

September 18, 2003

**What is the Rational Care Theory of Welfare?  
A Comment on Stephen Darwall's *Welfare and Rational Care***

Fred Feldman  
Department of Philosophy  
University of Massachusetts at Amherst  
Amherst MA 01003

email: [ffeldman@philos.umass.edu](mailto:ffeldman@philos.umass.edu)

## What is the Rational Care Theory of Welfare?\*

When we speak of a “good life” there are several different things we might mean. We might mean a *morally good life*. We might mean a life *good for others*, or *good for the world in general*. We might mean a life *good in itself for the one who lives it*. This last may also be described as the life high in individual welfare.

In their attempts to help readers focus on the concept of *individual welfare*, a number of philosophers have appealed to what we may call the “Crib Test”.<sup>1</sup> Imagine that you are the proud and loving parent of a newborn baby. Imagine that as you look into the crib where the baby is peacefully sleeping, your heart is filled with parental affection and concern. You hope that things will turn out well for your baby. You might express your hope by saying, ‘I hope that this baby gets to live a *good life*’. The idea behind the Crib Test is that when you speak of a “good life” in this context, you are speaking of the baby’s individual welfare. You are hoping (according to the Crib Test) that he gets a life that’s good for him – high in welfare. So the Crib Test is designed to help us identify the concept of individual welfare and distinguish it from other things that might be meant by talk of ‘the Good Life’.

The Crib Test is familiar and helpful, so far as it goes. It points us in the direction of the concept of individual welfare. Most of us wouldn’t put too much stock in it, since a misguided or confused parent might hope for nearly *anything* for his child, even with a heart filled with love. But Darwall adopts the vague idea behind the Crib Test and elevates it to a position of central importance in his book. In his hands, this rough idea becomes a fundamental doctrine in metaethics. He calls this the ‘Rational Care Theory of Welfare’ (RCTW). Approximately a third of Darwall’s book is devoted to the presentation and defense of this theory.<sup>2</sup>

Here I aim to make just one main point. It is that in spite of its central place in this book, RCTW remains seriously obscure. Darwall states the theory in several different ways. He suggests that they are intended to be interchangeable, but the statements are obscure. Some of them seem to be open to several different interpretations. It’s not clear that the different formulations really come to the same thing. In the end, there is no form of RCTW that is both plausible and plausibly attributable to Darwall.

---

\*A comment on Stephen Darwall’s *Welfare and Rational Care* (Princeton and Oxford: Princeton University Press, 2002).

<sup>1</sup> Robert Adams made use of this test in *Finite and Infinite Goods: A Framework for Ethics* (Oxford: Oxford University Press, 1999). I first used it in ‘On the Advantages of Cooperativeness’ *Midwest Studies in Philosophy*, XIII (1988), p. 308. I used it again in ‘The Good Life: A Defense of Attitudinal Hedonism,’ *Philosophy and Phenomenological Research* 65 (2002). Others have used it as well.

<sup>2</sup> Darwall gives a provisional exposition and defense of RCTW in Chapter One, which is new. Chapter two as a whole is devoted to the theory. Chapter two is a trivially revised version of Darwall’s “Self Interest and Self Concern” which was published in 1997.

## I

Darwall explicitly says that RCTW is a doctrine in metaethics.<sup>3</sup> It is intended to give an account of the concept of individual welfare; it is supposed to display the meaning of the expression ‘p would be good for S’; it is intended to be a definition or analysis of the concept of individual welfare.<sup>4</sup>

Here are some passages<sup>5</sup> in which Darwall seems to be stating RCTW:

A1. “... a person’s good is constituted ... by what one (perhaps she) should want *insofar as one cares about her*.” (p. 4)

A2. “*What is for someone’s good or welfare is what one ought to desire and promote insofar as one cares for him*.” (p. 7)

A3. “... what it is for something to be good for someone *just is* for it to be something one should desire for him for his sake, that is, insofar as one cares for him.” (p. 8)

A4. “...the claim that what we ought to desire for someone’s sake is what is good for him *is* a tautology.” (p. 10)

