Deconstructing Welfare: Reflections on Stephen Darwall's *Welfare and Rational Care*

SUSAN WOLF
University of North Carolina at Chapel Hill

In his book *Welfare and Rational Care*, Stephen Darwall proposes to give an account of human welfare. Or rather, he offers two accounts, a metaethical and a normative account. The two accounts, he suggests, are somewhat supportive of each other though they are logically independent.

Substantively, I agree with by far the majority of what Darwall has to say about human welfare. His views, both about how to think about individual welfare and about what constitutes it, strike me as much more plausible than the hedonistic and preference-based views of human welfare that dominate both philosophy and popular thought, and his arguments for these views are much stronger than those of his main opponents. The issues I shall take up, therefore, having to do with the remarks in his book with which I find myself uncomfortable, may well be seen as quibbles from the perspective of a larger philosophical context. Moreover, I am not sure that when my discomforts are aired, they will turn out to reflect genuine disagreements with Darwall at all, rather than points on which I am confused or unclear, concerning details of his view that he has not especially emphasized. On the other hand, I am not sure that my discomforts are mere quibbles either. We shall see.

My concerns cluster around two claims Darwall makes in the course of elaborating his view. One set of concerns has to do with the idea that the concept of human welfare is agent-neutral; the other has to do with the claim that the attitude of caring central to his view is a natural kind. After airing these concerns, I shall raise a further question about the interpretation of the project of the book as a whole. Specifically, I shall suggest that, although the book's official point is to propose and defend a particular analysis of individual human welfare, its insights might be better captured and its opponents more readily conquered if its point were conceived somewhat differently – not as an attempt to identify the concept of welfare with that of rational care, but rather as a proposal to shift our attention away from the former and toward the latter in both our philosophical and practical thought.

© 2006 Cambridge University Press
Utilitas Vol. 18, No. 4, December 2006
doi:10.1017/S0953820806002196
Printed in the United Kingdom

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

IS INDIVIDUAL WELFARE AN AGENT-NEUTRAL CONCEPT?

Darwall's principal thesis, which constitutes the core of his metaethical account of welfare, is that a person's good is to be understood as what it would be rational for someone who cares about that person to want for her for her own sake. What is perhaps most immediately striking about this account of a person's good is that it is, to use Darwall's term, agent-neutral. That is, in saying that a person's good is to be understood, most fundamentally, in terms of what it would be rational for *anyone* who cares about that person to want for her, Darwall conspicuously refrains from giving the person whose good is in question any special say in the matter. On Darwall's view, I have no special or privileged status in determining what is or is not in my interest. I may be more privy than others to facts about myself that are relevant to identifying my interest – I may know what I am allergic to, what phobias I have, what turns me on – but then again I may not. In any case, my epistemological situation would not be enough to give me the kind of privileged status that many, if not most, theorists of human welfare have wanted to accord their subjects.

Against some familiar and important accounts of human welfare according to which agents (or subjects) do have special standing, Darwall's arguments are entirely convincing. Thus, the account that identifies what is good for someone with *what the person wants* is to be rejected because a person may well and reasonably want things that have nothing to do with her own good; she may even, like Tarzan, want something despite realizing that it involves sacrificing her own good. Thus, for example, I want my children to flourish; I want peace in the Middle East; I want a reduction in the amount of misery in the world. But whether the attainment of these ends will be in my interest – that is, whether it will foster my welfare or good – is not guaranteed.

Moreover, even a more restricted agent-relative account that identifies a person's good with what that person wants *for herself* will not yield a satisfactory analysis of individual welfare, for a person may not know what is good for her. What she wants for herself may not be what it is best for her to have. Thus, as Darwall argues, it seems that an account of a person's welfare must involve a reference not to what the person wants for herself, but to what it would be rational for her to want for herself. But now one might wonder what difference it makes whether we consider what the person would rationally want for herself or instead ask what it would be rational for anyone who cares about that person to want for her. Once one recognizes that I can be wrong about my own good, it seems a short step from there to the position that I have *no* privileged status in the determination of my own good.

