, projects, and commitments (Shaw 275). AC is impartial in two senses. Firstly it says that agents are evaluated impartially, regardless of who the agent is. It is wrong for Jim to steal just as, and for the same reasons, that it is wrong for George to steal. This kind of impartiality is endorsed by most moral theories. Most deontological theories, for example, say that, for all people, it is wrong to murder.

The second kind of impartiality has to do with how the theory views other moral patients. AC views all persons as equally worthy and deserving of well-being. Each is to count for one, nobody for more than one. Most moral theories disagree with this. Scheffler, for example, advocates for a form of agent-centered morality where one ought to care more for a loved one than for a perfect stranger (Scheffler

Restrictions 6). In the chapter on Personal Relations, I look more deeply at this question, and attempt to show that, under AC, the greater benefit a loved one receives from *your* attention directs your attention towards that loved one. In this way, AC can accommodate our belief in caring for our loved ones, while upholding the idea that we are all equal moral patients. The special relations question can be considered a sub-branch of the integrity question in this way.

A second reason for the connection between AC and the integrity objection has to do with the Sidgwickian promotion of a form of government-house utilitarianism that has led to concern that Consequentialism “is not a theory of the rational authoritativeness of its moral requirements” (Hurley 25). If this concern is valid then AC might not see its precepts as reasons for action, or as Williams puts it, as personal commitments. If AC does not provide personal commitments then following AC might threaten the commitments we do have, and our integrity.

However, if the government-house form of AC does not provide moral commitments then there is no reason to see it as violating the commitments we do have. Either it provides a countervailing personal commitment, or it leaves our commitments unopposed and our integrity intact. However, it appears to me that AC does indeed provide its own personal commitments and Hurley’s concern is based on a misconception of government-house utilitarianism. Simply because this theory allows room for the imperfections of humans in order to promote the good does not

mean that it promotes these imperfections. An *ideal* agent, under government house utilitarianism, will follow AC directly and take its precepts as rationally overriding.³²

The third reason the integrity objection attaches itself to AC has to do with the perceived overdemandingness of AC. Indeed, Dorsey sees the integrity objection as a branch of the overdemandingness objection (Dorsey 9). Because AC is thought to demand perfect maximization of well-being, and bringing about perfect maximization will entail tremendous sacrifice, often of an agent's deepest personal commitments, it is thought that AC undermines our integrity. However, the scalar form of AC (argued for in the chapter on overdemandingness) does not draw the line of immoral behavior at perfect maximization. Under this system, we could find a person's behavior moral and praiseworthy because of the amount of well-being it brings about, even though this amount may fall short of the maximum. Of course promoting more well-being will always be considered more morally praiseworthy, so AC morality will surely be in support of this. However, just because AC says that acting in a perfectly moral way will entail tremendous sacrifice does not mean AC demands that we abandon our deepest personal commitments. After all, it seems intuitively plausible that acting with the utmost morality will require great personal sacrifice.

A fourth reason the integrity objection relates to AC has to do with how AC views well-being promotion as the ultimate end, with all other good deeds serving only as means to this end. It will undermine integrity to treat one's morality simply

³² And in any case, a plausible form of AC should eschew the notion of government-house utilitarianism as discussed in the section on expectation effects p. 53-59.

as operating “global causal levers” that seek only to promote some bizarre abstract ideal. This side of the objection has been put in various ways by different authors.

Bernard Williams, for example, objects that AC does not respect the “seperateness of persons” when it puts stress on the impartial view (Williams *Luck* 3). AC, he maintains, treats everyone as perfectly equal parts in an overall pool of well-being. It betrays our genuine moral thinking to construct a theory where everyone, be it your wife or a perfect stranger, is “inter-substitutable” (Williams *Luck* 15). He uses an example, borrowed from Charles Fried, in which an agent has a choice of saving one of two people, and one of them is his wife. In such a case, Williams and a sophisticated form of AC agree: the agent should save his wife. However Williams objects to *how* AC reaches such a conclusion. As he says, “the consideration that it was his wife is certainly, for instance, an explanation which should silence comment” (Williams *Luck* 18). In other words, the special relationship alone should suffice as a moral claim, without any underlying justification required. Indeed, invoking any underlying justification (as in AC) “provides the agent with one thought too many” (Williams *Luck* 18), and evinces the inhuman cruelty of which AC is often accused.