B5. “... welfare must be an explicitly normative concept. My proposal will be that it is the concept of what we would rationally desire for someone insofar as we care for her, or equivalently, what it is rational to desire for her for her sake.” (pp. 11-12)

B6. “In order for a thing to be in a person’s interest it must be something someone who cared about the person would *rationally* want for him, insofar as she cares.” (p. 48)

C7. “Something is for someone’s good if it is what that person would want for herself, as she actually is, insofar as she is fully knowledgeable and experienced *and* unreservedly concerned for herself.” (p. 31)

D8. “A person’s welfare is, I claim, the object of a desire spawned by concern for that person.” (p. 24)

## II

Some of these statements seem to tie welfare to what is or would be “rationally desired”. Others tie it to what *should be* or *ought to be* desired. Darwall says (e.g., on pp. 8-9) that these formulations are equivalent. However, we must be careful here since there are ways of understanding the concepts of (a) what it is *rational* for a person to desire and (b) what a person *ought* to desire that would make these formulations come apart.

---

<sup>3</sup> ‘It is a position in the metaethics of welfare concerning the concept of a person’s good...’ p. 12.

<sup>4</sup> Darwall’s substantive normative thesis about what in fact enhances individual welfare is the so-called “Aristotelian Thesis”. I discuss the Aristotelian Thesis in Appendix D of my forthcoming *Pleasure and the Good Life*.

<sup>5</sup> I recognize, of course, that these passages have been torn from their original contexts in the book. I also recognize that it’s possible that some of these might not have been intended as actual statements of RCTW. Perhaps they are just statements of other, closely related doctrines. All emphasis is taken from the original texts. This list of quotations is not intended to be complete. Many other similar passages could be cited.

A few formulations of the theory seem to tie a person's welfare to what in fact *is* desired for that person. This introduces a different idea and gives rise to different versions of RCTW. Some of them seem so implausible that they probably should be dismissed. Consider, for example, D8. D8 makes no mention of rationality or obligation. Instead, it seems to identify a person's welfare with the object of a desire (rational or not) that is in fact "spawned" by concern for the person. I take "spawning" to be a causal notion. Thus, D8 suggests:

D. p is good for S =df. our care for S causes us to want p for S.

Surely this does not express any idea Darwall really meant to endorse. If a person were sufficiently confused or misguided, caring for S could give rise to a desire for almost anything for S. Furthermore, suppose none of us cares about S. Then there is nothing such that our care for S causes us to want it for S. Then D seems to imply that nothing is good for S. I am confident that D8 and similar passages elsewhere in the book are mere slips of the pen.<sup>6</sup>

Passages like B5 and B6 raise other problems. According to them, something is good for a person if one who cared about him *would rationally desire it* insofar as he cared. When we say that someone *would rationally desire* something, we seem to be saying that he would desire it and his desire would be rational. But the phrase 'would rationally desire' cannot be used in this way in B5 and B6. For if they were so understood, those remarks would imply that if p would be good for you, then if I were to care about you I would desire p for you and this desire would be rational. But it seems obvious that this is not something Darwall means to say. He recognizes that in some cases, though p would be good for you, caring for you would not give rise to any desire (rational or otherwise) on my part for p for you. There is a difference between saying (a) that if I were to care about you, then *it would be rational for me to want p for you*, and saying on the other hand (b) that if I were to care about you, then *I would rationally want p for you*. Darwall says (b)-like things in B5 and B6 and elsewhere in the book, but I am confident that they should not be taken too literally. I suspect that he meant to affirm (a)-like things.

The passage identified as C7 seems a bit of an aberration, too. This is the only passage I could find in which Darwall explicitly brings in considerations of "full information". The remark occurs in a context in which he is discussing the similarities between his own view and some other more familiar views. Perhaps he just meant to suggest that his own view is in many ways similar to this specific form of the full information theory.<sup>7</sup> It would be unfair to take C7 as a serious expression of a view that Darwall wants to defend.

### III

That leaves the "A" quotations. If we focus on these passages, we will naturally conclude that RCTW can be formulated as a definition, or conceptual analysis, as follows:

A: p is good for S =df. insofar as we care about S, we ought to want p for S.

