Economics and Philosophy, 14 (1998), 249–281

WELL-BEING AS THE OBJECT OF MORAL CONSIDERATION

DAVID SOBEL
Bowling Green State University

An adequate moral theory must take (at least) each person into account in some way. Some think that the appropriate way to take an agent into account morally involves a consequentialist form of promoting something about her. Others suggest instead that morality requires a Kantian form of respecting something about an agent. I am interested here in pursuing the former line. When we pursue the broadly consequentialist line we come to this question: what should we promote on the agent's behalf when we are taking her into account morally?

Consequentialists have typically, if not unanimously, answered that we ought to promote an agent's well-being. The plausibility of this answer depends on what well-being is and different accounts of well-being paint importantly different pictures of its nature. In this paper I will consider the plausibility of this answer when we have in mind a preference account of well-being.

There are, very crudely speaking, two models for preference accounts of well-being that have found favor. The first model, seemingly popular in the decision theory literature, holds that all of one's (axiom-obeying) preferences are connected with one's well-being. Such a model

I gratefully thank Elizabeth Anderson, David Copp, Janice Dowell, Allan Gibbard, Don Hubin, Dan Jacobson, Jim Joyce, Ned McClennen, Peter Railton, Arthur Ripstein, Wayne Sumner, Sergio Tenenbaum, David Velleman, and Mike Weber for their help on this paper. I would like to especially thank Justin D'Arms, Steve Darwall, and Connie Rosati for their extensive help and encouragement. I presented an earlier version of this paper at the 1997 *Utilitarianism Reconsidered* conference. I benefited there from comments by Richard Arneson and my commentator James Griffin. I also presented drafts to the philosophy departments at Bowling Green State University, The University of Michigan, and North Carolina State University. I am grateful to these audiences for many useful suggestions.

makes no distinction between preferences whose satisfaction is connected to one's well-being and those that are not. This model insists that preferences that, for example, stem from team-spirit, patriotism, or morality are just as firmly connected with one's well-being as are more narrowly self-serving preferences.

This model makes no room, as I will show, for clearheaded self-sacrifice of one's well-being for moral reasons. This rather implausible understanding of well-being reveals well-being to be an inappropriate object of moral promotion. There are two reasons for this that I will stress below. First, such an account leads to incoherence when conjoined with a consequentialist or social-choice framework that requires self-sacrifice. Second, taking this understanding of well-being into account morally will penalize those with altruistic or moral tendencies while rewarding those whose concerns are more narrowly personal.

The second model, now dominant in philosophy, claims that one's preferences must be radically informed before they are reliably connected with one's well-being. Further, and more importantly for our purposes, this second model typically allows that some of one's informed preferences, for example, moral preferences, have no special connection to one's well-being.[1] Thus to get at just an agent's well-being, on this model, we must screen some of her concerns. But which concerns should we screen? This question is not convincingly addressed in the literature. We will see that any plausible preference account must screen not only blatantly moral preferences, but additional important elements of an agent's concerns as well. This more plausible model reveals well-being to reflect too little of what matters to an agent to represent her adequately to the group for moral purposes. A person might well not endorse being taken into account by having only her well-being promoted.

What would it be for the consequentialist to take a person into account morally in a way that she reflectively endorses? The suggestion will be that when we take an agent into account morally, we should promote what the agent informedly wants us to promote for her sake.

[1] This philosophical position is not new. One can find it in Mill (1979, Chapter 2); Sidgwick (1981, pp. 105–15); Brandt (1979, pp. 10, 113, 329); Hare's explicit agreement with Brandt's position can be found in Hare (1981, pp. 101–5 and 214–6) as well as Senor and Fotion (eds.) (1988, pp. 217–8); Griffin (1986, pp. 16–26); Rawls (1971, pp. 407–24); Gauthier (1986 Chapter 2); Darwall (1983, Part II); Harsanyi (1982, p. 55); Railton (1986, p. 9). Some have doubted that Mill held a subjectivist account of well-being, claiming that his competent judges test should be read as an epistemic tool for determining an agent's good rather than an account of what makes it the case that something is good for an individual. But Mill's words are conclusive. Just as he is about to introduce the competent judges test he tells us that he is addressing the question of 'what makes one pleasure more valuable than another'. (Mill, 1979, p. 8.) Gauthier and Harsanyi are less clear than the others that elements of an agent's concerns must be screened to get at just the agent's well-being.

Each person should control, if not the weight that her concerns receive in moral deliberation, at least which of her concerns get weight insofar as she gets weight. Welfarist consequentialists cannot offer agents this control.

The proposal I offer attempts to remedy the inadequacies of exclusive focus on well-being for moral purposes. The proposal is this: we should allow the (informed) agent to decide for herself where she wants to throw the weight that is her due in moral reflection, with the proviso that she understands the way that her weight will be aggregated with others in reaching a moral outcome. I will call this the 'autonomy principle'. The autonomy principle, I claim, provides the consequentialist's best prospect for taking people into account morally in a way that they endorse.[2] I do not claim that such a version of consequentialism can avoid all the problems that people have found with other variants of consequentialism. Rather, I will argue that a consequentialist view that respected the autonomy principle has decisive advantages over other versions of consequentialism, most notably welfarist versions.

A Dilemma

Much of my case for the claim that well-being is not the appropriate object of moral concern rests on my ability to show that well-being cannot be all it seems. What rationally matters to a person and her well-being can, and typically do, come apart. If this is so we would have to choose which merits our moral attention and which does not. Much of the intuitive force of the thought that the way to take a person into account morally is to promote her well-being, it seems to me, is owed to the presupposition that well-being and what matters to a person do not come apart.

Would you rather that 10,000 acres of rain forest be preserved or that your career get an important boost? Commonsensically we think that a person could prefer the former without this signaling that rain forest preservation promotes that person's well-being more than career advancement.[3] Many of us talk as though it is possible for the rich to

[2] The autonomy principle allows a person to choose for herself what she wants promoted when she is taken into account morally. It is thus guaranteed of being endorsed in this conditional sense: if I am going to be taken into account morally by having something promoted for my sake, then I endorse promoting these values for my sake. However, the agent need not endorse being taken into account morally by having something promoted (in a consequentialist sense). Seemingly no sensible ethical theory could allow us each to decide for ourselves the manner in which we are taken into account in the latter sense. Thus the autonomy principle's ability to secure the agent's conditional endorsement for the manner in which she is taken into account morally is the most that can be hoped for from an ethical theory.

[3] Perhaps it would be more accurate to say that commonsensically we would respond to such questions by asking: 'Prefer in what way?' I will show that a plausible preference

vote for progressive taxation policies because they prefer a more equitable distribution of wealth to one that better serves their own interests. But preference accounts of well-being suggest that there is an important connection between what we prefer and our well-being, at least once we are adequately informed about the options.[4] How should preference accounts of well-being handle cases like the preference for rain forest preservation? I will argue that this apparently innocent question leads us down a path which raises significant difficulties for preference accounts of well-being and the consequentialist ethical theories that would make use of them.

Preference accounts of a person's good as they are typically used by consequentialists face a serious dilemma. The dilemma is this: either the account takes everything that matters to the person to constitute a preference the satisfaction of which contributes to the agent's well-being or it does not.[5] If the account does this it has exhausted what matters to the person in constructing the agent's well-being. Thus a person could not care about anything beyond the extent to which it serves her interests. Self-sacrifice would be impossible.[6] The problem with this first horn of the dilemma is that there is something conceptually amiss with the thought that 1) all our concerns are given weight via our well-being in the input into the consequentialist calculus, yet 2) we can be rationally motivated to promote the output of the consequentialist calculus even when the output differs from the agent's own input.

On the other hand, if the account allows that one can have concern for something beyond the extent to which that concern furthers the

account of well-being must concern itself with the different ways that we care about things and not just how much we care about them.

[4] I will generally ignore questions about the appropriate epistemic vantage point from which one's preferences are alleged to be correlated with one's well-being. I do, however, think that there are real difficulties with preference accounts of well-being stemming from this issue. See Sobel (1994) and Rosati (1995).

Importantly for the arguments that follow, I will assume that the agent whose good is in question deliberates from an idealized epistemic perspective. Thus, complaints that the agent's preferences do not track her good because she is not sufficiently appreciative of her options, or that she makes mistakes concerning the causal implications of her acts, are meant to be out of place. Only if there is such a perspective is the project of constructing an agent's well-being from her preferences plausible. When I speak of an agent's rational concerns I mean concerns that exist after some such epistemic idealization.

[5] Some might not want to call all aspects of an agent's motivational set mere preferences. Perhaps, for example, cases in which the agent acts on principle should not be thought of as reflecting a preference. In this paper I will call all aspects of what matters to the agent preferences. The terminology does not matter. The important issue here is whether one cares about things in importantly different ways and if these differences cause the satisfaction of one's concerns to impact differently on one's well-being and on others' moral obligation to help achieve these satisfactions.

[6] By self-sacrifice I mean the deliberate choosing of what makes one's own life go less well. That is, self-sacrifice requires sacrifice of one's well-being.

agent's well-being, then well-being does not capture the whole of what matters to the agent. And if this is allowed, then a primary reason to focus our moral attention only on people's well-being would seem to be undermined. We ought to wonder why we should focus only on the agent's well-being when much of what matters most to the agent is not served by serving her well-being and when she perhaps prefers that we promote the broader category of what matters to her rather than just her well-being.