It could be said that the reasoning involved in AC justification of concern for special relations focuses on “extrinsic,” “derivative” “instrumental” or “secondary” reasons (Brink *Forms* 413). It is wrong to search for “instrumental” reasons for whether or not we should help our loved ones, but instead should focus on the immediate intrinsic reasons that say we should help a loved one simply because they are our loved one, without any reference to external, derivative utility analysis.

J.O. Urmson provides an example of secondary and primary reasons for moral action (Urmson 114). We can cut a wire to prevent a person from being electrocuted. We attach no special moral weight to cutting a wire, but we do to saving a life. So the saving of a life is said to be primary and the cutting of a wire secondary. However, for AC, we only cut the wire to save a life *to raise overall well-being*. Only well-being promotion is a primary moral reason. So for AC there is “no difference in the moral status of wire cutting or life saving” (Urmson 114). However there is a crucial difference. Saving a life will very often promote well-being greatly, cutting a wire almost never will. So, in terms of expected AC, the saving of a life has much greater moral significance. Urmson comes to recognize this – there is a difference in “the frequency with which (these acts) have secondary moral significance” – however he does not see this as an important difference (Urmson 115). The reason this is not an important difference for Urmson is because, in the end, AC will always judge acts solely in respect to their relation to well-being promotion. This means that an AC agent would be “prepared to lie, rape, steal and betray to achieve even a minimal increase in general utility” (Urmson 115). This is put quite unfairly. After all, for an AC agent to do such a thing will require a tremendously horrific result of not doing so. Urmson fails to highlight this. Indeed, once there is such a prospect, the deontologist could be accused of callousness for refusing to intervene in the face of such horrific consequences.

Objectors on the grounds of integrity often seem to neglect the importance of the well-being that is being promoted in AC. They treat the AC agent like one who tries to uphold some bizarre ideal without any thought towards his fellow man. An

analogy can be made between the AC agent, as Urmson sees him, and certain religious zealots, who help others, not out of any direct feeling of love or beneficence towards them, but only in order to follow the orders of their God (Kofman November 6). To see matters in this way is to misunderstand AC. It does not seek to fulfill some bizarre abstract ideal. Rather, that ideal, well-being, should be thought of as the very essence of what is good for mankind. Urmson misunderstands this in the same vein as someone who insists they hate money and wealth, but love jewelry and expensive clothing. Except, in the case of well-being, it never has to be cashed in for human welfare, it just *is* that welfare.

When we consider that AC seeks to promote, not some foreign inhuman entity, but precisely human well-being (however we decide to define it) then the criticism about the “derivativeness” of AC loses much of its force. Imagine if a wife were to complain to her partner, “You do not care about me. You only care about my well-being.” It would be hard to imagine what she could mean. True, AC demands that we consider each person equally, without seeing anyone as more worthy of well-being. But this does not entail that we should casually neglect our special relations, if we keep in mind that we are in a special position to add to their well-being, as I have tried to show in the chapter on Personal Relations. On the other hand, it does show that, under certain extreme circumstances, we should neglect our family members in order to bring about more overall well-being. But this is a fact that any sensitive moral agent should appreciate. Indeed Williams himself, believes it is absurd “that if there is some friendship with which his life is much involved, then a man must prefer any possible demand of that over other, impartial, moral

demands” (Williams *Luck* 17). Here the question begging of the integrity objection against AC presents itself again: we should consider important moral concerns with integrity, however AC does not provide important moral concerns, and so they will not be considered with integrity, and therefore AC does not provide important moral concerns.

