

THOMAS HURKA

A KANTIAN THEORY OF WELFARE?

Two main foundations have been proposed for the side-constraints that deontologists think make it sometimes wrong to do what will have the best effects. Thomist views agree with consequentialism that the bearers of value are always states of affairs, but hold that alongside the duty to promote good states are stronger duties not to choose against them.¹ Kantian views locate the relevant values in persons, saying it is respect for persons rather than for any state that makes it wrong to kill, lie, and so on.² The central innovation of Stephen Darwall's *Welfare and Rational Care* is to extend this Kantian idea from side-constraints to the concept of welfare, or of what is good for a person.³ As a good-to-be-promoted, welfare is usually understood as located in states of affairs. Darwall agrees that a person's welfare involves her being in certain states, but argues that the value in these states derives from her value as a person. More specifically, his "rational care" theory of welfare equates a person's welfare with those states it would be rational to want for her insofar as one cared for her for her sake, so an attitude to her is primary and to her states is derivative. Whereas standard theories take the concept of welfare to come first and define care as a desire for that, Darwall reverses this ordering.

Alongside this account of the metaethics of welfare, Darwall defends a substantive or normative theory about which states are in fact good for persons. He proposes that welfare consists primarily in "valuing activities," ones that involve appreciation of objects of independent worth; such activities include playing music and raising children. This theory, too, has Kantian elements, since the appreciation is of objects

rather than of states of affairs and is also of their worth. I will examine Darwall's two theories of welfare, asking of each how Kantian it is and how far, when it is Kantian, it is plausible.

1. WELFARE AND CARE

Welfare is a very specific value-concept, which does not include everything worth promoting. Most obviously, the elements of a person's welfare must be states of her. Even if she desires something outside her, as in an example of Darwall's a woman wants a war-ravaged city rebuilt (pp. 43–45), its obtaining cannot benefit her unless it affects her in further ways. Nor do all good states of her contribute to her welfare. On a retributive view, a vicious person's suffering pain is good, but it is not good for him; nor, according to Darwall, is a person's doing her duty good for her (p. 30). While some deny that perfectionist goods such as knowledge can contribute to welfare,⁴ others disagree, so there are disputed areas. But all views place some conceptual limits on what can be good for a person, and one task of a metaethics of welfare is to explain why these limits come where they do.

Other theories attempt this by directly restricting the states relevant to welfare, but in Darwall's theory the explanation must come from facts about care. If certain things cannot, conceptually, contribute to someone's welfare, it must be because it could not make sense to want them from concern for her. Darwall does make claims of this sort, but they are not very persuasive. Consider the woman who wants her money spent on rebuilding a damaged city rather than on saving her life. Darwall says that while respect for her may lead us to prefer the rebuilding, care requires us to prefer her life (p. 44). But is this true? Why could it not be a form of care as well as respect to want whatever she wants? Or consider the doing of duty. Why could care for a person not express itself in wanting that she act as she ought? At the least, Darwall's claims are less obviously true than the ones about welfare

they are meant to imply, such as that states outside a person cannot benefit her. And how can the less obvious explain the more compelling? It is hard not to suspect that in considering Darwall's cases we read our intuitions about welfare back into our judgements about care, counting something as care only when it is directed at what we already understand as welfare.

A similar point applies to substantive claims about what welfare consists in. Imagine that someone says he cares for you deeply and therefore wants you to suffer as an end in itself. Now, desiring your suffering is clearly inconsistent with caring for you, but can we explain why? On Darwall's theory we cannot: that a desire for suffering is inconsistent with care is an irreducible fact. But there seems an obvious explanation: wanting your suffering cannot express concern for you because suffering is bad for you, whereas wanting your happiness can because happiness is good. Again judgements about welfare seem prior to judgements about care rather than vice versa.⁵