The problem for these consequentialists[7] is that either they leave no room for self-sacrifice or they make such room by excluding things that rationally matter to us from being a part of our well-being. In the former case we are incapable of living up to the kind of morality they espouse.[8] In the latter case we begin to be puzzled about why it had seemed so obvious that well-being is the sole appropriate object of moral concern. We ought to be puzzled, given the lack of help in the literature, about which subset of an agent's preferences is connected to the agent's well-being and why exactly that subset deserves special moral attention.

This paper is long and I pursue several peripheral issues. Thus it is more crucial than usual to have and keep in mind a map of the overarching structure of the paper. The rest of this paper will have three parts. In Part 1 I argue that consequentialists need to reject accounts of well-being that treat all of one's preferences as being connected with one's well-being. In Part 2, I will consider how an advocate of preference accounts of well-being might try to respond to the arguments of Part 1 by identifying a proper subset of one's preferences that is connected with one's well-being (e.g., the agent's non-moral preferences). I will argue that extant methods of attempting to separate out the appropriate subset are inadequate. But even if we were successful in our search for the right subset of preferences, we would immediately be confronted with the questions I press in Part 3. In Part 3 I wonder why the favored subset should be taken to fully morally represent the agent to the group. I will

[7] My criticisms in this paper are not relevant to all consequentialist positions. When I talk about consequentialists in this paper I have in mind those who 1) claim that the rightness of acts, rules, etc., is determined by the extent to which those acts, rules, etc., promote a specified dimension of value, 2) claim that the extent to which this specified dimension of value is promoted depends on the extent to which individuals have that dimension promoted, and 3) claim that the extent to which an individual has this dimension of value promoted is determined by the extent to which that agent's preferences are satisfied. I take it that one could be a consequentialist while denying 2 and/or 3. However, the combination of 1, 2, and 3 represents the broad path of consequentialists these days.

[8] This is not exactly correct. Consequentialists need not suggest that the moral option requires self-sacrifice when doing what is best for the group is also best for the individual. I discuss this further below. Additionally, if we were psychological egoists then, I guess, we would be good consequentialist agents since we would choose the option, of those actually available to us, which maximized the group's good. I ignore this method of compliance with consequentialism's demands.

argue that once we appreciate that what matters to a person and what makes her life go best can come apart, the question of why well-being is what matters morally takes on a new kind of urgency. I will consider reasons to focus our moral attention on well-being, the broader notion of what matters to a person, and a narrower notion than well-being such as basic needs. I also offer, in Part 3, some reasons to prefer my autonomy principle version of consequentialism to other versions, especially welfarist versions.

1. INCLUSIVE ACCOUNTS OF WELL-BEING

Well-being is intended to be a measure of how well an individual's life is itself going for her, considered apart from other kinds of value, such as moral value, that a life could have. If we accept certain versions of hedonism, the concept of well-being is relatively unproblematic. Some versions of hedonism have it that all and only pleasurable sensations can make a contribution to an agent's well-being. On such a view, it is relatively straightforward what contributes to an agent's well-being. Pleasurable sensations that are part of my consciousness contribute to my well-being, other things do not. Of course issues could arise concerning who I am, or would be, under certain alterations which could complicate the question of which pleasurable sensations affect *my* consciousness. But this seems merely to imply that the concept of an agent's well-being will be no less problematic than the concept of the self. There are further complications in the notion of self-interest, beyond complications arising from ambiguities in the concept of individual identity, when we move beyond the simplest hedonistic theories of value.

Hedonists need not suggest, although frequently they do, that all we rationally care about is the sensation of pleasure. Hedonists therefore have no special difficulty in understanding how we could have an 'internal reason'[9] to promote states of affairs that do not promote our well-being.[10] Such hedonists can claim that there is such a thing as coming to care about others for moral reasons and allow that in some cases this moral concern moves clear-headed and informed people towards actions that are less than optimal in terms of their own well-being. Hedonists can allow all this because they can make room for

[9] Bernard Williams coined this term in Williams (1981).

[10] Gibbard (1990, p. 18), writes, 'rationality, in the ordinary sense, often consists not of using full information, but of making best use of limited information'. I think we should preserve this 'ordinary sense' of rationality and hence find that prudential behavior can be rational and non-maximizing of self-interest when, for example, the agent lacks information. The difficulty, therefore, in finding rational the non-maximization of self-interest only arises when the agent is in the favored epistemic state to determine her interests.

something mattering to a person beyond its capacity to improve her well-being. The hedonist can acknowledge that things besides our pleasure rationally matter to us, but insist that what matters to us in those ways is not part of our well-being. Yet some preference accounts of well-being do not make it clear how they can similarly make room for rational action which does not maximize one's own interests. This is because they assimilate any reason for caring about something into a reason which promotes one's well-being when that thing comes about.

Let us call a preference account of well-being 'inclusive' if it takes any 'all things considered'[11] preference for x over y, regardless of the reason for that preference or the way it is held, to imply that x makes the individual's life go better than y.[12] 'Exclusive' accounts of well-being, then, exclude some of a person's rational preferences from having a connection to the agent's well-being. Inclusive accounts of well-being use up all of one's internal reasons for action in developing a conception of well-being and leave the agent no room for non-well-being related motivations for action. The inclusive account need not suggest that every preference is consciously aimed at making the agent's own life go better. One mistakes the direction of explanation here if one thinks of one's well-being being antecedently set and creating constraints on how much one may care about others. Rather, by the inclusivist's lights, it is how much one cares about others that determines how much their doing well benefits you. However, the extent to which an option serves one's interest is read directly from the strength of the preference for it. If this were an adequate account of well-being there would be no conceptual room for rationally caring about things out of proportion to the impact they have on one's well-being.

Decision theorists typically, at least tacitly, embrace an inclusive account of an agent's well-being.[13] Social choice theory is the branch of decision theory that deals with aggregating individuals' preferences into fair social decisions. My case against inclusive accounts is most obviously directed against decision theory and social choice theory. There are several reasons for this. First, decision theory is the most

[11] Unless I explicitly say otherwise I mean 'all things considered' preferences when I talk about preferences. Of course one might appropriately say that an agent has a desire for x, and hence an internal reason to x, even when she has a stronger desire to y in that circumstance. However, here I am only concerned with one's all things considered preferences and the all things considered internal reasons they generate.

[12] I am assuming here that the agent's preferences obey the reader's favored set of decision-theoretic axioms.

[13] I should qualify this. It is rather that when decision theorists are best interpreted as offering an account of well-being, they typically seem to intend an inclusive account of well-being. Too frequently it is unclear if the decision theorist's utility function is meant to represent well-being (as opposed to, for example, choice worthiness given one's information).

widely known and accepted preference account of well-being. Second, decision theorists often invite the inclusive interpretation of their theory. Decision theorists could adopt an exclusive account of well-being (in ways I will discuss in Part 3), nonetheless they typically either ignore the issue or embrace the inclusive interpretation. Third, social choice theorists use the decision theorist's inclusive account of well-being as the object of aggregation at the moral stage. Although my case in this section is most obviously applicable against decision theory and social choice theory, I believe the moral of the case to apply more broadly to any attempt to use an inclusive account of well-being as the object which morality demands that we promote.

INCLUSIVISM AND MOTIVATION TO BE MORAL

The inclusivist interpretation of well-being leads to an absurd picture of moral motivation. Consider two different pictures of how an individual could be motivated to promote aggregate well-being. On the first picture what is best for the group is best for the individual. On this picture we could imagine that either the individual's good and the common good contingently coincide or that the agent prefers that the aggregate be promoted under that description. On the second picture the agent could be motivated to promote the aggregate even when doing so was less good for herself. The agent might, for instance, be impartial between peoples' well-being without the aggregate therefore becoming best for her. That is, the agent's well-being stays put and deviates from what maximizes for the group, yet the agent is motivated to pursue the aggregate nonetheless. Many consequentialists accept this latter picture, treating it as an account of what we do when we take up the moral point of view. It is this picture that the inclusivist denies. Thus the only way for our inclusivist consequentialist to understand an impetus from the agent's motivational set towards the morally required act is to follow the first picture.

Some who adopt an inclusivist understanding of an agent's well-being, cut morality's demands down to size so that the morally required act does not conflict with the agent's (sophisticated) pursuit of her interests.[14] Our consequentialist is different. For her, the morally required act is defined independently from, and has no conceptual connection with, what is best for the deliberating agent. If what is good for the agent is also what is best for the group, this will be because of special features of the agent or the situation. This happy predicament is by no means guaranteed.

Let us now consider, then, what the inclusivist consequentialist must say about a particular case. There is a cake that is to be divided between

[14] See Gauthier (1986).

Desdemona and Iago. Iago is unproblematically egoistic and wants all the cake. Desdemona, on the other hand, is more fair-minded. She says that while she would eagerly take all the cake if the portion that she did not take would go to waste, she thinks it would be best if the cake were evenly split between the two. Desdemona is not especially fond of Iago. Her sole reason for the even division, she tells us, is that it would be fair. How should we divide the cake according to the inclusivist consequentialist?