---

<sup>6</sup> Footnote 18 on p. 111 makes this clear.

<sup>7</sup> Darwall says, 'If...any informed desire standard can serve as a plausible criterion of welfare, I think it will be something like the following.' He then states C6. (p. 31)

This raises several questions: What does ‘ought’ mean here? What does ‘care’ mean here? What does ‘insofar as’ mean here? Who are “we”? Let’s start with what may seem the easiest question: what is meant by ‘insofar as’?

I looked up the phrase ‘insofar as’ in several dictionaries. They all agreed that in its current standard use, ‘insofar as’ means the same as ‘to the extent that’.<sup>8</sup> To say that A is F “insofar as” B is G is to say that there is a sort of proportionality between A’s being F and B’s being G. Consider the extent, or “distance”, to which B is G; the statement then says that A is F to that same extent, or “distance”. Thus, if we say that a person will improve as an athlete *insofar as* he trains harder, what we mean is that the degree to which he will improve is proportional to the degree, or amount, or extent, that he trains. The more he trains, the better he will get. But it should be obvious that Darwall could not have been using ‘insofar as’ in this standard sense in any of his formulations of RCTW. To see this, just consider the strange thing that A would say if it were so understood:

Ax1: p is good for S =df. to the extent that we care about S, we ought to want p for S.

Ax1 says that p is good for S if and only if there is a kind of proportionality between the extent of our care for S and the extent of our obligation to want p for S. If this is indeed the intended interpretation of RCTW, the doctrine can be expressed as follows:

Ax2: p is good for S =df. for any extent, n, if we were to care about S to n, then we would have, to extent n, an obligation to want p for S.

Ax2 depends essentially on the idea that there are degrees, or extents, to which we can have obligations. I am not sure that I understand that idea. Clearly, it makes no sense if the obligation in question is any sort of “absolute” obligation. On the other hand, if we assume that the sort of obligation invoked here is one that comes in degrees, or amounts, then it probably would be reasonable to assume that a variety of factors could influence the degree to which someone has an obligation. Suppose we had several different reasons to want p for S. Suppose some of these were based on considerations other than our care for S. Perhaps, for example, we had promised to want p for S. Then the extent of our obligation to want p for S might be greater than the extent of our care for S. In the opposite direction, if we had independent reasons to avoid wanting p for S, then the extent of our obligation to want p for S might be smaller than the extent of our care for S. In general, it seems to me that the fundamental idea behind Ax2 is simply wrong. Surely there can be things that are good for S, but where the extent of our care for S does not march in lock-step with the extent of our obligation to want those things for S. Furthermore, the doctrine yields unacceptable results in the case in which we do not care about a worthy person. Suppose we have every reason to care about S and to want p for S (since it would be good for him). But suppose we just don’t care about S. Then the extent of our care for S would be much smaller (zero) than the extent of our obligation to want p for S. Ax2 is again falsified.

It might appear that Ax2 invokes a proportionality between the wrong items. Maybe instead Darwall meant to say that when p is good for S, there is a proportionality between the extent of our care for S and the extent of the *want* that we ought to have for S’s getting p. In other words:

---

<sup>8</sup>*Oxford English Dictionary (online edition); The American Heritage Dictionary of the English Language, 2000 (4<sup>th</sup> ed.); The Merriam-Webster Online Dictionary (10<sup>th</sup> ed.); The Oxford Paperback Dictionary (New Edition, 1983); The American College Dictionary (Random House, 1964); Webster’s New Collegiate Dictionary (1953); etc .*

Ax2': p is good for S =df. for any extent, n, if we were to care about S to n, then we would have an obligation to want, to extent n, p for S.

Ax2' has some bizarre implications. For one, it implies that if in fact we don't care about you, then, even if p would be good for you, we have an obligation to want *to degree zero* for you to get p. This seems to me to say that it would be our duty to avoid wanting p for you. Why would we have such a strange obligation? Since Darwall does not clearly commit himself to Ax2', I suspect that this is not a fair interpretation of his meaning.<sup>9</sup> In general, it's just hard to understand how proportionality could figure in RCTW.