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

Darwall seems to have taken that step, for, as I have already noted, the metaethical conception he proposes is purported to be completely agent-neutral.

This short step, however, is not entirely innocuous, for there is a difference between the view that one's good is to be understood metaethically in terms of what it would be rational to want for oneself (or, more precisely perhaps, what it would be rational to want for oneself insofar as one cares about oneself) and the more completely agent-neutral view that one's good is to be understood in terms of what it would be rational for anyone to want for that person insofar as he cares about her. The difference between the two views comes out in connection with substantive conceptions of human welfare about which reasonable people disagree.

Thus, for example, consider the question of whether it is in one's interest to know the truth, independently of whether that knowledge enhances or detracts from one's overall long-term happiness. Some of us care a lot about this: we place a high non-instrumental value on 'living in the real world'; we would prefer to know the unpleasant truth even at the cost of considerable suffering. But not everyone feels this way. Some would prefer to remain ignorant, or be deceived. The question we are interested in is: Which would be better for the person – to know the truth, or to remain ignorant or be deceived?

We may grant, with Darwall, that this question can in some way be identified with the question 'what would it be rational to want for that person insofar as one cared about her?' But how shall we answer it? On the one hand, we might acknowledge that it is a difficult question about which people disagree, and take this as evidence that more discussion is needed in order for reasonable people to reach consensus. Still, we might hold that whatever the answer turns out to be – whether it is better for someone to know the truth or not to know it – the answer will be the same for everyone. This seems to be the answer we are committed to if we understand the concept of a person's good as completely agent-neutral. On the other hand, we may think that this is a case in which there is no right agent-neutral answer to what is in a person's interest. And we may think that, in light of this, the views or the values of the person whose life is in question are determinative of what counts as good for her. If she cares more about knowing the truth than about being happy, and it is not irrational of her to do so, then, according to this second view, it is better for her to know the truth despite the increased suffering. If on the other hand, she is one of those who would rather be spared the suffering than have the knowledge, then it is better for her to be spared. In other words, according to this view, what is in her interest depends, at least within this limited domain, on what she would rationally want for herself. This

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

view, therefore, does give the agent, or the subject, a privileged status in determining her own good, and is therefore not completely agent-neutral.

Darwall's view, as stated, seems to opt for the first rather than the second position. However, once the possibility of this sort of case is raised (the sort of case, that is, in which reasonable people's informed preferences for themselves might not converge), it is unclear whether we should opt for this, and if we should, it is unclear why.

The kind of case I have been discussing is one for which the question, 'What would it be rational for a person to want for herself?' may plausibly be thought to have no uniquely correct agent-neutral answer. Related questions about the agent-neutrality of welfare arise in connection with cases in which there is a right answer but the agent in question does not possess it.

Many questions about what is in one's self-interest are controversial. We have already mentioned one – is it better to know a painful truth or to be happier but deceived? We might also wonder whether it is better for oneself to live a shorter but more admirable life or a longer but somewhat shameful one (should one, for example, prefer death to dishonor, understanding the question as one about one's self-interest)? Or we might wonder about Darwall's own claim, the claim of his normative account of welfare, that the best life is one in which the agent is in appreciative rapport with agent-neutral values, as opposed to a life, for example, in which the agent is simply engaged with something that satisfies and fulfills him, no matter whether rapport with agent-neutral values is involved.

I have suggested that if there is no uniquely right agent-neutral answer to these questions, we may think that the agent (or subject) whose good is in question has a special role in determining what, in connection with these questions, would be best for her. But we may think that the agent (or subject) has a special role even in cases in which there is a right agent-neutral answer. For example, I agree that something like Darwall's substantive view about the best kind of life is right. As Darwall points out, rapport with agent-neutral values has the capacity to be a source of joy and fulfillment that is not likely to be found by other means. 'What makes these pleasures loom so large in our welfare,' he writes, 'is the sense that, through them, we are connecting with things that matter' (p. 95). Indeed, I believe that connecting with things that matter has the capacity to add not only joy to one's life but meaning as well.[1] Therefore, it seems to me that people should want such rapport for themselves – it is rational for people to want this, and

[1] I argue for this in 'Happiness and Meaning: Two Aspects of the Good Life', *Social Philosophy & Policy* 14.1 (Winter, 1997).