Consider Gibbard's observations about this case. He writes,

> With a cake there is a natural compromise between their respective first choices: we can split the difference. That is to say, we might give three-quarters of the cake to Iago and one-quarter to Desdemona. That way, they each get an amount of cake that is halfway between what they would get if Iago could dictate and what they would get if Dedemona could dictate.
>
> Desdemona should object, it seems to me – but on what grounds? [...] One ground for objection [...] seems especially clear: there is simply no parity between Iago's selfish first choice and Desdemona's commitment to a fair outcome. [...] This is indeed the outcome Desdemona herself would choose if she had the power, but that does not mean that we are partial to her when we select it. We simply judge that she too would be impartial given the choice, and he would not.[15]

But such sensible thoughts are unavailable to the inclusivist consequentialist. Indeed it is hard to see how the inclusivist can avoid the conclusion that Iago's share ought to be bigger than Desdemona's. In fact, the situation is worse than that for our inclusivist consequentialist. For, imagine that Desdemona is powerfully motivated by her understanding of morality, and her understanding of morality is an inclusivist consequentialist one. She will then herself conclude in the above example that she ought to give Iago a larger share. But since, according to inclusivism, this new preference that results from the consequentialist thought is tied to her interests she must reapply the same consequentialist thought, adjusting for where her well-being now lies. Again and again she will find that what seemed like a fair compromise only moments ago must now be revised in response to her current interests. This process will result in Desdemona getting a vanishingly small piece of the cake and Iago's share approaching all of it. If preferences motivated by the same kind of concern for others that gets consequentialism going get counted as part of one's well-being, then the concerns of those with moral preferences can get fully washed away.

This process of spinning moral concern into self-interest will only come to an end when Desdemona most wants Iago to get all the cake.

[15] Gibbard (1987).

Only then could she be motivated towards what the inclusivist consequentialist tells us is the moral outcome. Thus in order for Desdemona to be moral when confronted with the Iagos of the world, she must fully efface the needs and wants she had prior to the moral situation (Desdemona could consistently want to have all the cake if the uneaten portion went to waste) and make what is best for the Iagos of the world also best for herself. I take it that at the very least it is obvious that there are ways of being moral in such situations other than so effacing oneself. Thus I claim the inclusivist premise leads us to absurd pictures of what it takes to be moral. If we are to make room for other ways of being moral in such cases we must reject the inclusivist premise.

A parallel argument creates similar trouble for the inclusivist in the intrapersonal case. Imagine that we are attempting to maximize well-being over two different time-slices of the same person. Suppose one of the perspectives is motivated to be prudent (that is, motivated to maximize her well-being across time) and the other is not. We will again face a situation in which the prudent perspective's share will inappropriately erode and the imprudent's share will inappropriately be bolstered, unless we can distinguish between concerns connected with one's own well-being (at a time) and concerns that exist thanks to one's broader concerns.

Who Can be Moved by Consequentialist Morality?

The enterprise of consequentialism and social choice theory presuppose that there are moral and immoral ways of treating people's preferences. They prescribe that individual's preferences get fairly aggregated into the social decision, that is, they recommend a method of social division on moral grounds. Weirdly, the picture the normative theory of social choice suggests, is one in which inclusivism is true of everyone except the social choice theorist while she has her social choice theorist hat on. Kenneth Arrow provides one notable expression of this thought in approvingly quoting Bergson.

> According to this view the problem is to council not citizens generally but public officials. Furthermore, the values to be taken as data are not those which might guide the official if he were a private citizen. The official is envisioned as more or less neutral ethically. His one aim in life is to implement the values of other citizens as given by some rule of collective decision-making.[16]

Thus special and unlikely sorts of people have to be invented who could care about and be moved by the collective decision-making rule.

[16] Arrow (1963, p. 107).

Ordinary citizens would have already had all of their concerns summed up in the input to the aggregation system, and hence will not be able to be motivated towards the aggregate.

There is a problem for the social choice theorist prior to the problem of interpersonal comparisons of utility. It must be determined if the stated preferences reflect moral concern for others.[17] The same flavor of moral motivation that gets social choice theory going also informs many of our everyday choices. If the preference is motivated by the same kind of concern for other's well-being that the social choice theorist is going to reapply at the stage of aggregation of interests, then it would seem inappropriate to give such a preference weight as part of the agent's well-being as well. In cases such as this, individuals have done some of the social choice theorist's work for them by attempting to treat the interests of all, or at least some, as of equal weight to their own preferences. If the interests of others is given what Sen calls non-sympathetic weight in one's preference ranking, then the level of satisfaction of one's preferences will not reflect one's well-being.[18]

The social choice theorist's likely response would be that the person who allows her moral views to influence her motivations is, to that extent, making her morally preferred option a better option for herself in terms of her own interests. This construal of acting on principle as merely another kind of preference or inclination defines Kantian positions out of existence. But the point I would insist on here is that social choice theorists and consequentialists do not think of their system

[17] Surprisingly Arrow (1987, p. 727), writes in reply to Gibbard's (1987), formulation of this problem that it has always bothered him, but not in print. Arrow concludes that 'penalizing altruists hardly seems reasonable. But I find it hard to state a coherent position'.

[18] Sen (1982). Sen distinguished between concern for the well-being of others out of sympathy and concern for others out of commitment. Sympathy, for Sen, is the concern you feel for others when their hurt hurts you. If you do not yourself feel pained by the pains of others yet 'you think it wrong and are ready to do something to stop it, it is a case of commitment'. Sen is pointing out the distinction between the case in which something bad happening to another counts as a harm to one's own well-being, and the case in which we are moved to help the other, but not because in doing so we are doing what makes our own life go best. But drawing this distinction is complicated. Sen seems to suggest that when a harm to another produces a sympathetic resonance in me this should count as a harm to my well-being. And when there is not this resonance but some more abstract, less visceral, and impersonal response which provokes us to help there is less or no harm to our well-being, yet we may have a powerful motivation to act (e.g., in the name of morality). But I would have thought that a more participatory emotional reaction could reflect a distinctively moral reaction as well. Often to be pained by the pain of another is to be already, in part, responding to the situation morally. At least, I think we can say that some pleasures and pains are moral in the sense that they are felt because of a person's moral character. Sen relies on the commonsense difference between sympathy and commitment more than he argues for it. Although I find appeals to commonsense convincing here, others, including many decision theorists, do not.

of aggregating interests as merely reflecting a self-interested preference that they have. Rather they seem to conceive of it as a fair method of collective decision making. But they have made no room for the distinctive kind of pro-attitude we can have towards actions we think morally justified.

Consequentialists who make use of such an inclusive notion of well-being should be unsurprised when they find they have left no motivational space for an agent rationally to care about their notion of morality. Indeed, ironically, such consequentialists would seem unable to give voice to their distinctively moral concerns once they adopt an inclusive notion of well-being. Their tracts extolling the virtues of impartially caring about the sum of peoples' interests must look to us as either the expression of an unlikely personal preference on their part, or as wistful pining for a distinctive moral mode of concern for which they have failed to make conceptual room. Issues about why one should be moral will not disappear if we reject inclusive accounts of well-being, but we will have at least removed an important barrier to the coherence of any answer to that question. One might wonder why we should sacrifice our well-being for other things that matter to us, but this is a far more tractable problem than why we should sacrifice our interests for something that does not matter to us at all.

A Problem and a Proposal

I have argued that consequentialists cannot sensibly make use of an inclusive account of well-being. Now I want to try to lay bare what goes wrong when we advert to an inclusive account of well-being as that which we should take into account when we take an agent into account morally. To do so I need to distinguish between different conceptions of what it means to prefer x to y. First, one could mean that, if it were up to you, you would bring about x rather than y (ignoring cases of weakness of will). Call this 'causally preferring'. Second, you could mean that you wish x would come about rather than y. Call this 'non-causally preferring'. Both causally and non-causally preferring are examples of preferences for outcomes. Third, you could mean that with a (in this case, consequentialist) method of determining the rightness of acts already in place, you would put forth x rather than y as best representing where you want to throw your weight in the (consequentialist) framework. Call this 'preferring as input'. Each of these understandings of preference can come apart from the other in everyday cases. I will here focus on the distinction between preferences for outcomes and preferences for input.

To see the distinction I have in mind, think of a group deciding what movie all should see. Sometimes people say, 'I would like us to see

movie A' and what they mean is, 'I think, all things considered, we ought to go see A'. This would be a kind of preference for a certain outcome – that the group see movie A. Other times when people say, 'I would like us to see movie A' they mean, 'I vote for A, but let's see the movie that democratically wins the vote'. This would be a preference meant to be used as input into a more or less well-understood aggregation procedure.

Think back now on poor Desdemona. She might well be disgruntled upon learning of the reason for the unequal division of cake in the case mentioned earlier. She might well complain that she did not understand what question was being asked of her. Had she understood the procedure for making the decision she would have stated a different preference. It is not that she does not have the preference that she originally gave, it is rather that she does not think that that preference is the one that is appropriate as input into the consequentialist calculus. This is different from the case of someone who only wants half the cake but realizes that only by asking for more than they want will they get what they want. This latter case is one in which strategic voting calls for misrepresenting one's preferences. Desdemona, on the other hand, has changed her answer in response to a better understanding of what is being asked of her. Her answers were always honest.[19] She is, in effect, asked to allow moral concerns to take care of themselves at the stage of aggregation and worry about what matters especially from her point of view at the input stage.