Some dictionaries suggest that 'insofar as' has a sort of degenerate use in which it means pretty much the same as 'given that', or 'since'.<sup>10</sup> Under this interpretation, the A formulations might be taken to mean:

Ag: p is good for S =df. given that we care about S, we ought to want p for S.

Ag is an even stranger thesis. Clearly, there are many people about whom we don't care. Suppose S is such a person. Surely there are things that are good for S. Suppose p is such a thing. It would be preposterous to say in this case that *since we do care about S* we ought to want p for S. This flies in the face of the stipulated fact that in this case we do *not* care about S. In other words, this formulation of RCTW has a weird implication: if anything is good for someone, then we care about that person.

The natural assumption here (one that I made immediately upon reading the passages quoted above) is that Darwall uses 'insofar as' to mean 'if'. In one possible "iffy" interpretation<sup>11</sup>, the definition means this:

Ai: p is good for S =df. if we were to care about S, then we would have an obligation to want p for S.

This, clearly enough, would make questions about the goodness of p for S independent of questions about whether we actually care about him. The fact that we don't actually care about him (if that's the case) would be irrelevant. We could say: 'Of course we don't care about S and we don't have any obligation to want anything for him. But if we were to care about him, then we would have obligations to want things for him. These hypothetical or conditional obligations to want things for S determine what's good for S. What we would have an obligation to want under those counterfactual circumstances are the things that are in fact good for him.' This formulation is reminiscent of various formulations of the so-called "full information" forms of preferentism.

---

<sup>9</sup> I am grateful to Kristen Hine for encouraging me to consider this possible interpretation of RCTW.

<sup>10</sup> A discussion of this possible sense of 'insofar as' can be found under 'inasmuch as' in the *Oxford Universal Dictionary on Historical Principles*, p. 975.

<sup>11</sup> We could take the 'if' to indicate material conditionality, or strict conditionality, or perhaps even the kind of conditionality involved in statements of conditional obligation. Darwall never suggests that he means any of these. I focus on what may seem the most likely interpretation – the interpretation under which 'if... then...' is taken as a subjunctive conditional.

This approach might seem interesting and might be worth pursuing. However, Darwall seems to reject it. He says:

Insofar as we care for someone, we ought to be guided by his good. So far, these reasons are merely hypothetical. The idea, however, is not that the fact that one cares about someone makes considerations of his good reasons for one. The reasons are not conditional on one's caring. If that were so, they would be canceled once one ceased to care. They are conditional, rather, on a hypothesis one accepts or is committed to *in* caring, namely, that the cared for is *worth* caring for. (p. 8; see also pp. 48-9 where Darwall seems to repeat the claim)

Notice Darwall's remark that 'the reasons are not conditional on one's caring'. If we were to take this remark seriously we would have to reject Ai as an interpretation of Darwall's view, since Ai explicitly does make the obligation conditional upon the care.<sup>12</sup>

Note that there is a hint of a reformulation of RCTW in the final quoted sentence. It suggests that the claim is really this: something, p, is good for a person, S, provided that one's obligation to want p for S is conditional upon S's being worthy of care. In other words:

Awc1: p is good for S =df. if S is worthy of care, then we ought to want p for S.

It is not likely that this is a suitable expression of Darwall's thesis. One problem is that he seems to affirm it in only two passages – the ones just cited. Another problem is that if every person is worthy of care (as Darwall says on pp. 78, 83 and elsewhere), the antecedent would seem to be doing no work. Why not just say that a person's good is whatever we ought to want for him? We could rephrase it as:

Awc2: p is good for S =df. we ought to want p for S.

In this case, obviously, it would be strange to use the term 'care' in the name of the theory. Furthermore, it would apparently imply that we have obligations to want specific things for specific persons even though we never met those persons and even though we could not conceive of the things.

It might seem that all of this discussion of 'insofar as' is relatively trivial. Perhaps it appears that we understand Darwall's view well enough. I think that would be a serious mistake. To see the importance of this little phrase, let's use 'C' to abbreviate 'we care about S', and let's use 'OW' to abbreviate 'we ought to want p for S'. What, then, is the logical structure of the right-hand side of the A formulations? is it the claim that C *is proportional to* OW? or is it the claim that C is proportional to the amount of W such that OW? or is it that if C *were true, then* OW would be true? or is it that *given that* C is true, OW is true? or is it something else entirely? Insofar as we don't understand the logical structure of the definiens, we don't understand the proposed theory.