These difficulties are exacerbated by a difference between Darwall's theory and the Kantian view it is modelled on. As Darwall recognizes (p. 14), Kantian respect is for persons as having certain properties, typically rationality and freedom, and these properties explain why the side-constraints they ground have the content they do, such as forbidding acts that subvert others' rationality. But Darwall never ties care to any specific properties of persons. (At one point he says caring for someone involves relating to her as a being with a welfare (p. 14; see also 69, 70), but that suggestion would clearly make his theory circular.) This, first, makes it difficult to see how care can generate the conceptual and substantive limits on welfare. If care were directed to a person as having certain properties, any states unconnected to those properties could be excluded from her welfare. But if care has no such focus, where can the limits come from? Second, this feature of Darwall's view makes it hard to understand what care as he conceives it is. If care is for persons apart from any of their properties, is its object a featureless substrate or Lockean

I-know-not-what? If so, how is care even possible? In personal relationships we often say we love a person “for herself,” meaning that although we admire her for certain properties we would not abandon her for someone with the same properties to a higher degree. But the best explanation is that we love her in part for properties no one else could share, namely those of having participated with us in a shared history.⁶ And this explanation is not available to Darwall, who thinks we can care for people we have never met. So if care is not directed to any of a person’s properties, what is its object?

Darwall has various positive arguments against reducing care to a desire for a person’s welfare. One says it is possible to desire another’s good on a whim or by fancy, without caring for the person herself (p. 2). But surely one can also care on a whim, for example, be struck by the curve of a person’s nose and love her intensely for a short period. Darwall sees care as involving a perception of the other as valuable independently of one’s caring; this too, he argues, distinguishes care from any whimsical desire (pp. 70–71). But can one not make whimsical evaluations, for example, decide suddenly that biology is categorically valuable and then equally suddenly change one’s mind? Darwall also argues that care involves a whole complex of emotions and sensitivities, so that if someone I care about is suffering, “I will be disposed to emotional responses, for example, to sadness on his behalf, that cannot be explained by the mere fact that an intrinsic desire for his welfare is not realized” (p. 2). But reductionist theories of welfare can agree, taking the term “care” to apply only when one has the full range of positive attitudes, including desiring, pursuing, and feeling appropriately pleased, by all or most of another’s good. Care does seem distinct from a single desire for part of a person’s welfare; it is not so obviously distinct from this full range of attitudes.

Darwall also directly addresses the objection that care cannot be understood except as directed to someone’s welfare. This circularity objection fails, he says, if care is a natural psychological kind. Just as we can talk of water without having a definition of “water,” so we need not define care if

it has a similar status (p. 50). He then gives a rich account, informed by much empirical psychology, of how sympathetic concern differs from related states such as empathy and what he calls proto-sympathetic empathy (pp. 54–72). But it is unclear how this discussion bears on the circularity objection. Although we need not know water's nature to refer to water, it does have a nature: it is H₂O. And it must have a nature if our uses of "water" are to have a common referent. But then care, too, must have a nature if it is a natural kind, and we can inquire what that nature is. It could in principle be neurological, but Darwall never suggests this possibility and it is more consistent with his discussion to see care as irreducibly psychological, involving an intentional attitude to some distinctive object. Then the question arises, as before, whether this object is not a person's antecedently understood welfare. The analogy with natural kinds may even tell against Darwall's view. He cites the psychologist C. Daniel Batson, who defines care as a motivational state whose goal is "increasing the other's welfare" (p. 67). The standard view is that the natures of kinds are determined by empirical scientists, to whose discoveries the rest of us must defer. If an empirical psychologist relates care to an independent concept of welfare, should Darwall not follow suit?

A metaethical theory of welfare must explain the concept's relativization, how it concerns what is good for a person rather than simply good. Darwall attempts this in a novel way, by deriving claims about welfare from claims about care for a person. I have questioned this Kantian move, on the ground that we cannot understand care except in terms of a prior concept of welfare. But there are other interesting aspects of Darwall's theory, for example, about the normativity of welfare.

2. WELFARE AND NORMATIVITY

Welfare is something we have reason to promote, in either or both of ourselves and others. Some philosophers treat the concept of welfare as descriptive, either equivalent to pleasure

or desire-fulfilment or *sui generis* but still non-normative.⁷ On this view any requirement to pursue welfare must come from outside the concept itself. But Darwall sees welfare as intrinsically normative, so the claim that something is good for a person itself provides a reason to promote it. The reason, however, is of a distinctive type, giving what Darwall claims is a superior account of welfare's normativity.