Desdemona thought she was being asked how she would divide the cake if it were up to her. Her answer to this question differs from the answer she would give if she understood the use that the consequentialist divider would make of her preferences. So here, what Desdemona prefers as an outcome comes apart from what she prefers as input. The Desdemona example shows that our moral theory needs to be sensitive to this distinction.

The best proposal I know for overcoming the problems in the

[19] I am trying to point towards a distinction between intrinsic desires as input and instrumental desires as input. The later are derivative from desires for outcomes, whereas the former are not. I would think the autonomy principle would do well to try to focus only on intrinsic preferences as input. One might also suggest that what gets moral weight via the autonomy principle must be something that actually has an intrinsic home in one's motivational system. Thus, for example, one might say that a person can veto any aspects of her motivational set from receiving weight, but she cannot give weight to things that are no part of her intrinsic motivational set. The general issue here is how to avoid problems with strategic voting. These problems seem to arise because what one intrinsically prefers is not manipulable in the same way as where one chooses to throw one's moral weight. I believe the best version of the autonomy principle would grant the agent maximal control over where to throw her moral weight compatible with avoiding strategic voting problems.

Desdemona case is to take into account those preferences that Desdemona would put forward if she knew what use her answer was going to be put to.[20] I call this the autonomy principle. The proposal is to take her into account morally using her preferences as input. This suggestion has many merits and problems, and I will not try to discuss them all here, although I will have more to say about it in Part 3. There will be difficulties with strategic voting on this approach.

The autonomy principle claims that people should be granted complete autonomy in deciding where to throw the weight allotted to them in moral reflection. One could think of the proposal in either of two ways. First, one could see it as showing how one's well-being need not be, and typically would not be, the only appropriate object of moral concern, since people could rationally throw their weight towards rain forest preservation or whatever. Secondly, one could argue that where an agent chooses to throw her weight in moral reflection deserves the name well-being.

I find this latter proposal implausible. It would force us to say that whether or not the satisfaction of a certain preference contributes to the agent's well-being depends on her views about what kinds of preferences appropriately make moral demands on others generally. Nagel and Scanlon find that certain kinds of preferences do not appropriately make moral demands on others.[21] But there is no suggestion that the sort of idiosyncratic or optional preferences which do not make moral demands on others are not the sort of things which can contribute to our lives going better and worse. Further, a person's goal could be to accomplish something without asking for any assistance. Thus, this goal would not receive weight via the autonomy principle. Nonetheless, how well one's life goes could clearly be affected by how well one does in accomplishing such goals. Thus, the autonomy principle should not be thought to capture the agent's well-being. So, while my proposal is perhaps well suited to rectifying what has gone wrong in Desdemona's case, it does not rectify the problem by offering a plausible account of well-being. This proposal suggests that the answer to the question of how to take another into account morally lies elsewhere.

[20] This suggestion, which is the core of the autonomy principle, surfaced from conversations I had with Justin D'Arms. Although I have not seen such a framework explicitly presented before, I think we can see it implicit in some presentations. For example, Goodin (1995, p. 142) writes that 'there is a deeper dynamic, inherent in the very nature of the collective decision process, which induces people to launder systematically their own preferences, and to express only a small subset of their preferences in the form of political demands'. Goodin apparently thinks it appropriate to take people into account using only the resulting 'self-laundered' preferences.

[21] Nagel (1986, Chapter 9) and Scanlon (1975).

2. EXCLUSIVE ACCOUNTS OF WELL-BEING

Once we recognize that the satisfaction of some of an agent's preferences is not part of her well-being, we must search for a subset of preferences which constitutes her well-being, if we are to continue to hold a preference satisfaction view of well-being. Many influential advocates of preference accounts of well-being accept that inclusive accounts are false. J. S. Mill argued that, 'Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure'. Sidgwick suggested that we focus only on 'what a man desires for itself – not as a means to an ulterior result – and for himself – not benevolently for others'. Richard Brandt claims that only 'self-interested' preferences are connected with one's well-being. Peter Railton thinks we should focus on 'non-moral' preferences. James Griffin is forced to offer an explicitly circular account of what constitutes the right subset of preferences. Derek Parfit rejects the 'Unrestricted Desire-Fulfillment Theory' in favor of the 'Success Theory' which 'appeals to all of our preferences about our own lives.'[22] However, these authors are not as helpful as they could be in getting us to see the shape of the subset of preferences that they have in mind. In the most systematic writings in this area, Mark Overvold argued that the desires which are connected with well-being are those such that the agent's 'existence at t is a logically necessary condition of the proposition asserting that the outcome or feature obtain at t'.[23]

What is the worry that causes these authors to restrict the set of preferences which constitutes one's well-being? Unfortunately the preceding authors, and the literature in general, are less than clear in answering this question.[24] One worry is that some of our preferences are motivated by the same kind of concern for others that motivates some to be consequentialists. Should the satisfaction of such preferences count as improving one's well-being? What about moral concern for others of a non-consequentialist kind? The consequentialist could suggest that once we exclude consequentialist preferences, the rest are neatly correlated with well-being. But this seems to distort the way in which intuitively

[22] Mill (1979, p. 259); Sidgwick (1981, p. 109); Brandt (1979, p. 329); Railton (1986, p. 20); Griffin (1986, p. 22); Parfit (1984, p. 494). It is not clear that Parfit counts as an advocate of the preference approach to well-being.

[23] Overvold (1982).

[24] Railton (1986, p. 30) allows that, 'it may turn out that an ideally informed and rational individual would want to seek as an end in itself (were he to step into the place of his present self) the well-being of others as well as himself'. Griffin (1986) writes that, 'The trouble is that one's desires spread themselves so widely over the world that their objects extend far outside the bounds of what, with any plausibility, one could take as touching one's own well-being'.

moral but non-consequentialist preferences are held. The same reasons to exclude consequentialist preferences seem to suggest that other kinds of moral preferences should be excluded as well. But how do we separate moral from non-moral concern for others? Are all non-morally motivated preferences correlated with well-being? What about the person who takes some into account in a consequentialist way, say, members of his country, but gives less weight to everyone else. Is this a moral preference? If not, is it directly correlated with one's well-being? We will see that such preferences, typically born of group identification, are difficult to neatly fit into the categories available to the consequentialist.

Just Eliminate the Moral Preferences?

Perhaps the most obvious initial tack to take here would be to insist that it is only moral preferences which fail to be part of one's well-being. The thought is that if we extract the agent's moral preferences we will be left with the set of preferences that constitutes one's well-being.[25] If this project were workable it would remain plausible that all that matters to an agent would receive expression either in her well-being or at the stage of aggregation of interests. If this were the case, then all that mattered to one non-morally would adequately receive expression in one's well-being and, the consequentialist might suggest, all that represents genuine moral concern would receive expression at the stage of aggregation. Welfarist consequentialism can give expression only to what matters to us in one of these two ways. Thus it is perhaps unsurprising that many consequentialists would be tempted by a 'just eliminate the moral preferences' picture. This picture would not force these consequentialist to sweep some aspects of what matters to the agent under the rug such that they receive no expression in the consequentialist's moral system. However, the attempt to depict all of our non-moral motivations as being neatly correlated with our well-being is strained.

There are two questions at issue here: 1) do we get an accurate account of well-being when we strip away an agent's moral preferences? And 2) does everything that matters to an agent receive adequate expression when we use the resulting account of well-being as that which gets fed into the consequentialist aggregation procedure? I believe the answer to both questions is 'no'.

First, we do not get an adequate account of well-being with the 'just

[25] Kant (1956) appears to endorse a 'just eliminate the moral preferences' picture of happiness. He writes on p. 20 that, 'All material practical principles are, as such, of one and the same kind and belong under the general principle of self-love or one's own happiness'. Reath (1989) argues that this appearance is misleading.

eliminate the moral preferences' approach. There are non-moral preferences that are not correlated with well-being. The clearest instances of this are cases where the agent identifies with a group such as a nation, religion, team, department, etc. In such cases we seem capable of caring about the success of the group beyond the extent to which the group's doing well constitutes a benefit for the agent. Alternatively we could focus on cases in which the agent has the recognizably moral way of taking others into consideration, but takes a subset of the group that consequentialism takes into account into account in that way. These are cases in which the agent treats the well-being of some but not all others in the way consequentialism recommends. In this way an agent could be motivated to self-sacrifice in the name of maximizing the subset's welfare. Clearly some such concerns are not happily characterized as moral (we could make the subset that gets consequentialist concern very small) yet they do not have a tight connection to the agent's well-being. Thus I am claiming that should I develop a preference to maximize the well-being of the group of people named David, this would be neither a moral preference, nor directly connected with my well-being.

Second, all of an agent's concerns do not receive adequate expression on the 'just eliminate the moral preferences' approach. There is the problem of distinguishing genuinely moral preferences from other kinds of preferences. How plausible will it be to express adequately all non-consequentialist but seemingly morally-based desires as either constituting part of the agent's well-being or as being adequately expressed by the consequentialist's aggregation process? No doubt the non-consequentialist moralist will rebel at the attempt to express what matters to them in the ways available to the consequentialist. But this, the consequentialist could say, is just to say that the consequentialist disagrees with the non-consequentialist, and will have to offer an error theory of non-consequentialist moral thinking. This error theory will have to find a neat way of parsing up non-consequentialist moral concern into desires which are correlated with the agent's well-being and desires which receive adequate expression at the aggregation stage. That is, the consequentialist will have to reinterpret the allegedly morally motivated but non-consequentialist desires as, most likely, personal preferences whose satisfaction is correlated with how well the agent's own life goes. Thus to the extent that the morally motivated but non-consequentialist desires do not receive expression by consequentialist aggregation, it must be claimed that they are merely personal preferences which contribute to the agent's well-being.