A final way the integrity objection could be interpreted is by focusing on Williams’ idea that AC involves “one thought too many.” It seems that AC does ask for a strange sort of moral thinking that breaks down into strange, over complex, deliberations about well-being promotion. Indeed, it should be agreed that there is something overly mechanistic, verging on cold-heartedness, in an agent who seeks to add justifications to their decision to save their wife. However, while an agent should not need to seek justifications in this case, that alone does not show these justifications to be incorrect. Morality is a human endeavor, and like many such endeavors, such as aesthetics or comedy, we see something wrong when someone tries to break down their reasoning instead of just living in the moment. Consider the case of beauty. I may find someone beautiful, and this is enough. It takes me out of the human sphere to overanalyze why I find someone beautiful. If I were to subject my assessment to further analyses, such as considerations of how this person fits into my biological imperative to procreate, along with cultural and psychological factors, then I could indeed be accused of having quite a few thoughts too many. Inadequate as these considerations are in an instinctive judgment, that alone does not mean this analysis is false. Consider also the case of a funny joke. We are right to object to “stuffed shirts” who have to rigorously justify the elements

involved in a joke in order to find it funny, and indeed when a joke has to be explained it often loses its spontaneous charm.³³ Again, though, our disdain for this disconnectedness does not show the justifications to be untrue, but instead shows an inappropriate, paralyzing over-analysis that deprives us of the immediacy required for such human endeavors. Returning to the case of saving one's wife or a perfect stranger, we can agree with Williams that undergoing a rigorous utilitarian analysis would be inappropriate in such a case, removing us too far from the intuitive passions of our lives. However, we could say that this justifying is correct, but that we understand it intuitively and instantly, and this is the reason why we object to someone who proceeds differently. Indeed in the case of morality this type of thinking seems especially out of place. We make many obvious moral choices in the course of a day. If an agent felt compelled to subject every single moral choice to a rigorous analysis according to a moral theory (however true this theory may be) the result would lead to a paralysis of action and perhaps even many mistakes.

Occasionally, however, our intuitive moral thinking runs into problems (related, perhaps, to a deep unshakeable bias) and we are unable to clearly see what we ought to do. After all, it is common enough that someone does an awful thing, only because they thought they were acting morally. When such mistakes occur, then it is appropriate for an agent to slow down the process and submit the question to the rigorous theoretical analysis AC allows for. This approach was advocated for

³³ Another example would be that of walking: we instinctively learn to walk with ease without actively considering the intricate elements of motion, balance, etc. even though these components do play a part and are instinctively taken into account. When someone undergoes trauma and has to "relearn" to walk, then they would have to actively consider these elements.

by Hare with his distinction of “the archangel and the prole” (Hare *Method* 45).³⁴ Although the need for this analysis might not arise very often, it is still a strength of AC that it allows us to reevaluate our moral thinking by an independent standard, namely the overall promotion of well-being. This should not interfere with the notion that “the morality of common sense is at least unconsciously utilitarian” (Sidgwick 424).

In conclusion, it appears the integrity objection against AC suffers from circular logic. Either integrity permits all acts that the agent truly takes to be a personal commitment (regardless of the act’s immorality) or integrity will have to include within it the commitments of an overriding morality. Williams, rightly, does not accept the former possibility. If he accepts the latter possibility, then for him to claim that AC does not provide an overriding morality he must have already assumed its falsity as a moral theory. Therefore the integrity objection does not demonstrate the falsity of AC, but already assumes it. There are, however, special reasons for which the integrity objection attaches itself to AC in particular, as opposed to other moral theories. It seems, though, that either these reasons are based on a confusion or they do not, under investigation, legitimately raise concerns over AC’s threat to our integrity. However, the impartial treatment of our special concerns does seem to conflict with our considered moral commitments. On the other hand, total partiality for these relations also conflicts with our moral commitments, and so AC does correctly identify an important moral feature in these dilemmas.

³⁴ As discussed in the chapter on Expectation Effects p. 55.

Conclusion

In this thesis I have looked at five objections against AC and have attempted to show that in large part they do not pose a serious threat to the theory.

In the case of personal relations I looked at certain unique features of these relations, especially the profound need we have to be helped by loved ones. These unique features explain that cultivating and preserving close personal relations will have a powerful Consequentialist backing. In this way our intuitive sense of personal relations will be “caught in the impersonal net of utility” (Sen *Introduction* 6). It appears that AC is able to explain not just the importance we give to these relationships, but can also explain when we should scale back these sentiments, as when it amounts to a “pernicious prejudice,” like in the case of nepotism (Scheffler *Relationships* 194).