The rational care theory says a person's welfare is what one ought to want insofar as one cares for her. For Darwall the "ought" here is hypothetical, connecting care and desire for someone's good in the same way Kant's hypothetical imperative connects desire for an end and desire for the means to it. He understands Kant's imperative not as a conditional with an imperative consequent but as a command to make a conditional true, one that can be satisfied either by not desiring an end or by choosing the means to it.⁸ Similarly, claims about welfare forbid as incoherent the combination of caring for someone and not desiring her welfare. This makes the claims intrinsically normative, but in a distinctively hypothetical way.

As so understood, claims about welfare do not provide simple reasons to act, since they can just as well be satisfied by not caring for anyone. They would generate such reasons if they were supplemented by categorical demands to care, as in Kant's theory a general hypothetical imperative is supplemented by categorical demands to pursue one's own perfection and the happiness of others.⁹ Darwall sometimes suggests this move, saying "We have reason to care about our own good . . . because we have reason to care about ourselves" (p. 53; see also 83). But elsewhere he rejects it, saying "practical reason includes no intrinsic requirement that we care either about others or about ourselves" (p. 37). And the move would undermine the superiority he claims for his theory. He says that metaethical theories equating welfare with pleasure or the items on an objective list cannot explain the normativity of welfare, since we can always ask why something's being pleasant or on the list creates reasons (p. 45). His objection seems to be that these theories require external

and ungrounded requirements to pursue the constituents of welfare. But if his theory requires an external requirement to care for persons, it is in essentially the same boat.

So where does Darwall think our reason to pursue welfare comes from? He denies that this reason depends on the fact of our caring, but often says it is conditional on a hypothesis we accept in caring, namely that the object of our care is worth caring about (pp. 8, 37–38, 48, 70–71). But this just invites the question whether this hypothesis is true. If it is, we are back at an external, categorical requirement to care. If not, why bother making our actions consistent with a false hypothesis? At two points Darwall suggests that we can believe the hypothesis is true on the basis of an argument in his earlier book *Impartial Reason* (p. 114n7, 116n39).¹⁰ But that argument, too, is hard to understand. It says our reason to care for persons depends on our capacity to care for them. But we also have a capacity to hate persons, which presumably does not generate a reason to hate. And what can explain the difference other than that hate fails to recognize an independent value persons have that grounds an independent reason to care?

I do not see why Darwall does not affirm a simple categorical requirement to care for persons. This would undercut his claims for his theory's superiority, but this is no loss if every normative view must make some ungrounded claims. And it would combine with his hypothetical-imperative analysis of the normativity of welfare to yield an elegant account of the source of our reasons to pursue welfare that parallels Kant's account of our reasons to pursue ends. Though not distinctive in its type of foundational reasons, it would be distinctive in its structure. A categorical requirement to care would also remove an incentive for some puzzling things Darwall says about what care is; I now turn to that topic.

3. IS CARE KANTIAN?

Darwall understands care in a distinctive way. Not only is it directed initially at persons rather than states, but it involves

seeing them as worth caring about or valuable, so one has categorical reason to promote their good (pp. 8, 38, 48). Moreover, the value one sees is agent-neutral, giving everyone reasons to care: “caring involves seeing the cared for as worthy of care and, consequently, involves seeing their welfare as giving (agent-neutrally) reasons to anyone” (pp. 48–49; see also 53, 70–71, 83).

This is a demanding conception of care, in particular because it excludes two weaker possibilities. One is that care involves seeing a person as valuable but only agent-relatively, or for oneself. One’s thought in caring need not be that others lack a reason to care; one just sees a reason for oneself to care without considering whether others have similar reasons. The second possibility is that care involves no evaluative thoughts at all, either agent-neutral or agent-relative. Most of us think it possible to desire a state of affairs either because one sees reason to do so or apart from any sophisticated thought about reasons. Thus, one can desire knowledge either because one thinks it is good or because one is simply curious. The parallel possibility is that one cares for a person without any thought about her value but from simple emotion. Darwall excludes this possibility by insisting that care involve evaluative thoughts, and excludes the agent-relative possibility by insisting that these thoughts go all the way to agent-neutrality.