The deontological Kantian and the Christian moralist think it wrong to kill one to save several. They prefer that such killings not be done. How should we understand the connection between the satisfaction of such a preference and an agent's well-being? The non-consequentialist

no doubt conceives of such preferences as expressing genuine moral concern in much the way that consequentialists think that consequentialist aggregation constitutes the appropriate form of moral concern. The existence of non-consequentialist but seemingly moral motivations is awkward for the 'just eliminate the moral preferences' approach. If the non-consequentialist moral concerns are allowed not to be connected with one's well-being, then they must be moral preferences. But they would not receive adequate expression in the aggregation stage.

The consequentialist is committed to saying that non-consequentialists are wrong about morality, and, hence, taking other people into account morally in the wrong way. It is something of a further step, it might seem, to suggest that the non-consequentialist's moral convictions are just as self-serving as blatantly egocentric preferences. This seems to deny non-consequentialist moral theories their status as rival ethical theories. To refuse to acknowledge that non-consequentialist moralities can be as distinct from one's well-being as preferences born of consequentialist concern is to badly misinterpret and underestimate the kinds of motivations that move some non-consequentialists.

If reasons flowed from only two sources, as some ethicists seem to assume, one the self-interested reasons for action and the other the detached reasons from the 'point of view of the universe', then a 'just eliminate the moral preferences' approach would be workable. In such a case we could simply eliminate those reasons which arise from the detached perspective and have left the set of preferences that are connected with one's well-being. In fact however, many reasons for action flow from intermediate positions. We are specially concerned that our country, our family, or our department do well. Such reasons are not impartially motivated but they are also not neatly connected with one's well-being.[26]

Overvold

Working out the details of a plausible exclusive preference account of well-being is a scandalously neglected task. As far as I know Overvold's account is the one tolerably developed theory of how we could separate preferences whose satisfaction contributes to a person's well-being from those whose satisfaction does not.[27] Overvold instructs us to focus on those preferences in which the agent's 'existence at t is a logically necessary condition of the proposition asserting that the outcome or feature obtain at t.'[28] He considers the criticism that one's desire that

[26] The arguments in this section offer support for claims merely asserted in Sobel (1997).

[27] But see also Darwall (1997) for the beginnings of an interesting alternative.

[28] Brandt (1979, pp. 331–2) considers Overvold's proposal, finds it inadequate, and confesses that he lacks a theory which satisfactorily draws the distinction between moral

one's spouse be happy logically implies one's existence, since one must exist in order to have a spouse that could be made happy. To avoid this problem Overvold adds the condition that 'the reason for the desire is due to one's essential involvement in the state of affairs.'[29] Overvold might instead have insisted that the logical entailment of existence cannot be contingent on the semantic formulation of the preference. With these two conditions in hand we reach the seemingly happy conclusion that the satisfaction of one's preference that one's spouse be happy does not constitute part of one's well-being, but insofar as the preference is that one be around to witness one's spouse's happiness or be the cause of it, the preference's satisfaction does constitute part of one's well-being.

Some preferences seem to get their strength through combining self-interested and non-self-interested concern. Someone attracted to Overvold's position could suggest that the extent to which we should count such intermediate concerns as constituting the well-being of the agent is determined by disambiguating the extent to which the chooser simply prefers that the group do well from the extent to which the person has her existence logically implied by the preference that the group do well. Only the strength of the latter component of the preference should be taken as reflecting the extent to which the satisfaction of the preference that the group do well is in the agent's intrinsic interests. As Thomas Carson remarks, on Overvold's proposal we would need to 'subtract the purely other-regarding element from such desires.'[30] Very roughly, the question to ask the agent is how much more important it is to her that she cause, witness, etc., the desired state of affairs over and above the importance she attaches to the state of affairs occurring. Answering this question would then, if Overvold is right, determine the extent to which the preference's satisfaction improves the agent's well-being.

In general Overvold's criterion would seem to wrongly categorize many agent-centered moral injunctions. If one's goal is to do one's duty or to keep one's promise, this would seem to implicate one's existence in a way that wanting more valuable states of affairs to come about would not. These agent-centered injunctions cannot be fully captured by the thought that one finds murder bad and, therefore, one should minimize it. Rather the agent-centered injunction is often personal (e.g., I will not kill), stemming, as Darwall has suggested, from the inside-out rather than from the outside-in.[31] But such agent-centered goals are standardly

and non-moral goods. Kavka (1986, pp. 40–4) does roughly the same thing, but Kavka claims to 'partially explicate the distinction'. Griffin (1986, p. 316 n25) admits that desire accounts have 'difficulty distinguishing between selfish and selfless action'. Raz (1986, Chapter 12) finds this difficulty overwhelming.

[29] Overvold (1982). See also Overvold (1980 and 1984).

[30] Carson (1993). [31] Darwall (1986).

taken as moral constraints on action, not personal preferences for one's own well-being. To suggest that such goals are really self-serving would have to involve arguing for a radical Nietzschian rethinking of the point of societally upheld values.

Thus the first problem for the Overvold account is that some preferences which do entail our existence are not happily characterized as correlated with our well-being. One might also suspect that some goals that do not imply one's existence can be connected to one's well-being. Consider, for example, the goal that one's estate be well managed after one's death.

But suppose we grant that such preferences are not part of the agent's well-being as Overvold's criterion implies. What is to be done with the leftover preferences? The beauty of the 'just eliminate the moral preferences' account was that it tried to find a place for everything that mattered to the agent. However, Overvold's account does not. Now, this is not an objection to Overvold's account of well-being. I have argued in Part 2 that some of what matters to us is not part of our well-being. Rather this seems to be a problem for consequentialist ethical theories that make use of Overvold's account. Any such attempt will leave some of what matters to people out of account morally. Now, by itself this does not seem problematic to me. Scanlon and Nagel have plausibly argued that certain aspects of what matters to us do not make moral demands on others. However Scanlon and Nagel identified those types of wants which were to get no moral weight and offered a rationale for their being excluded. The consequentialists who would use Overvold's account of well-being, must similarly identify the sorts of preferences which receive no expression in their account, and make a case that such preferences do not deserve the status that preferences which are part of the agent's well-being have.

There are reasons to be skeptical about the prospects of neatly separating out the set of preferences that has the right connection to well-being. For example, an ordinary moral upbringing inculcates a feeling of self-respect and assurance, if not awe, in acting morally. Such happiness would seem to be neither part of one's non-moral good since it depends on our moral concern for others, nor is it irrelevant to one's well-being for it clearly can affect one's self-respect.

The consequentialist, as we saw earlier when she was confronted with the phenomenon of belief in non-consequentialist moral systems, has no happy method of dealing with preferences which are neither for one's well-being, nor motivated by an impartial concern for the well-being of all, nor merely a combination of the two. Such concerns, which have been relatively ignored because they do not fit happily into the consequentialist's framework, constitute an important part of what matters to us. An ethical theory which lumps such concerns with self-

serving preferences or ignores them will distort or ignore much of what matters to us.

3. HOW SHOULD WE TAKE A PERSON INTO ACCOUNT MORALLY?

When one conceives of well-being as encompassing and giving expression to everything that matters to a person, there are two reasons to think that the way to take a person into account morally is to promote her well-being. The first reason is the welfarist thought that promoting the agent's well-being makes the agent's life go better and morality is crucially about furthering the true interests of persons. The second reason is the autonomy-based thought that we should allow people to decide for themselves what matters to them and how they wish to use the weight that is their due in moral reflection. But when one sees the problems with inclusive accounts these two reasons come apart.

In rejecting inclusive accounts, we create conceptual room for an agent to care about things that are not part of her well-being or to care more about them than the extent to which they further her well-being. This severely threatens the harmony of the two reasons offered above. It is no longer obvious that a person would best express what matters to her by putting forth only her well-being as demanding moral concern from the group. To insist that well-being is the appropriate object of moral concern for everyone, is to refuse to grant agents the autonomy to throw their weight in the way they think best expresses what matters to them. It is to focus on an aspect of the agent's motivational set and exclude other aspects from consideration in moral reflection, no matter how powerfully the agent identifies with the excluded aspects of her motivational system. If you do not care about rain forest preservation in the way one cares about things that are part of one's well-being, then, no matter how important rain forest preservation is to you, when you are taken into account morally there will be no direct moral pressure to preserve the rain forest. Welfarist consequentialists frequently insist that they are not paternalistically imposing their conception of what is valuable for an agent, but rather letting the agent determine for herself what she finds to be valuable for her. But in another way the consequentialist who claims that we take a person into account morally by promoting her well-being, is paternalistically restricting important aspects of what the agent cares about from receiving moral consideration.