For the overdemandingness objection, I began with a similar method to highlight real world scenarios that show how efficiency of well-being promotion will lend itself to more agreeable forms of beneficence. Insofar as aiding our loved ones will gain a Consequentialist backing, as shown in the previous chapter, complete and impartial beneficence could be made more congenial. However, real world cases of third world suffering, and even the highly inefficient methods of charitable aid available to us, will call for tremendous sacrifice on the part of an optimizing AC agent. This leads to two possible defenses for AC. Firstly, one could maintain that, given the real world inordinate suffering of third world countries, morality must demand tremendous sacrifice, and so the overdemandingness (or, rather, the

demandingness) objection does not amount to an objection at all, but supports the validity of AC (since morality should be very demanding). Still one might object that the degree of sacrifice asked for by AC is still unrealistically overdemanding, and that demanding so much might lead some to give up on morality as a whole. At this point I argued for a scalar form of AC (put forward by Michael Slote) that judges rightness according to promotion of well-being but does not draw the line at total maximization, since to do so would lead to highly counterintuitive moral assessments. In this way AC need not call for an overly extreme moral sacrifice, but still follows the same AC standard (whereby optimization is best). Using this form of Scalar AC, we can both deflect fears of overdemandingness while, by the same standard, explain the praiseworthiness of extreme sacrifice for moral causes.

The next chapter dealt with the expectation effects objection. In relation to promise keeping in particular I tried to counter the idea that the effects will be totally “invisible,” so that AC will not lead to the unforeseen collapse of important institutions, which would result in the theory’s incoherence. However there may also be collateral effects of breaking promises, comparatively “invisible,” that do lead to a possible collapse and therefore incoherence. For promises, it appears that a “thumb in the scales” method, like that attempted in the first chapter on personal relations, will not solve the problem. So instead the solution might lie in focusing on Expected (as opposed to Actual) AC. Such an approach might call for the inclusion of “rules of thumb” in AC. In this way we can meet the expectation effects objection while maintaining an organic approach to AC that does not seek to restrain AC thinking in the way other solutions (such as government-house AC) would. Still, an

over-reliance on Expected AC might lead to counterintuitive results, so it should work alongside an Actual AC form of the theory.

For the justice objection, I undertook a similar approach to the one used in the chapter on expectation effects. I looked at Mill's argument for placing a very strong utilitarian (or ACist) weight on the dictates of justice. Next I examined Rawls and Nozick's arguments for the intrinsic nature of the claims of justice, which undermine Mill's argument for upholding justice due to its instrumental importance within AC. In consideration of arguments from Ross, Kamm and Thomson I attempted to refute these claims by denying the complete inviolability of rights, and by showing that well-being promotion is often a just motivation for certain rights violations. I next considered the unfair distribution that Rawls observes in AC. Although this leads to some lasting concerns for AC, particularly in terms of promoting sadistic pleasures, I go on to show that a thoughtful application of Diminishing Marginal Utility militates against many of the more severe infractions critics see in an ACist unjust distribution. Finally I looked at the possibility of reconsidering the AC maximand to include the upholding of rights, an idea argued for by Sen.

Lastly, for the integrity objection, I analyzed Bernard Williams' chapter in *Utilitarianism For and Against*, which outlines the problem of integrity. In thoroughly going over the two thought experiments he lays out, I argued against the assessments he makes, and attempted to show that they do not demonstrate that AC is counter-intuitive and undermines our sense of moral integrity. From here I went on to argue for the circularity of Williams' integrity objection, since the objection is

based on the idea that AC opposes our moral integrity if we already take AC to be fallacious, which is precisely what the integrity objection is meant to demonstrate in the first place. In this respect the integrity objection pertains no more to AC than it does to other moral theories. I looked at possible reasons, then, for why the integrity objection seems especially applicable to AC.

Through examination of these five objections I hope that a new, more workable and more plausible form of AC emerges. For all the objections, by highlighting the methods of efficient well-being promotion, as well as a fair consideration of how to take well-being as a maximand, it can be shown that an organic form of well-being promotion need not lead us to reject AC. This method contrasts with a more cumbersome and corrective Consequentialist approach like that of Rule-Consequentialism, which seeks to obstruct the unencumbered well-being promotion of AC.