Darwall does not argue for his conception of care so much as assert it. But why must care involve the extra elements he requires, rather than taking either of the two simpler forms? This question is especially pressing since people do not in fact care agent-neutrally. They do not show equal concern for all persons, but care more for their family and friends than for strangers, and even feel it is right to do so. Darwall recognizes this, saying “individuals may have more reason to care for themselves or close relations than they do for strangers,” though he adds that “neither is possible without the (third-person) capacity to care” agent-neutrally (p. 53). But imagine that someone cares to some degree for everyone but more for her child. Her extra concern for her child cannot be

agent-neutral: she cannot believe that her child matters more from the point of view of the universe. It must instead involve either a thought about value just for her or no evaluative thought at all. And this raises the question whether Darwall's added remark about the third-person capacity is true. If care can be partly free of agent-neutral evaluations, why can it not be entirely free? Why can it not involve only agent-relative or no evaluative thoughts? I am not questioning whether there is an agent-neutral reason to care; there surely is. But Darwall gives no reason to believe this reason must figure explicitly in the content of every instance of care.

In addition, Darwall's conception of care will strike some as positively unattractive. Bernard Williams says that if a husband's thought in saving his wife rather than several strangers is "She's my wife, and in situations like this one is permitted to save one's wife," then he has "one thought too many"; Michael Stocker similarly says he would not think much of a friend whose only motive for visiting him in hospital was to do his duty.¹¹ These comments concern motivation by duty, but they extend to motivation by evaluative thoughts in general; Stocker would surely be no more impressed by a friend who visited because he saw Stocker as agent-neutrally valuable.¹² The objection expressed here does not apply in all contexts, but it seems specially relevant to care for persons, where many will join Williams and Stocker in preferring emotions that relate directly to their object rather than indirectly through an evaluative thought. In requiring such a thought for care, therefore, Darwall requires it in a context where many find its presence positively objectionable. The requirement is another Kantian feature of his theory, since Kant famously held that only motives expressing evaluative judgements have moral worth. But it follows Kant on a point where many find his view repellent.

Darwall may have an incentive for his restrictive view of care in his account of the normativity of welfare. He believes, plausibly, that there is an agent-neutral reason to pursue everyone's welfare, and this reason would follow directly if there were a categorical requirement to care for everyone, as

I urged in the previous section. Moreover, the reason would follow even if care were understood permissively, so it can involve only agent-relative or no evaluative thoughts. But Darwall is reluctant to affirm a categorical requirement to care, preferring to derive the reason to pursue welfare from something within the caring person's mind, namely a hypothesis she accepts in caring. This means the agent-neutrality of the resulting reason must figure in her mind, so the hypothesis becomes one of agent-neutral value. And this generates all the difficulties I have surveyed, of not matching how people actually care and of making care unattractively intellectualized. It also threatens to generate an infinite regress. For Darwall, caring involves appreciating a reason to care. But unless people are required to accept a false hypothesis, the second instance of care must also involve appreciating a reason, so caring involves appreciating a reason to care based on appreciating a reason to care based on appreciating ... — a sequence that cannot be completed.¹³ All these problems would disappear if Darwall affirmed a simple categorical requirement to care for all. While retaining his distinctive view that care is directed at persons and still deriving a reason to pursue everyone's welfare, he could then allow additional forms of care, for example, simple emotional ones. That, I believe, would result in a more realistic and attractive picture of what care is.

4. THE SUBSTANCE OF WELFARE

Darwall also proposes a normative theory of which states are good for persons. He claims that the primary constituents of welfare are activities (in a broad sense) that involve appreciating objects of independent worth: these include playing or listening to music, which appreciates the value of beauty, seeking knowledge, and raising children. To have their full value these activities must combine two features. Their object must be genuinely valuable; caring about trivial ends does little to improve one's life. And one must positively appreciate the object; engaging in a worthy activity without love for it

has little value. On their own these objective and subjective features contribute some to welfare, but by far the greatest value comes when they are combined.¹⁴

Darwall does not argue at length for this theory, saying simply “When I ask myself what kind of life it makes sense to want for my children, it just seems obvious to me that it is a life in which they engage in activities whose merit and relation to worth they themselves appreciate” (p. 103). One may question whether appreciating worth is good *for* a person rather than simply good; the issue here is the conceptual limits on welfare discussed above. That aside, Darwall’s normative theory is in broad outline attractive: a life that combines objectively valuable activities with appreciation of them for themselves does seem intuitively desirable.¹⁵ But he again gives his theory a Kantian cast, which again in my view detracts from its appeal.