Consider now the autonomy principle. This is the thesis that the appropriate object of moral concern must be endorsed by the agent as such given knowledge of how those preferences will be conjoined with others' preferences in moral aggregation. The fundamental idea behind the autonomy principle is that we should take people into account

morally in a way that they rationally endorse. It is an odd sense of acting for my sake which can lead to acting contrary to what I rationally want. Welfarists can console themselves that they are taking a person into account in the sense of taking that person's interests into account, but it remains obscure why this counts as adequately taking that person into account. A non-welfarist version of consequentialism which respected the autonomy principle would have less difficulty explaining why giving weight to what they do constitutes taking the agent into account morally.[32]

Imagine you are a waiter and you ask me what I want to eat. I say I want the salad. Cheekily you ask me what I would want to eat if health considerations were put to the side. I say I would then want ice cream. Then you bring me ice cream claiming not to have imposed your view of what I should eat on me because, after all, I said I wanted ice cream, health considerations aside. In this case what seems to have gone wrong is a violation of the autonomy principle. The agent did not endorse using only those kinds of preferences for that role. It was also a failure of the autonomy principle that got us into trouble in the Desdemona case. Once we see that the consequentialist must opt for an exclusive account of well-being, we see that welfarists (those who focus only on well-being in taking people into account morally) have to reject the autonomy principle.

A counter-intuitive consequence of taking people into account morally by focusing only on their well-being is that when the agent acts in ways that only affect herself, she still can be morally forbidden from doing what she rationally most wants to do. I have argued that what a person cares about can differ from what is good for her. If this is so, then an agent on a desert island would be morally obligated, according to welfarist consequentialism, to promote her well-being rather than promote what most matters to her. The autonomy principle can avoid this result. The welfarist consequentialist could try to avoid this conclusion by dropping her customary symmetrical moral treatment of self and other. That is, she could suggest that one's duty to promote one's own well-being differs from one's duty to promote the well-being of others. But this threatens not only the maximization aspect of the consequentialist position, but the anti-agent-relative reasons stance as well.

[32] Of course, some might only endorse being taken into account in ways that have their preferences dictate to the group what should be done. Thus, a strong version of the autonomy principle is obviously false. The strong version would say that we not only have to allow people autonomy about how they use the weight that they are allocated in moral reflection (call this the weak autonomy principle), but also that we have to grant people autonomy over how much weight they are allocated in moral reflection. Thus it is only the weak version which is plausible.

Autonomy pressures push us away from focusing our moral attention on well-being in different ways. A kind of autonomy pressure not yet mentioned would likely push us towards taking into consideration something narrower than an agent's well-being. Here the idea would be that we must pay attention to how an agent cares about a thing in determining whether or not its being wanted places any corresponding moral pressure on the rest of us. For example, if the agent did not take the want to put such pressure on others, we might be thought in some sense to distort the want by taking it to have such a status. If this were a kind of autonomy we wanted to respect, then we likely would be pushed towards a picture in which not even all aspects of an agent's well-being would make moral demands on others.

The autonomy principle also respects the above autonomy pressure. Once we grant the autonomy principle it is up to the agent whether to focus on basic needs, well-being, or everything that matters to her. There are perhaps conflicting pressures towards the broader and narrower objects of moral attention, but these are pressures for the agent to adjudicate, and it will not be up to anyone else to implement a univocal answer for everyone on these questions.

What matters to us and what makes our lives go well are often different things. We are forced to choose between them in deciding what matters morally. In much the same way, an exclusive account of well-being creates problems for the traditional theory of prudential rationality. The traditional idea is that rationality is simply a matter of efficiently pursuing one's ends. This is sometimes paraphrased as though it were equivalent to the thought that rationality is a matter of promoting one's well-being. The claim which connects these two theories is that what one wants is, to that extent, that which is best for one. The claim is false. Proponents of instrumental rationality must also choose whether to go along the autonomy path or the welfarist path.

But is it Consequentialism?

One might challenge the claim that the autonomy principle I have proposed deserves to be thought of as a variant of consequentialism. It might be claimed to be constitutive of consequentialism that it recommend the promotion of 'the good' and I am not especially tempted to argue that what the autonomy principle recommends that we promote deserves to be called 'the good'. Those who accept this constraint on what counts as a variant of consequentialism should think of my claim as being that a quasi-consequentialist view that respected the autonomy principle is superior to a genuine consequentialist view that does not. I am not much concerned with what we call the view I offer here. However, I do think my proposal captures many of what have been

thought to be the attractive features of consequentialism, while, undoubtedly, inheriting many of the features that have been widely criticized. Notice, for example, that the autonomy principle 1) avoids agent-centered restrictions and permissions, 2) gives no intrinsic moral importance to the distinctions between causing and allowing or intending and foreseeing, and 3) invokes a maximizing conception of one's moral obligations.

The standard way of characterizing consequentialism is to say that a moral theory counts as an instance of consequentialism if and only if it defines the good prior to the right and the right in terms of the good. This way of understanding what makes consequentialism distinctive goes back at least to William Frankena's *Ethics*, and no doubt it was partially popularized by John Rawls, in his *A Theory of Justice*, where he explicitly picks up Frankena's definition.[33] Interestingly, although Frankena and Rawls's above definition helped shape the understanding of consequentialism, they were actually defining 'teleology'.

For the sort of definition Frankena and Rawls offer above to be truly helpful, we would need to be able to place some constraints on what could count as a theory of the good. We need a way of characterizing a moral reason for choosing an option which is not an appeal to the goodness of that option if we are to coherently divide consequentialist views from non-consequentialist views using the standard method of demarcation.

The issue of isolating what can count as an appeal to goodness and what cannot has not seemed all that pressing, because the most popular versions of consequentialism recommended the maximization of well-being. In fact they had a picture of well-being in mind that looked pre-moral in the sense that it was inappropriate to complain that an intuitively immoral element could not be part of a rational agent's well-being. Well-being was assessed by these consequentialists in a non-moral manner; that is, no appeal to moral considerations was invoked in shaping the understanding of an agent's well-being. Hence it was plausible to say that they built up a notion of the good, which was just aggregate well-being, which did not rely on antecedent moral notions of what was right. For such a framework, it was plausible to claim that the consequentialist constructed moral value from non-moral value.

The simplest case where the good is defined prior to the right, and the right is defined in terms of the good, is the familiar case of uncensored well-being being the object of moral promotion. But some have wanted to screen elements from well-being before morally

[33] Frankena (1963, p. 13); Rawls (1971, p. 24). Frankena writes that teleological ethical theories, as opposed to deontological ones, claim that 'the basic or ultimate criterion or standard of what is morally right, wrong, obligatory, etc., is the nonmoral value that is brought into being'.

recommending its promotion. John Harsanyi, for example, for some time urged the elimination of nasty elements of our well-being before its moral promotion.[34] Samuel Scheffler claims that an account of the good could be distribution sensitive in the sense that it gave more weight to the interests of the downtrodden than those that are doing well.[35] James Griffin allows that the moral penetrates the prudential in the sense that 'one has not got a specification of the prudential at all without a pretty full account of what moral demands there are on us and how they are to be accommodated.'[36] Robert Goodin argues that 'our paramount goal should be to protect people's self-respect and dignity, and that these are offended by the social sanctioning of mean motives of others that take place when perverse preferences are allowed to enter the social decision calculus.'[37] Others, including David Braybrooke, have urged that the satisfaction of basic needs is what morally must be promoted in a consequentialist manner.[38] In these cases the object of moral promotion, I think it safe to say, has been shaped by moral considerations prior to the recommendation to maximize. Yet these authors claim to be consequentialists.[39]

Now exactly what counts as 'the good' having been shaped by moral considerations is somewhat opaque. But because prominent self-styled consequentialists seem to flaunt the requirement that what we are morally to promote be independent in this way, and because the most influential recent arguments for consequentialism appeal to its lack of agent-centered restrictions or the way it follows from a universal precriptivism[40] rather than its promoting of the good, one might think that the standard definition of consequentialism mentioned at the opening of this section somewhat behind the times.[41]

Is Well-Being Too Broad to Serve as the Object of Moral Consideration?

Nagel, Scanlon, Dworkin, Harsanyi and others each argue that well-being is not the appropriate object of moral concern.[42] In each case the

[34] Harsanyi (1982). [35] Scheffler (1982, pp. 70–9).

[36] Griffin (1986, p. 131). [37] Goodin (1995, pp. 145–6).

[38] Braybrooke (1987). Of course consequentialism's insistence on symmetry between self and other will be a real problem for such a view. More on this below.

[39] This is not true of Scheffler, whose book after all is entitled *The Rejection of Consequentialism*. Nonetheless, he does think that the distribution-sensitive account he offers still counts as an account of the good and not merely of the morally considerable.

[40] Parfit (1984); Hare (1981).

[41] Broome (1991, Chapter 1), has an excellent discussion of the above issues. He allows that consequentialism has largely come to be defined in terms of agent-neutrality (rather than the promotion of non-moral goodness), so he re-appropriates the term teleology for the ethical views that he is especially interested in (i.e., those that take the notion of goodness to be ethically primary).

[42] They do not seem to dispute preference satisfaction accounts of what makes one's own

idea is the same: one's well-being outstrips the appropriate object of moral concern. Morality, they contend, requires us only to respond to a subset of others' well-being, either genuine needs, non-anti-social preferences, or 'personal' preferences. They argue that well-being is too broad to serve as the appropriate object of moral concern. I have been suggesting that one could also reject well-being as the appropriate object of moral concern on the grounds that it is too narrow.[43] It is important to notice that the autonomy principle is but one method of overcoming the basic problem with well-being that is urged here; namely that it is too narrow and hence fails to give weight to the full range of our concerns. One might well not find the autonomy principle itself compelling, while yet appreciating the concerns with well-being argued for here, which animate a search for a broader object of moral concern.