---

<sup>12</sup> Nishi Shah has suggested (in personal correspondence) that it would be possible to revise Ai so as to avoid this difficulty. We could say: 'p is good for S =df. if we were to care about S, then it would be the case that either we ought to want p for S or else we ought to stop caring for S.' Space constraints preclude discussion of this proposal. I am very grateful to Shah for extensive, detailed commentary on an earlier draft.

## IV

Now I want to turn to the concept of obligation (or the “ought” or “should”) that appears in many of Darwall’s statements of RCTW. Darwall says very little about it. In one place he is talking about the ‘should’ in formulation A3. He says, “The relevant sense of ‘should’ again, is its most general normative sense.”<sup>13</sup>

One interesting question about any ‘ought’ is this: does that sort of ‘ought’ imply ‘can’? From the fact that someone ought (in that sense) to do something, does it follow that that person can, or is able to, do that thing? It appears to me that Darwall’s ‘ought’ does not satisfy this “ought implies can” principle. Suppose I have been brought up in a strict religious tradition. According to this tradition, let us imagine, marrying outside the faith is “unthinkable”. No matter how much I care about someone, I cannot bring myself to want him to marry outside his faith. Indeed, the more I care about him, the more difficult it would be for me to want such a thing for him. If the ‘ought’ in A implies a corresponding ‘can’, then, since I cannot want him to marry outside his faith, it is not true that I ought to want this for him. As a result, if Darwall’s ‘ought’ implies ‘can’, it would follow that marrying outside the faith is not good for him. That seems silly. Maybe marrying outside the faith would be ideal for him. Furthermore, if many others who care about him do want him to marry outside the faith, Darwall’s theory might even seem to have inconsistent implications in this case.

Another interesting question about any alleged ‘ought’ is this: does the ‘ought’ express a form of “absolute obligation” or does it express a form of mere “prima facie obligation”? My impression is that Darwall intends his ‘ought’ to express a weak form of prima facie obligation. This impression is based on hints and suggestions he makes in various places. For example, in some places he suggests that when he says that we ought to want p for S, what he means is merely that it “makes sense”<sup>14</sup> for us to want p for S, or that it would be rational for us to want p for S. Let us consider the idea that a version of RCTW can be expressed in this way:

Apf: p is good for S =df. if we were to care about S, then we would have a prima facie obligation to want p for S.

I think Apf is much too weak to capture any interesting notion of welfare. Consider this case: Suppose our friend S is sick. Unfortunately, we know very little about medicine. Suppose a learned medical scholar comes to us, and tells us that S will be instantly cured if she gets a dose of medicine M. Then it seems to me that we have a prima facie obligation to want S to get a dose of medicine M. That is, we have *some reason* to want S to get this medicine; it *makes sense* under the circumstances for us to want S to get the medicine. But of course it’s consistent with all this to suppose that medicine M is the worst possible thing for S. It might be fatal, given her condition. But Apf implies that it would be good for her.

Although I do not have the space to try to establish this point here, I think it could be shown that for anything – good, bad, or indifferent – there are imaginable circumstances in which it would “make sense” for those who care about S to want that thing for S. One factor is epistemic: Given sufficiently large and misleading stores of evidence, care for S could even make it reasonable for us to

---

<sup>13</sup> P. 8. It’s odd that he says ‘again’ here, since this is the first (and only?) time that he mentions that the ‘should’ has this “most general normative sense”.

<sup>14</sup> p. 9.

want her to get a dose of arsenic. Another factor is causal: we might realize that a small dose of negative welfare would shock S out of her complacency; it might encourage her to freely adopt some better patterns of behavior. If this is right, then Apf would imply that just about anything – good, bad, or indifferent – would enhance S’s welfare.

The variant of A that makes use of absolute obligation does not have this implication, but seems just as implausible. Consider this:

Aabs: p is good for S =df. if we were to care about S, then we would have an absolute obligation to want p for S.