AC is often, unfairly, seen as overly callous or (and?) requiring overabundant beneficence. However, we can avoid this conclusion by highlighting three aspects of AC demonstrated in my arguments against the five objections. Firstly we must consider seriously the goal of AC, which is to promote, not some inhuman ideal but the very human element of well-being. Secondly, it is important to carefully consider how best to efficiently promote well-being, such as through Diminishing Marginal Utility or by understanding the profound effect personal relations have on our well-being. Thirdly, what is morally expected of us will be decided, not according to full well-being maximization, but only by a certain level of well-being promotion (in the context of all available alternatives) that is deemed acceptable. In

highlighting these aspects of AC we can soften the five objections and avoid a crude caricaturizing of the theory.

A crucial question remains, however. Although by highlighting these three aspects we arrive at a theory that accurately describes how to act morally, it remains to be seen, whether or not, through AC, we arrive at these judgments for the right reasons. One might argue, for example, that the reason we should treat people fairly is not explained through some roundabout appeal to well-being promotion but simply because, on a basic level, we should treat people fairly. It is appealing to explain the call to fair treatment (and other moral rules) according to a fundamental principle such as well-being promotion but, as W.D. Ross puts it, “loyalty to the facts is worth more than a symmetrical architectonic or a hastily reached simplicity” (Ross ch. 1).

Now one initial response AC can make is to claim that it is not a coincidence that when we attempt to promote well-being (with a mind to efficiency of promotion, and the human nature of well-being) we arrive at the same judgments our moral intuition observes. Instead, this shows a connection between our moral beliefs and well-being promotion. However, one can always argue against AC justifications, as Williams does in *Utilitarianism For and Against*, that they work backwards from already arrived at intuitive judgments. AC may make use of justifications “so implausible that it would scarcely pass if it were not being used to deliver the respectable moral answer.” There is always a worry that AC justifications fall into this question-begging trap.

Furthermore, demonstrating parallels between the judgments of AC and our intuitions does not alone prove a foundational role of AC in this regard. Still, it is an important indication that AC seldom contradicts with our intuition, and that in many diverse moral scenarios, we seem to be able to explain our intuitive judgment by reference to well-being promotion.

Another response to this question would be to look more closely at the profoundly human nature of well-being. In the integrity chapter I argued that, as opposed to seeing well-being as a bizarre and foreign thing, we should see it as the essential element of human worth and goodness. In this respect, there is good reason to believe, beyond noting parallels, that promoting well-being relates on a fundamental level to all our moral beliefs.

I believe the AC theory that emerges after consideration of the five objections is a much more plausible and interesting one for future consideration. Still there are lasting objections that must be faced by this refined form. Some of these have been looked at in the thesis. For example, in the justice chapter I argued that AC fails to explain the importance of a retributive sense of justice. Also in the justice chapter I raised fears about including too much within the well-being maximand (for reasons of brevity however, I never provided a thorough analysis of the component parts of the well-being maximand). Including certain values within the maximand, such as fairness, might be seen as an implicit concession to the norms of other moral theories, and so would undermine the distinct aspects of AC.

There also remain other worthy objections not raised in the thesis. Even if we accepted many diverse components within the well-being maximand, we must

consider how to treat these distinct elements, leading to the difficult question of whether or not their incommensurability would prevent this. Nozick presented troublesome objections about the ACist disregard of the authenticity of our experiences (Nozick 42). Additionally, Derek Parfit's "mere addition paradox" raises concerns about the AC treatment of potential lives and well-being.

Ultimately, the refined version of AC presented in this thesis may not be immune to important criticisms, but I hope to have shown that certain objections do not pose as serious a threat. By clarifying AC in respect to these objections, and facing new objections in this clarified form, we might advance the theory to a point where it sheds an important light on our overall idea of morality.