Well-being sits in an unhappy middle position as the object of moral concern. If we must maintain the autonomy principle, then it is a contingent matter if people choose to be taken into account by having their well-being promoted. If we need not always respect the autonomy principle, then Nagel, Scanlon, *et al.* are surely right that the kinds of cases where we will want to stray from granting it are the one's in which we think we should focus on something narrower than well-being; cases in which the appropriate object of moral concern is basic needs or non-

life intrinsically go well, but rather dispute the use of this notion of well-being to represent what society has a duty to promote. Nagel (1986, Chapter 9) argues that some things which make my life go well or badly, such as intense pain, produce agent-neutral reasons. However, other things which make my life go well or badly produce only agent-relative reasons. He writes, on page 167, that, 'If I have a headache, anyone has a reason to want it to stop. But if I badly want to climb to the top of Mount Kilimanjaro, not everyone has a reason to want me to succeed'. For a somewhat similar view see also Scanlon (1975). There Scanlon agrees that the strength of desire, perhaps even informed desire, should not be taken to measure the extent to which others are morally bound to help. He writes on pp. 659–60 that, 'The fact that someone would be willing to forgo a decent diet in order to build a monument to his god does not mean that his claim on others for aid in his project has the same strength as a claim for aid in obtaining enough to eat'.

Harsanyi has consistently argued for the exclusion of some aspects of one's well-being from serving as the input that is used to determine the moral outcome. Once Harsanyi thought that only malevolent preferences should be excluded from each agent's input into the social decision. However, in (1988) he argues, following very closely Ronald Dworkin's (1977) position, that all one's 'external preferences' (those preferences for the 'assignment of goods and opportunities for others' – whether malevolent or benevolent) must be excluded. Again the thought seems not to be that the satisfaction of external preferences cannot affect one's well-being, but rather that such preferences do not make moral demands on others.

[43] While granting the autonomy principle does not guarantee that a broad spectrum of what matters to the agent will get moral weight, it does not exclude any aspect of the agent's motivational set from being a possible object of moral attention. In this sense the autonomy principle is broader in scope than well-being or basic needs.

anti-social preferences. Surely the primary reason to think that we should give more weight to the agent's preference to climb Mt. Kilamanjaro or to become a great pianist than her physical pain and malnourishment was that the agent cared strongly about these things. But if we think we need not always respect the autonomy principle, then surely this is because of thoughts such as this: society is not under the same kind of obligation to help a person climb a mountain that it is to help a person get enough to eat even if the former better promotes the agent's well-being. The best reasons we have to reject the autonomy principle are reasons which also carry us past well-being to narrower notions of the object of moral concern such as basic needs or non-anti-social preferences.[44]

If we are to decide that it makes most sense to take only the agent's well-being into account, we will need some reason not to allow the agent to throw her moral weight around as she informedly sees fit. One possible argument in this direction might take inspiration from an influential argument that suggests that we must not focus on subjective elements of an agent's motivational set in taking her into account morally, for this will inevitably result in an excessively demanding and unjust scheme. Giving weight to subjective concerns, it is suggested, will result in an excessively demanding scheme because a person's subjective concerns are many, whereas her urgent or basic needs are few. Further, it is suggested, focusing on the subjective elements in giving a person moral weight will be unjust, because some will develop expensive tastes and taking such preference structures into account will result in these people receiving an unjustly large slice of the social pie.[45]

But neither of these concerns is telling in this context. For we may wonder how we are to provide an agent with a fixed slice of society's pie. Should we provide it to the agent in the manner of her choosing, or should we provide it in some other way? Thus, for example, we could wonder if a fixed amount of funds must be used to provide for an agent's basic needs, or may instead be directed elsewhere should the agent so desire. In considering this question there can be no issue of

[44] But notice that the consequentialist has difficulty accepting a relatively narrow object of moral concern like basic needs. This is because the consequentialist holds that one should take oneself into account morally in the same way one takes others into account. So the basic needs consequentialist would give us very odd moral instructions on desert island cases. Seemingly such consequentialists would have to argue that we were morally required in such cases to promote our basic needs even at the expense of other things that matter more to us. Consequentialism's insistence on this symmetrical self/other treatment helps explain why well-being has been so attractive to the consequentialist as the object of moral consideration. It is because one's own well-being has seemed as tempting as that which one should promote insofar as one is acting solely for one's own sake.

[45] See Scanlon (1975) and Rawls (1982).

demandingness or unjust shares and hence no special concern about our responsibility for our subjective ends. The autonomy principle can be seen as answering the question of how to allocate a fixed slice of the pie. That is, we can consider how to take a person into account morally without worrying about how to balance the weight of different peoples' concerns. Clearly there is a different kind of rationale available for letting an agent's own concerns dictate how she make use of a given slice of society's pie than there is for letting her concerns dictate the size of her slice.

These problems about balancing are reminiscent of the problem of interpersonal comparisons of utility. We can, I am suggesting, make some progress without solving these problems. We could simply ask which ranking from each individual we should be looking at, without settling the issue of how to determine how to weigh the different rankings against each other. I suspect that any plausible solution to the problem of interpersoal comparisons of well-being that the welfarist might use to determine each agent's share could easily be modified to do the same work for the champion of the autonomy principle. (It should be noticed that the popular method of making interpersonal comparisons by taking on the motivational systems of others into a single motivational set will not lead to an interpersonal comparison of well-being, as opposed to an interpersonal comparison of what matters to people, unless welfarist restrictions on what gets weight are added.[46])

Allowing the agent to direct her moral weight however she chooses respects the agent's autonomy. The weight is, in a sense, the agent's. Shouldn't she be allowed to have it reflect what she most cares about, even if this diverges from what makes her life go best? Granting the autonomy principle results in decisions that, to the fullest extent possible, reflect what each agent cares about. Further, as we have already seen, there is no way of taking into account the full array of what matters to a person when we focus exclusively on the agent's well-being as the object of moral concern.

We are familiar with not taking the agent's word for how she should be morally represented to the group (when, for example, we did not use the agent's actual preferences to represent her morally) but this was because we had recourse to a notion of a self more in touch with what the agent really wants. But our reasons for not allowing the agent to determine her own ranking here can have no such motivation. We have been assuming all along that the agent that expresses the view that what matters to her differs from what make her life go best, is in the position

[46] See, for example, Hare (1981) for an example of this method of interpersonal comparisons being mistakenly thought to yield interpersonal comparisons of *well-being*. The mistake is common.

which is the reader's favorite for accurately determining what she really cares about.

A Problem with Autonomy?

Why focus on well-being? To the extent that we take people into account by promoting what matters to them rather than their well-being we will pass up opportunities to make peoples' lives go better. Sen, in *Inequality Re-Examined*, endorses two reasons to focus on an agent's well-being (at least in some contexts) rather than what he calls her 'agency aspect'[47] The first reason is exactly the one that Scanlon puts forward to show that we should not take well-being as the object of moral concern. Sen writes that, 'society may be seen as having a special responsibility to make sure that no one has to starve, or fail to obtain medical attention for a serious but eminently treatable ailment. On the other hand, this carries no implication that the society must take an equally protective attitude about the person's agency goal of, say, erecting a statue in honor of some hero he particularly admires ...'[48] But, as Scanlon argues, this point surely argues precisely against paying attention to well-being and towards the narrower notion of basic needs as the suitable object of moral attention. Sen is right to see here an argument against the autonomy principle, but wrong to see an argument for well-being as the appropriate object of moral concern.

The second reason Sen offers to focus our moral attention on well-being is that 'a self-sacrificing idealist who is ready to sacrifice fully his own well-being for some "cause" does not thereby make it okay for others to ignore his well-being so long as the "cause" is not harmed'.[49] This is a powerful objection to the autonomy principle. Sen himself does not think that arguments of this sort show that well-being is all that should get weight in all contexts. Rather he sees such an argument as showing that well-being can, in some contexts, be the most important object of moral consideration. But even in such cases Sen's argument is a far more powerful case against the autonomy principle than for well-being. Surely Sen is right that it would be an important objection to the autonomy principle if it led to morally sanctioning sadistic harming of such idealists on the grounds that nothing that deserves moral consideration is being harmed. This might make us think that some aspects

[47] Sen (1992, pp. 69–72). On p. 56, he writes, 'A person as an agent need not be guided only by her well-being, and agency achievement refers to the person's success in the pursuit of the totality of considered goals and objectives'. On p. 69 he considers the suggestion that 'Treating the person herself as the best judge of how she may be viewed by others, it might look as if the agency aspect would tell all that is relevant for others to know'. I consider below Sen's reasons for rejecting this suggestion.

[48] Sen (1992, pp. 70–71). [49] Sen (1992, p. 71).

of the idealist's interests deserve moral weight even when those aspects would get no weight through the autonomy principle.

But it is less than clear that we should think that those aspects of the idealist's interests which always deserve moral weight are the parts which constitute the idealist's well-being. The situation does not look nearly so objectionable if more optional or voluntary aspects of what the idealist cares about get no weight when the autonomy principle would give them no weight. If the idealist decided to throw her weight towards rain forest preservation rather than towards getting great coffee, we might well think it appropriate that her getting great coffee gets no weight at all, even if getting the coffee would make her life go better. In general, the force of Sen's case of the idealist seems to be that people's basic needs deserve consideration even when the autonomy principle would not grant them consideration, rather than the thought that well-being itself deserves moral consideration. Thus perhaps we should think that a person's basic needs always make moral demands on others, but that beyond that it is the autonomy principle, rather than well-being, that determines the shape of our moral obligations to others.