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

it is regrettable if they do not. Nonetheless, it may be that a particular individual does not care about or value such rapport, and that his life would not be happier were he to have it. In that case, although it would be rational for him to want rapport with agent-neutral values, if he does not want it, and having it would not in fact make him happier, then perhaps it would not be better for him if he had it.

It is not clear to me that Darwall would disagree. In saying, 'the best life for human beings is one of significant engagement in activities through which we come into appreciative rapport with agent-neutral values . . .' (p. 75), he may simply be saying that there is reason for people to want such lives for themselves. He may not mean to commit himself to the claim that even if they do not want it (and it does not, unbeknownst to them, contribute to something else that they do want), it would be in their interest to have it anyway. But if that is his view, then it appears that the concept of human welfare is not completely agent-neutral after all.

Let me move on to another way in which agent-neutrality is invoked in Darwall's discussion – specifically, in his description of what caring for someone involves. According to Darwall, the perspective that is crucial to the concept of a person's welfare, viz., the perspective of one who cares about the person, can be identified with the perspective of one who sees the person as herself agent-neutrally valuable. Darwall writes, 'to one caring, considerations of welfare present themselves as agent-neutral . . . From the perspective of sympathetic concern, what benefits the cared for seems not only good for him; it seems a good thing absolutely (agent-neutrally) that he benefit in this way' (p. 15). In another passage, he writes that 'in caring about someone . . . we see her good as giving reasons to anyone, or at least to anyone who can care' (p. 24). And in yet another, he writes that 'insofar as one cares about oneself, one is committed to accepting a "hypothesis" (that one is worth caring about), and the rational dictate to promote one's own good is conditional on that' (p. 38).

These passages do not resonate with my experience of caring. When I care for someone, I do, as Darwall emphasizes, want what I take to be her good, and I want it for her sake, and not in any way that is conditional on her good having any good effect on me. That is to say, when I care for someone, her good is for me a final end. I want it directly. That something will benefit her is itself a reason for me to do it. But it is not at all obvious to me that I accept any hypothesis about the object of my care being worthy of care, and I do not seem to commit myself to any view about what reasons others may have to promote her good.

It is not that I think, as for example Harry Frankfurt seems to, that it is of no interest whether the objects of my care are worth caring

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

about.[2] To the contrary, I believe it is desirable to care about things that are worth caring about rather than things that are not. If I came to believe that the objects of my care were not worth caring about, it would signal a problem for me. Nonetheless, there are many caring people who, I suspect, don't think one way or another about whether the objects of their care are worthy of it, and it is easy at least to conceive that one care about someone whom one explicitly thinks is not worth caring about. In such a case, one wants that individual to fare well for her own sake, even though one does not think this would be good absolutely. In any event, if you ask someone why she wants a particular person to fare well, the fact that she cares about that person (or, alternatively, that she is the person's mother or friend, from which it can be inferred that she cares about her) would seem to name a sufficient reason. To talk about her worthiness would be, as in Bernard Williams's famous example, to offer one thought too many.[3]

IS DARWALL'S CARE A NATURAL KIND?

Darwall's identification of a person's good with what it would be rational for one caring about that person to want for her for her own sake is intended to be a metaethical analysis of the concept of individual welfare. In other words, Darwall wants to argue that the concept of welfare is parasitic on the phenomenon of care for someone, and cannot be understood independently of it. In this, Darwall stands opposed to what might have seemed the common-sense position – namely that the idea of a person's welfare, far from being dependent on the concept of care, figures into that concept insofar as part of what it is to care about a person is to have his or her welfare as a final end. Darwall defends his view against the common-sense alternative in the book's first chapter (see especially pp. 8–12).