Jus as in Darwall’s metaethics the primary attitude is to a person, so in his normative theory the appreciation with most value is of an object or activity rather than of any state including it; thus, a musician should appreciate his playing itself rather than the fact that he is playing (pp. 87–93). We can again question whether we understand appreciating an activity except in terms of believing that the state where it exists is good or desiring that state. But even granting Darwall his psychology, we can question his normative preference for attitudes to activities over attitudes to states. Consider the following remarks from a rock-climber: “It’s exhilarating to come closer and closer to self-discipline. You make your body go and everything hurts; then you look back in awe at the self, at what you’ve done, it just blows your mind.”¹⁶ The climber’s primary delight is in something propositional – that he completed a difficult climb – yet it is hard to see how that makes his activity less valuable. Or consider the hockey player Raymond Bourque, who extended his career by a year so it could contain at least one Stanley Cup. What is remotely objectionable about his motivation? Darwall’s position is made more difficult by the fact that he takes appreciation to be of objects for the properties that give them worth (p. 88). So imagine

that the climber appreciates his climbing for its difficulty. Since this property is shared by other climbs, how different is his attitude from one of pleasure that he completed some difficult climb? If he has made a particular climb, he may feel specially attached to its features; thus, he may be specially fond of the particular route he took. But these feelings will be secondary to his primary satisfaction that he did some difficult climb. If they were not, and he cared more about his particular climb than about the fact that he did something difficult, his attitude would seem fetishistic, caring more about features of the climb that have no worth than about the ones that give it value.

In addition, Darwall insists that the appreciation that makes for welfare be of its object as having worth (pp. 76, 89–93). This again reflects the Kantian view that only attitudes expressing evaluative thoughts can be good, but now avoids some difficulties that faced Darwall's earlier application of this view. Since he now uses two value-concepts – what the object has is worth, what appreciating it makes for is welfare – there is no infinite regress. And there seems no room for a Williams–Stocker objection: climbing rocks because one thinks it a worthy activity is not positively objectionable. But we can still find Darwall's view too restrictive for denying significant value to appreciations involving only simple emotion. We need to imagine someone who loves an activity for its good-making properties but without thinking of them as good-making, for example, a climber who loves climbing for its difficulty without connecting that property to ideas of worth, or a scholar who pursues knowledge from simple curiosity. Darwall denies that these attitudes have the special merit of “valuing activity,” but this seems an excessively intellectualized view. Why should pursuing a worthy object from an emotion directed to its worth-giving properties be any less good if the emotion is not itself about worth?

Finally, Darwall makes the appreciation of worth a vital component of welfare. He does not go as far as Kant, who holds that only a good will is unconditionally good; he allows that an objectively excellent activity unaccompanied by

appreciation has some value. But he thinks the addition of appreciation makes the activity vastly better, and we can wonder whether that is true. Consider someone who engages in an objectively excellent activity only as a means, say, a scientist who makes fundamental discoveries only in order to win a Nobel Prize or a hockey player who plays only for money. It would clearly be better if these individuals cared about their activities intrinsically, but would it be vastly better? That seems implausibly high-minded. The excellence of the activity seems the main value, the appreciation of it less important. That certainly seems to be Aristotle's view. Darwall credits the ideal of "valuing activity" to Aristotle, but Aristotle holds that pleasure in a good activity completes it "as the bloom of youth does ... those in the flower of their age," that is, as an appropriate addition but not the principal good.

The idea that valuing value itself has value is deeply attractive and was held by early 20th-century philosophers such as G. E. Moore, Hastings Rashdall, and W. D. Ross. But they did not interpret the idea in Darwall's heavily Kantian way. They allowed that attitudes to states of affairs can have high value; insisted, even vehemently, that these attitudes need not involve evaluative thoughts; and at least occasionally held that an attitude to an object has less value than its object. The result was a less restrictive and therefore more plausible theory of individual good than Darwall's Kantian one.