Suppose you care deeply about the welfare of your child. Suppose the evil demon can read your mind. The demon tells you this: “If I discover that you are wanting that child to fare well, I will subject the child to terrible torture.” It seems to me that under these unusual circumstances your overriding obligation would be to avoid wanting your child to fare well. If you were to want that, the child would suffer. Clearly, then, you do not have any absolute obligation in this case to want your child to fare well. (I grant that you would still have a *prima facie* obligation to want him to fare well, but it seems to me that that obligation would be overridden by the countervailing obligation to avoid wanting him to fare well.)

The evil demon case is obviously far-fetched. However, the point is relevant to real-life cases. Our wantings are events in the world. They have consequences. Sometimes these consequences bear on welfare. Thus, for example, I might want my son to do well in Little League. Wanting this might make me behave badly at games. Under the circumstances, it might be better all around if I didn’t want it. And in such a case my concern for my son’s welfare might be precisely the thing that makes it important for me to “keep my emotional distance”. Here I have a good reason, based on considerations of my son’s welfare, to refrain from wanting something that would be good for him. Surely there could be a case in which reasons such as these are decisive – a case in which all things considered I ought to avoid wanting what’s best for him.

In several places Darwall suggests that when he says that we ought to want things, what he means is that it would be rational for us to want those things<sup>15</sup>. This suggests a seemingly different interpretation of RCTW. On this view, the theory amounts to something like this:

Ar: p is good for S =df. if we were to care about S, then it would be rational for us to want p for S.

The concept of rationality is open to a variety of interpretations.<sup>16</sup> One is fundamentally prudential: to say that it is rational for us to do something is to say that doing it would maximize our welfare, weighted for probability. Clearly, Darwall cannot have any such conception of prudential rationality in mind. The circularity is too blatant. And in any case the resulting doctrine is pretty obviously false. Suppose it’s highly probable that we will be best off if we somehow manage to get ourselves to want p for S. Then, in this sense, it would be rational for us to want p for S. Clearly, however, it’s *our* welfare that is being enhanced here, not S’s.

---

<sup>15</sup> Some examples were cited above in B5 and B6. There are many others in the book.

<sup>16</sup> Nishi Shah reminded me that Darwall says that rational care theories of welfare ‘can be neutral on which theory of rationality and rational desire is correct.’ (p. 48)

Talk of rationality suggests something epistemic.<sup>17</sup> To say that it is rational for us to want something for a certain person might be to say that, given our evidence about the person and the thing, it would “make sense” for us to want that thing for the person. But no such sense could be relevant to Darwall’s enterprise, since (obviously) a person with bad evidence could have reason (in this sense) to want someone he cared about to take a dose of arsenic even though arsenic would not in fact be good for her.

It might seem that the fundamental idea behind Darwall’s thesis is just this: there are some things that are just “appropriate” or “fitting” objects for us to want for those we care about. By their nature, these objects “call for” or “require” us to want them for our loved ones.<sup>18</sup> While this idea does seem attractive, it is not clear that it was the idea that Darwall had in mind. Furthermore, it seems to make the concept of care irrelevant. Suppose p would be good for S. Suppose this shows that by its very nature, p is an “appropriate” or “fitting” thing for us to want for S insofar as we care about S. Then wouldn’t p be an “appropriate” or “fitting” thing for us to want for S *even if we didn’t care for S*?

## V

The final element in RCTW is the concept of *care*. Darwall devotes Chapter III to the concept of care. One of his main goals is apparently to distinguish the concept of care from a number of closely related concepts, such as empathy and mere “emotional contagion”. At the outset of the chapter, Darwall seems to be defining ‘care’<sup>19</sup> when he says:

“... *sympathetic concern* or *sympathy*. It is a feeling or emotion that (i) responds to some apparent obstacle to an individual’s welfare, (ii) has that individual himself as an object, and (iii) involves concern for him, and thus for his welfare, for his sake.” (p. 51)

Darwall recognizes that he cannot make use of anything like this as a definition or analysis of care. To do so would be viciously circular, since any such definition would make use of the concept of welfare, and he wants to define welfare in terms of care. He cannot at the same time define care in terms of welfare.<sup>20</sup>