However strong the reasons are for thinking that the concept of welfare is derived from the phenomenon of care, Darwall recognizes that one can defend this claim only if the phenomenon of caring can be identified independently of any appeal to the notion of welfare. To meet this requirement, Darwall argues that the attitude of caring for someone (or something) is a natural psychological kind. My second set of questions has to do with this claim. Roughly, I want to raise some doubts about whether the phenomenon of caring for someone *is* a natural kind – especially, whether it is exactly one kind – and whether,

[2] I discuss this in 'The True, the Good, and the Lovable: Frankfurt's Avoidance of Objectivity', *The Contours of Agency*, ed. Sarah Buss and Lee Overton (Cambridge: MIT Press, 2002).

[3] See Bernard Williams, 'Persons, Character, and Morality', in *Moral Luck* (Cambridge: Cambridge University Press, 1981).

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

if it is, it really yields a notion that corresponds to our, or even Darwall's, notion of individual welfare.

For much of my reading of Darwall's book, I took it for granted that a (normal) parent's care for her child offered a paradigm case of the kind of caring to which Darwall wants to appeal. Certainly, a parent's care for her child includes her wanting the child to fare well, and wanting it for the child's sake. Darwall himself uses the parent–child relationship in an example, pointing out that insofar as the parent cares for the child, she wants what is good for the child (say, a trip to the dentist, or the consumption of vegetables) rather than what the child wants for herself (namely, to get as far away from both the dentist and the vegetables as possible). Familiar as the attitude a parent typically has for her child is, we might wonder whether the attitude is literally a *natural* kind. That is, it seems reasonable to question how much of parents' attitudes to their children are products of their historical and cultural circumstances. Still, I want to put that question aside, and focus instead on whether we really can or should identify the child's welfare with what the parent who has such an attitude toward her child would rationally want for her.

One problem with this proposal has to do with the difficulty of separating what a parent wants for her child for the child's sake from what she wants for the child for some other reason. Parental love – even good, healthy parental love – is hardly a model of love that is disinterested or selfless. Because of this, there may be no way of deciding whether what a parent wants out of love for her child should be identified with what she wants for the child's sake that does not beg the question.

For example, parents often want their children to have certain qualities or experiences that give them a connection to their heritage or to the parents themselves. A parent may want, for example, that his child grow up with a sense of his Jewish or Irish or Chinese identity; the parent may hope that the child have a philosophical turn of mind or a musical bent. Why do parents want this sort of thing for their children? Is it for the child's sake? for the parent's sake? for the grandparents' sake? Speaking for myself, as one who does want such things for my children, I must confess that none of these answers seems obviously right.

A further difficulty comes out when we focus on some other things parents often want for their children, which have nothing, or at least nothing obvious, to do with wanting their children to feel or to have a special connection to themselves. For example, I want my children to be morally upstanding, if not outright admirable. I want them to be talented, and even physically attractive. Quite generally, I want them to be possessed of excellences, and I rejoice in the excellences they

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

already so evidently have. Of course, there may well be an element of self-interest in these desires as well – the more excellent they are, the more pride I can take in them, and that is a pleasure for me. But it also seems quite natural to think that I and other parents want such things mainly for our children's sake. The question is, insofar as we do want these things for their sake, does that imply that we think these things will contribute to their well-being?

Here is the problem. If I reflect on it, I find myself quite uncertain about whether my daughters' good looks, for example, will enhance their well-being. Though I agree with Aristotle that one cannot be maximally happy, or *eudaimon*, if one is repulsive, it is far from clear that in our culture being beautiful is good for a woman. Yet it is not obviously irrational for a parent to rejoice in her child's beauty, and, if she does so, it would seem to be out of love for her child – the child's beauty is something she seems to be happy about for her child's sake. Consider as well the parent of a child who is serious about athletics. What should the parent want for her child? Again, if one steps back and asks what would be best for the child from the point of view of the child's welfare, the answer is not obvious. Competitive sports, at least in America, are fraught with all sorts of pressures and expose the participants, especially at the most selective levels, to both physical and psychological risk. Sensitive, enlightened, and rational parents must worry about these pressures and risks. But this seems compatible with their nonetheless wanting their children to have the talent, the skill, the discipline, and the good fortune to succeed in their sport, and to want it for their children's sake. Does this mean that it is in the child's interest to succeed? That it is better for the child (in the sense of improving the child's welfare)? According to Darwall's analysis, that is exactly what it means. But there is room for doubt.