If such an emendation to the autonomy principle is to be ultimately persuasive, at least in some contexts, we must distinguish cases in which an agent fails to put forward her well-being or basic needs as objects of moral attention because she does not think herself worthy of such attention from cases like Sen's idealist. The former are cases in which the agent's preferences are not yet fully autonomous, perhaps due to the deleterious effects of having second-class status or being regarded as essentially a nurturer within the culture. In cases where detrimental societal attitudes prevent people from realizing that they are self-originating sources of claims we will no doubt question the autonomy of the preferences. In such cases we would have other routes, consistent with the autonomy principle, for criticizing the outcome.

The Autonomy Principle and Liberalism

Before concluding I want to try to ward off one possible misinterpretation of my positive proposal. It might be thought that essentially I am using the insights of liberalism, particularly Millian thoughts about the privacy of 'self-regarding actions', to argue against welfarist consequentialism. But this is not so. I have been arguing that agents ought to have a certain kind of autonomy over their inputs into the consequentialist machine, not that there is a sphere in which the agent's concerns should automatically trump other considerations concerning what the morally (or politically) acceptable outcome should be.

Liberals assert a fundamental ethical distinction between causing and allowing or intending and foreseeing. For the liberal 'self-regarding

actions' are, no doubt, meant to be situations in which one's actions cause no harm, not necessarily cases which allow no harm. Unless such an understanding of self-regarding actions is in play the liberal's notion of the self-regarding will be vanishingly narrow. Further, as I see it, this sort of liberal appeals to agent-centered restrictions. I may not violate your private moral space even to protect the private moral space of several others (though perhaps I may if the numbers get large enough).

The consequentialist denies that there is a fundamental moral distinction between causing and allowing. The consequentialist also thinks that agent-centered restrictions are paradoxical. Thus much of the liberal view, it seems to me, cannot be appropriated by the consequentialist, except, perhaps, via considerations of indirection.

My version of consequentialism does not rely on the liberal's notion of a morally protected sphere of self-regarding actions in which one should be free to make one's own choices. Perhaps it could be said that I am salvaging what can be salvaged of the liberal insights within a consequentialist framework (others would no doubt say I am perverting the liberal insight).

My desert island example offered earlier is a case in which my choice from among my feasible options is meant to have literally no impact (either in terms of causing or allowing) on anyone else's concerns. This is perhaps merely a logically possible case, but it does illustrate a point. I suggest that the consequentialist can and should accommodate the thought that the agent in this scenario ought to be morally permitted to do what she informedly most wants to do, even if this is not best for her well-being.

This is a thought that the consequentialist, I argue, can and ought to capture (since it does not involve invoking a causing/allowing distinction or agent-centered restrictions). But I argue that welfarist consequentialism fails to capture it because the agent, even if fully informed and aware, may have non-moral concerns besides her well-being.

CONCLUSION

I have claimed that well-being is not the appropriate object of moral promotion. When we get clear about what well-being is and see that many of our non-moral concerns have little to do with our well-being we see that we might prefer to throw the weight we get in moral deliberation elsewhere. My autonomy principle would grant agents the freedom to shape their input into the consequentialist calculas so as to best reflect their concerns.

A significant advantage of the autonomy principle not previously mentioned is that it requires commensuration of the input into the moral calculus for pragmatic reasons. If we are satisfied that the consequenti-

alist aggregation procedure has morally significant virtues, then the reason an agent must commensurate her concerns is that only in that way can she be taken into account in the way we deem morally best. Those who argue that well-being is the appropriate object of moral concern argue metaphysically rather than pragmatically for the commensurability of input into the aggregation procedure. Well-being, if commensurable, is not commensurable simply because it would be convenient for our favored ethical theory that it be so. The autonomy principle guarantees commensurability of the input not by finding our concerns commensurable, but by making their commensurability a prerequisite for our being taken into account morally in a philosophically preferred way.

Consequentialists have historically been committed to finding a single dimension of value that is the appropriate object of moral concern in all contexts.[50] Overwhelmingly consequentialists have turned to well-being for this all-purpose role. I have tried to point out a problem with the fixation on well-being. This inevitably involves making a case for taking something else into account morally besides well-being. Insofar as we go in for the thought that a single dimension of value must be the appropriate object of moral concern in all contexts, I think we do better to look towards the autonomy principle rather than towards well-being. I have not claimed that the resulting consequentialist position would be unobjectionable, but rather only that it is better than other versions, most notably the traditional welfarist version.

REFERENCES

Arrow, Kenneth. 1963. *Social Choice and Individual Values*, 2nd edn. Yale University Press

Arrow, Kenneth. 1987. 'Reflections on the Essays'. In *Arrow and the Foundations of the Theory of Economic Policy*. George R. Feiwel (ed.). New York University Press

Brandt, Richard. 1979. *A Theory of the Good and the Right*. Oxford University Press

Braybrooke, David. 1987. *Meeting Needs*. Princeton University Press

Broome, John. 1991. *Weighing Goods*. Basil Blackwell

Carson, Thomas. 1993. 'The desire-satisfaction theory of welfare: Overvold's critique and reformulation'. John Heil (ed.). Rowan and Littlefield Press

Darwall, Stephen. 1983. *Impartial Reason*. Cornell University Press

Darwall, Stephen. 1986. 'Agent-centered restrictions from the inside out'. *Philosophical Studies*, 50:291–319

Darwall, Stephen. 1997. 'Self-interest and self-concern'. *Social Philosophy and Policy*, 14:158–78

Dworkin, Ronald. 1977. *Taking Rights Seriously*. Harvard University Press

Frankena, William. 1963. *Ethics*. Prentice Hall

Gauthier, David. 1990. [1986] *Morals By Agreement*. Oxford University Press

Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Harvard University Press

[50] Only after an initial aggregation in which a single dimension of value is the appropriate object of moral concern in all contexts could indirect consequentialist strategies of adopting different attitudes towards different dimensions of value from context to context find a justification.

Gibbard, Allan. 1987. 'Ordinal Utilitarianism'. In *Arrow and the Foundations of the Theory of Economic Policy*. George R. Feiwel (ed.). New York University Press

Goodin, Robert. 1995. *Utilitarianism as a Public Philosophy*. Cambridge University Press

Griffin, James. 1986. *Well-Being*. Oxford University Press

Hare, R. M. 1988. 'Comments'. In *Hare and Critics*. Senor and Fotion (eds.). Oxford University Press

Hare, R. M. 1981. *Moral Thinking*. Oxford University Press

Harsanyi, John. 1982. 'Morality and the Theory of Rational Behavior'. In *Utilitarianism and Beyond*. Sen and Williams (eds.). Cambridge University Press

Harsanyi, John. 1988. 'Problems with act-utilitarianism and with malevolent preferences', In *Hare and Critics*. Senor and Fotion (eds.). Oxford University Press

Kant, Immanuel. 1956. *Critique of Practical Reason*. Trans. L. W. Beck. Macmillan Publishing

Kavka, Gregory. 1986. *Hobbesian Moral and Political Theory*. Princeton University Press

Mill, J. S. 1979. *Utilitarianism*. Hackett Publishing

Nagel, Thomas. 1986. *The View From Nowhere*. Oxford University Press

Overvold, Mark. 1982. 'Self-Interests and Getting What You Want'. In *The Limits of Utilitarianism*. Miller and Williams (eds.). Minnesota University Press

Overvold, Mark. 1980. 'Self-interest and the concept of self-sacrifice'. *Canadian Journal of Philosophy*, 10:105–18

Overvold, Mark. 1984. 'Morality, self-interest, and reasons for being moral', *Philosophy and Phenomenological Research*, 44:493–507

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

Railton, Peter. 1986. 'Facts and values'. *Philosophical Topics*, 14:5–29

Rawls, John. 1971. *A Theory of Justice*. Harvard University Press

Rawls, John. 1982. 'Social Unity and Primary Social Goods'. In *Utilitarianism and Beyond*. A. Sen and B. Williams (eds.). Cambridge University Press

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford University Press

Reath, Andrews. 1989. 'Hedonism, heteronomy, and Kant's Principle of Happiness'. *Pacific Philosophical Quarterly*, 70:42–72

Rosati, Connie. 1995. 'Persons, perspectives, and full information accounts of the good'. *Ethics*, 105:296–325

Scanlon, T. M. 1975. 'Preference and urgency'. *Journal of Philosophy*, 72

Scheffler, Samuel. 1982. *The Rejection of Consequentialism*. Clarendon Press

Sen, Amartya. 1982. 'Rational Fools'. In his *Choice, Welfare, and Measurement*. MIT Press

Sen, Amartya. 1992. *Inequality Reexamined*. Harvard University Press

Sidgwick, Henry. 1981. *The Methods of Ethics*. Hackett Publishing Company

Sobel, David. 1994. 'Full information accounts of well-being'. *Ethics*, 104:784–810

Sobel, David. 1997. 'On the subjectivity of welfare'. *Ethics*, 107:501–8

Williams, Bernard. 1981. 'Internal and External Reasons'. In his *Moral Luck*. Cambridge University Press