Darwall says that he can avoid this circularity if he can show that care is a natural kind. More exactly, he says this:

“However, we need not define care...if it is something like a psychological natural kind. Just as we can use a term like ‘water’ without a prior definition to refer to the natural stuff in the rivers and lakes for purposes of empirical theory, so likewise might we refer to care for purposes of a metaethical theory of welfare if it is a natural kind.” (p. 50)

---

<sup>17</sup> This sort of interpretation is also suggested by Darwall’s use of terms such as ‘warranted’ and ‘justified’ in several of these contexts (e.g., p. 9)

<sup>18</sup> Chris Heathwood pointed out that Chisholm often spoke in this way of the “fittingness” of certain emotional reactions to certain objects.

<sup>19</sup> For reasons he never explains, Darwall uses ‘sympathetic concern’ in this chapter instead of ‘care’, which he uses elsewhere.

<sup>20</sup> He explicitly says this on p. 12.

He then devotes about twenty pages to the attempt to identify and describe the phenomenon of care. He distinguishes it from some near relatives; he claims that it functions in some psychological explanations; he says that it is “natural”. And then, in the final paragraph of the chapter, he says:

“We may therefore take it that the phenomenon of care for someone for her own sake is available for metaethical theorizing, specifically, for a rational care theory of welfare.” (p. 72)

Darwall’s response to the charge of circularity amounts to little more than this mere hint. Of course there is something “natural” about care. That seems to mean that in the natural course of events people find themselves caring about others. Care is not a man-made phenomenon. It is not a mere figment of our imaginations. Perhaps Darwall is right when he claims that we can *refer* to the phenomenon of care even if we have no definition of the concept – even if we aren’t able to spell out the meaning, or intension, of ‘A cares for B’. But I am puzzled about why he thinks (if indeed he does) that these facts somehow bear on the question whether there is a problem about circularity in defining ‘welfare’ in terms of ‘care’, and ‘care’ in terms of ‘welfare’. He does not explain what he means by ‘natural kind’ in this context. He does not explain what he thinks it would take to show that something is a natural kind in the relevant sense. He does not give any account of a theory of meaning that would justify his claim that his use of ‘care’ in his account of welfare evades the charge of circularity. If the reader was dubious about the procedure at the outset, he or she will be at least as dubious after studying Chapter III.<sup>21</sup>

## VI

My conclusion is this: Darwall has not made clear what he means by ‘insofar as’, ‘care’, ‘rational’ and ‘ought’. Thus we don’t know what he means when he says that what’s good for a person is what we *ought* to want (what would be *rational* for us to want) for him *insofar as* we *care* about him. The terms are open to many possible speculative interpretations. Darwall’s thesis is implausible under all the possible interpretations discussed here. So until we have a clearer understanding of what this thesis is supposed to mean, substantive criticism seems premature.<sup>22</sup>

---

<sup>21</sup> Although virtually all of Chapter III is devoted to discussions of the concept of care, only two short passages are explicitly devoted to the argument about the naturalness of care and its relevance to the circularity objection. I think I have quoted just about everything that Darwall says on this topic in Chapter III. Earlier in the book (pp. 12-13) he hints that he will be making use of some such line of argument. However, in that earlier context he seems to emphasize the claim that ‘it is sufficiently evident to us that there is such a psychological state [as care]...’ This seems obvious but it’s hard to see why it is relevant to the circularity puzzle. Of course there is such a thing as care. But it’s not clear how the mere *existence* of such a state as care would bear on the question of circularity in definition. In one passage (p. 48) Darwall is discussing this circularity problem. He says ‘And even if we don’t know very much about what care is, we may know it when we see it.’ This also is plausible, but I find it difficult to understand how this claim relates to the claim about the naturalness of care. I also find it difficult to understand precisely how it might be thought to bear on the question about circularity.

<sup>22</sup> I am grateful to several friends who gave thoughtful criticism and suggestions. Chris Heathwood, Kristen Hine, Kris McDaniel and Nishi Shah have been especially helpful. This is not to say that any of them agrees with everything I say here.