These examples have rested on the assumption that a parent's love, and care, for her child constitutes a good paradigm for the natural kind of caring to which Darwall's account appeals. If we took some other relationship of intimate love as a paradigm of care – the love one might have for a life partner or a sibling or a best friend, for example – the issues might slightly shift. A different issue arises from the consideration of non-intimate relations. For it seems we can care for people we hardly know, or even for complete strangers (e.g., tsunami or hurricane victims who have lost their families, their towns, their limbs), and when we do so, we want their good for their own sake, too. But if this attitude that we sometimes have towards strangers and slight acquaintances is a natural kind of attitude, is it the *same* natural kind as the attitude we have toward our own children and loved ones? If not, which is the attitude that we should take as our ground for the concept of welfare?

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

I have already pointed out difficulties with choosing the attitude we have toward our children. If we opt instead for the beneficent attitude we sometimes have towards strangers, we face a different problem. Specifically, it seems to me that beneficent attitudes toward strangers do not typically extend as far as our attitudes toward more intimate loved ones. We care that they enjoy the basic goods, perhaps, of food, shelter, health, and companionship, that they not suffer horribly or be victims of premature death. But we may not care about whether they get to take piano lessons or travel or, more generally, whether they live lives of significant engagement of appreciative rapport with agent-neutral values. Nor is it obvious that this marks a failure of rationality in us.

If we identify individual welfare with that which it would be rational for one who has the attitude typical of our beneficent care for strangers to want for the stranger for her own sake, then it seems we are confronted with a trilemma regarding the adequacy of Darwall's analysis of welfare. If, on the one hand, we do not think it irrational for a caring person to care only that the cared-for be possessed of basic goods, then the non-basic goods (the rapport with agent-neutral values, for example) may not be a part of the concept of welfare. If, on the other hand, we stick to the idea that these goods are part of welfare, then either we must admit that it may be rational to care for some parts of a person's welfare but not all, or we need to provide an argument for why it would be irrational for a person to care about another's basic well-being (health, freedom from torture, and so on) and yet not care about the person's maximum welfare. The first of these latter two alternatives would loosen the identification between an individual's welfare and what it would be rational for someone to care about for that individual for his own sake. Since this directly conflicts with Darwall's main thesis, he may prefer the second. But I do not see how one could provide an argument for the irrationality of caring about some but not all aspects of a person's well-being without relying on the concept of welfare. In that case, however, the concept must be independently available, and cannot be a product of the phenomenon of care.

HOW IMPORTANT IS THE CONCEPT OF WELFARE ANYWAY?

The trilemma just sketched casts doubt on what seems rightly described as the main thesis of the book – viz., the claim that the concept of welfare just is the concept of what we would rationally desire for someone insofar as we care for her. As such, it may appear to strike at the very heart of Darwall's position, and to be an expression of a deeply antithetical philosophical point of view. As I mentioned earlier,

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

however, I am in sympathy with by far the majority of what Darwall has to say in the book. How can this be?

The answer, I believe, is that what constitutes the heart of a philosophical position is subject to interpretation. I shall conclude by suggesting a recasting of Darwall's position in a way that I believe captures and endorses most of the values and insights that motivate Darwall's view even as it gives up what is superficially Darwall's central thesis.

Specifically, it seems to me that rather than try to defend the identification of welfare and rational care, a sympathetic reader of Darwall's book would do better to de-emphasize the importance of the concept of welfare altogether. By shifting attention away from the concept of welfare, and toward the question of what it would be rational for a person who cares about someone to want for him, one could avoid some of the controversies brought up by the commentaries in this volume, while achieving the same practical and normative ends.