Works Cited

- Arneson, Richard. "Sophisticated Rule Consequentialism: Some Simple Objections" *Philosophical Issues* 15:1 (2005), 235-251.
- Bentham, Jeremy. *Anarchical Fallacies*. 1843.
- Bentham, Jeremy. *Introduction to the Principles of Morality and Legislation*. 1823.
- Blackburn, Simon. *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Clarendon, 2000.
- Brandt, Richard. *A Theory of the Good and the Right*. University Park: Prometheus Books, 1998.
- Brink, David O. "Impartiality and Associative Duties." *Utilitas* 13:2 (2001): 152-170.
- Brink, David O. *Perfectionism and the Common Good: Themes in the Philosophy of T.H. Green*. Oxford: Clarendon, 2003.
- Brink, David O. "Some Forms and Limits of Consequentialism." *The Oxford Handbook of Ethical Theory*. Ed. David Copp. Oxford: Oxford UP, 2006.
- Broad, C.D. "Self and Others." Herbert Spencer Lecture. Oxford University, 1953.
- Coady, Tony. "The Problem of Dirty Hands" *Stanford Encyclopedia of Philosophy*. Stanford, 2009.
- Dorsey, Dale. "How Not to Argue Against Consequentialism." KU.edu. Kansas University. Web. 10 June, 2015.
- Goodin, Robert E. *Protecting the Vulnerable: A Re-analysis of Our Social Responsibilities*. Chicago: Chicago UP, 1985.
- Grono, Nick. "The deterrent effect of the ICC on the commission of international crimes by government leaders." *The Law and Practice of the International Criminal Court: Achievements, Impact and Challenges*, The Hague, 2012.
- Hare, R. M. *Essays in Ethical Theory*. Oxford: Oxford UP 1989.
- Hare, R. M. *Moral Thinking: Its Levels, Method and Point*. Oxford: Oxford UP, 1981.
- Harris, John. "The Survival Lottery." *Philosophy* 50 (1975): 81-7.

- Hodgson, D.H. *Consequences of Utilitarianism A Study in Normative Ethics and Legal Theory*. Clarendon, Oxford, 1967.
- Hooker, Brad. *Ideal Code, Real World*. Oxford: Clarendon Press, 2000.
- Hurley, Paul. *Beyond Consequentialism*. Oxford: Oxford UP, 2009.
- Jackson, Frank. *A Defense of Conceptual Analysis*. Oxford: Clarendon Press, 1998.
- Kagan, Shelly. *The Limits of Morality*. Oxford: Clarendon Press, 1989.
- Kamm, Frances. *Intricate Ethics*. Oxford, Oxford UP, 2007.
- Kaplan, M. A. "Some Problems of the Extreme Utilitarian Position." *Ethics* 70 (1960): 228-32.
- Kapur, Neera Badhwar. "Why it is Wrong to Be Always Guided by the Best: Consequentialism and Friendship" *Ethics* 101:3 (1991): 483-504.
- Kofman, Daniel. "Moral Theory Convergence: Rule-Consequentialism and Moral Pluralism." University of Ottawa. Ottawa, Canada. Sept.-Nov. 2013. Lecture.
- Kymlicka, Will. *Contemporary Political Philosophy: An Introduction*, 2nd ed. Oxford: Oxford UP, 2002.
- Lyons, David. *Forms and Limits of Utilitarianism*. Oxford: Oxford UP, 1965.
- Lyons, David. *Rights, Welfare, and Mill's Moral Theory*. Oxford: Oxford UP, 1993.
- MacIntyre, Alasdair. *A Short History of Ethics*. New York: Macmillan, 1966.
- McCloskey, H. J. "Respect for Human Moral Rights Versus Maximizing Good." *Utility and Rights*. Ed. Raymond Frey, U of Minnesota Press, 1984.
- McNaughton, David. *Moral Vision*. New York: Blackwell, 1988.
- Mill, John Stewart. *On Liberty*. 1859.
- Mill, John Stewart. *Utilitarianism*. 1863.
- Mulgan, Tim. *The Demands of Consequentialism*. Oxford: Clarendon, 2001.
- Nagel, Thomas. *Equality and Partiality*. Oxford: Oxford UP, 1991.