5. CONCLUSION

Though it is short, *Welfare and Rational Care* is a rich book, with many novel proposals about welfare. (If only more philosophy books combined these traits!) I have suggested that many of these proposals are unified by expressing Kantian ideas, and have challenged them on that basis. But whether or not the challenges succeed, Darwall's is an original book that adds a strikingly new theory to the list of competing theories of welfare.

NOTES

¹ John Finnis, *Natural Law and Natural Rights* (Oxford: Clarendon Press, 1980).

² Alan Donagan, *The Theory of Morality* (Chicago: University of Chicago Press, 1977).

³ Stephen Darwall, *Welfare and Rational Care* (Princeton, NJ: Princeton University Press, 2002). Page references in the text are to this book.

⁴ L. W. Sumner, *Welfare, Happiness, and Ethics* (Oxford: Clarendon Press, 1996).

⁵ The objection here is not to the fact that Darwall's is a "buck-passing" theory of welfare, which equates claims about welfare with claims about reasons. It concerns how welfare, however understood, relates to care. Darwall holds that claims about reasons to desire someone's states derive from claims about care for her; the objection challenges this view.

⁶ Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 168; Thomas Hurka, 'The Justification of National Partiality', in Robert McKim and Jeff McMahan (eds.), *The Morality of Nationalism* (New York: Oxford University Press, 1997), p. 150.

⁷ The latter is Sumner's view; see *Welfare, Happiness, and Ethics*.

⁸ For earlier statements of this analysis see R. M. Hare, 'Wanting: Some Pitfalls', in R. Binkley, R. Bronaugh and A. Marras (eds.), *Agent, Action, and Reason* (Toronto: University of Toronto Press, 1971); Patricia S. Greenspan, 'Conditional Oughts and Hypothetical Imperatives', *Journal of Philosophy* 72 (1975), pp. 259–276; Stephen L. Darwall, *Impartial Reason* (Ithaca, NY: Cornell University Press, 1983), pp. 15–17, 43–50; and John Broome, 'Normative Requirements', *Ratio* 12 (1999), pp. 398–419.

⁹ Immanuel Kant, *The Doctrine of Virtue: Part II of The Metaphysics of Morals*, trans. Mary J. Gregor (Philadelphia: University of Pennsylvania Press, 1964), pp. 44–47.

¹⁰ Darwall, *Impartial Reason*, Part III, esp. pp. 160–63.

¹¹ Bernard Williams, 'Persons, Character, and Morality', in his *Moral Luck* (Cambridge: Cambridge University Press, 1981); Michael Stocker, 'The Schizophrenia of Modern Ethical Theories', *Journal of Philosophy* 73 (1976), pp. 453–466.

¹² I take it that what Williams and Stocker object to is not just evaluative beliefs but evaluative thoughts in general, including seeings-as that stop short of belief.

¹³ This is essentially Ross's objection to what he (perhaps mistakenly) takes Kant's view of rightness to be; see W. D. Ross, *The Right and the Good* (Oxford: Clarendon Press, 1930), pp. 5–6. I raise the same objection against some related views in Thomas Hurka, *Virtue, Vice, and Value* (New York: Oxford University Press, 2001), pp. 184–187.

¹⁴ John Stuart Mill, 'Utilitarianism', in J. M. Robson (ed.), *Essays on Ethics, Religion and Society*, in *Collected Works of John Stuart Mill*, vol. 10 (Toronto: University of Toronto Press, 1969), p. 211; William K. Frankena, *Ethics*, 2nd ed. (Englewood Cliffs, NJ: Prentice-Hall, 1973), pp. 89–92; Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), pp. 501–502.

¹⁵ I have defended a similar view in *Virtue, Vice, and Value*.

¹⁶ Quoted in Mihaly Csikszentmihalyi, *Flow: The Psychology of Optimal Experience* (New York: Harper & Row, 1990), p. 40; comments from a sailor illustrating the same point are on p. 55.

Department of Philosophy
University of Toronto
215 Huron Street, M5S 1A1
Toronto, ON
Canada
E-mail: tom.hurka@utoronto.ca