Thus, for example, I believe that many people will agree that it is rational to want one's loved ones to be morally admirable, and, more generally, to be possessed of excellences, and to want this, in some sense, for their own sakes, rather than for the sake of the one caring or for the world at large. Yet they may be uncertain about whether these virtues actually benefit the ones they love, whether, in other words, it is in their self-interest to be morally and otherwise admirable. What difference does it make whether we settle this controversy? What is important, it seems, is that those who care about someone (including the person himself) figure out what, as ones who do care about the person, they should want and do for him. If bypassing the issue of whether something contributes to his welfare allows us to do this more easily, it is hard to see how anything will be lost.

Darwall's own reasons for wanting to insist on the identification – reasons that became clearer to me as I listened to the discussion during the symposium in which an earlier version of these commentaries were presented – have to do with his view that the concept of welfare is a normative concept. 'It seems possible', he writes 'for two people who care about someone, S, to coherently disagree about whether something, X, is good for S, even though they agree completely about all the non-normative facts concerning X and S' (p. 11).[4] But this can only be so if the concept of welfare (of S's good) is a normative, as opposed to a purely descriptive, concept. If one accepts this conclusion, the identification of the concept of welfare with the particular normative concept that Darwall formulates has considerable appeal.

[4] The example he offers as a test case concerns S's illusory belief that S's novel has sold 10,000 copies when in fact it has sold only 12.

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

Though I agree with Darwall that the kind of disagreement he refers to *seems* coherent, and that its coherence would imply that the concept of welfare is normative, I am wary of putting too much pressure on this kind of example. Other considerations and other examples, drawn from this and the other commentaries in this volume, may incline us in the opposite direction, to the idea that the concept of welfare is descriptive or non-normative after all. Thus, for example, the association between the concept of a person's welfare and what contributes to his overall happiness is very strong. Defending the claim that something contributes to someone's welfare despite its making no contribution to his long-term happiness is an uphill battle – steeper, I think, than what for Darwall must be the equivalent claim that it would be rational for someone who cares about someone to want something for him despite its making no contribution to his happiness. Moreover, it seems plausible that the concept of welfare when applied to plants and non-human animals is non-normative. What is good for a plant is whatever makes and keeps the plant healthy, perhaps, and this seems to be a botanical matter. Understanding the concept of human welfare as similarly descriptive would allow us to treat the concept of welfare applied to any object the same way.

I do not wish to settle the matter of whether welfare is a normative concept here. Our ordinary uses of the term may be inconsistent, and so its use in philosophical contexts may require an artificial convention to decide to use the term one way rather than another. My point is simply that, although Darwall casts his work as a metaethical theory about the concept of welfare, many of his most substantive and controversial claims can be made in a different vocabulary. If we want to use 'welfare' as a normative term, then perhaps we should identify it with the concept of what it would be rational for someone who cares about someone to want for that person. If we want to use 'welfare' instead as a non-normative term, however, then we can make many of the same points by being clear that not everything we want for a person insofar as we care about him must be wanted because it is thought to contribute to his welfare or well-being.

In *What We Owe to Each Other*, Thomas Scanlon argued 'that the concept of well-being has less importance . . . than is commonly supposed. From a first-person point of view . . . the concept of well-being plays little role.'[5] He suggests, however, that the concept may have a greater significance from a third-person point of view, such as that of a benefactor. On the recasting of Darwall's view that I have offered above, we have some reason to doubt that the concept has all that much

[5] Thomas Scanlon, *What We Owe To Each Other* (Cambridge, MA: Harvard University Press, 1998), p. 142.

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press

significance even there. As Darwall has shown us, however, the concept of 'what it is rational, as one who cares about someone, to want for him' *is* significant – whether we identify that concept with the concept of welfare or not. This is a different concept from what that person wants for himself and from what it is rational for us to want for him, period. Often it is just the concept we need to describe and discuss what to aim for.

susanw@email.unc.edu

https://doi.org/10.1017/S0953820806002196 Published online by Cambridge University Press