- Nielson, Kai, 2000, "There is No Dilemma of Dirty Hands." *Cruelty & Deception: The Controversy over Dirty Hands in Politics*, Eds. Rynard, Paul and David P. Shugarman. Melbourne: Pluto Press, pp. 139–155.
- Nozick, Robert. *Anarchy, State and Utopia*. New York: Basic Books, 1974.
- Parfit, Derek. "Prudence, Morality and the Prisoner's Dilemma." *Proceedings of the British Academy* 65 (1979), 539-64.
- Parfit, Derek. *Reasons and Persons*. Oxford: Oxford UP, 1984.
- Pettit, Philip, and Michael Slote. "Satisficing Consequentialism" *Proceedings of the Aristotelian Society* 58 (1984): 139-176.
- Portmore, Douglas W. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford UP, 2011.
- Rawls, John. *A Theory of Justice*. Harvard UP, 1971.
- Rawls, John, "Two Concepts of Rules." *Value and Obligation*. Ed. Richard Brandt. New York: Harcourt, 1961. 230-238.
- Raz, Joseph. *The Morality of Freedom*. Oxford: Clarendon Press, 1986.
- Ross, W.D. *The Right and the Good*. Oxford: Clarendon, 1930.
- Scanlon, T.M., "Rights, Goals, and Fairness." *Consequentialism and its Critics*. Ed. Samuel Scheffler. Oxford: Oxford UP, 1988. 74-92.
- Scheffler, Samuel. "Agent-Centered Restrictions, Rationality, and the Virtues." *Consequentialism and Its Critics*. Ed. Samuel Scheffler, Oxford: Oxford UP, 1988.
- Scheffler, Samuel. *The Rejection of Consequentialism: A Philosophical Investigation of the Consideration Underlying Rival Moral Conceptions*. Oxford: Oxford UP, 1994.
- Scheffler, Samuel. "Relationships and Responsibility" *Philosophy and Public Affairs* 26: 3 (1997): 189-209.
- Sen, Amartya. "Rights and Agency." *Consequentialism and its Critics*. Ed. Samuel Scheffler. Oxford: Oxford UP, 1988. 187-223.
- Sen, Amartya and Bernard Williams. "Introduction" *Utilitarianism and Beyond*. Eds. Amartya Sen and Bernard Williams. Cambridge: Cambridge UP, 1982.
- Shafer-Landau, Russ. *The Fundamentals of Ethics*, 3rd ed. Oxford: Oxford UP, 2015.

- Shaw, William H. *Contemporary Ethics: Taking Account of Utilitarianism*. Oxford: Blackwell, 1990.
- Sidgwick, Henry, *The Method of Ethics*. Chicago: University of Chicago UP, 1907.
- Sinott-Armstrong, Walter. "Consequentialism" *Stanford Encyclopedia of Philosophy*. Stanford, 2003.
- Slote, Michael. *Common Sense Morality and Consequentialism*. London: Routledge, 1985.
- Smart, J.C.C. "An Outline of a System of Utilitarian Ethics." *Utilitarianism: For and Against*. Eds. J.C.C. Smart and Bernard Williams. Cambridge: Cambridge UP, 1973.
- Singer, Peter. "Is Act-Utilitarianism Self-Defeating?" *The Philosophical Review* 82 (1972), 94-104.
- Sobel, David. "The Impotence of the Demandingness Objection" *Philosophers' Imprint* 7:8 (2007): 1-17.
- Sprigge, T. L. S. "A Utilitarian Reply to Dr. McCloskey," *Inquiry* 8 (1965): 264-91.
- Stroud, "Moral Overridingness and Moral Theory." *Pacific Philosophical Quarterly* 79 (1998), 185-6.
- Tedesco, Matthew. "Intuitions and the Demands of Consequentialism" *Utilitas* 23:1 (2011): 94-104.
- Thomson, Judith Jarvis. *The Realm of Rights*. Harvard: Harvard UP, 1990.
- Unger, Peter. *Living High and Letting Die: Our Illusion of Innocence*. Oxford: Oxford UP, 1999.
- Urmson, J.O. "A Defense of Intuitionism." *Proceedings of the Aristotelian Society. New Series* 5 (1974-1975) 26-38.
- Williams, Bernard. "A Critique of Utilitarianism." *Utilitarianism: For and Against*. Eds. J.C.C. Smart and Bernard Williams. Cambridge: Cambridge UP, 1973.
- Williams, Bernard. *Moral Luck*. Cambridge: Cambridge UP, 1981.
- Woodcock, Scott. "When Will Your Consequentialist Friend Abandon You for the Greater Good?" *Journal of Ethics and Social Philosophy*. 4. 2 (2010): 1-23.