https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

CONTRACTUALISM, PERSONAL VALUES, AND WELL-BEING*

By Peter de Marneffe

I. Contractualism and Philosophical Utilitarianism

In "Contractualism and Utilitarianism," T. M. Scanlon draws a distinction between "philosophical utilitarianism" and his own contractualist view.[1] Philosophical utilitarianism, according to Scanlon, is the view that all moral facts are, fundamentally, facts about what will promote well-being. Scanlon holds instead that moral facts are facts about what actions would be permitted or prohibited by principles no one could reasonably reject. A person, S, can reasonably reject a principle, according to Scanlon, if and only if there is a good personal reason for S to want not to observe this principle or to want it not to be generally observed, and this personal reason has greater weight than any personal reason someone has to want S to observe this principle or to want it to be generally observed.[2] A personal reason for or against a principle is a reason for a person to prefer his own situation when this principle is or is not observed.[3] It follows that if all personal reasons were facts about well-being, there would be no substantive difference between philosophical utilitarianism and contractualism. All moral facts would fundamentally be facts about well-being, even on a contractualist view.[4]

* I thank the other contributors to this volume for valuable questions and comments, particularly David Brink, David Estlund, Sam Fleischacker, and Erin Kelly. Thanks are also due to Doug Portmore, Hannah Tierney, Justin Tosi, and Steve Wall for helpful feedback.

[1] T. M. Scanlon, *The Difficulty of Tolerance* (New York: Cambridge University Press, 2003), 129, 133.

[2] T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 195, 218–20.

[3] Ibid., 219. Scanlon writes here that personal reasons "have to do with the claims and status of individuals in certain positions" and seems to identify them with "the claims and interests of individuals." But he does not make explicit here what makes a claim "a claim of an individual in a certain position" or how this sort of claim differs from a claim that an individual might make on behalf of something else of value, such as the Grand Canyon or an old-growth forest. To clarify the nature of Scanlon's distinction between personal and impersonal reasons, I believe we must therefore explicitly introduce the distinction between reasons that individuals have to prefer their own situations when a state of affairs obtains and the reasons that individuals have to prefer the world when a state of affairs obtains. On this view, an interest or claim just is a good reason for someone to prefer his own situation when a state of affairs occurs.

[4] This is not to say that there would be no substantive difference between contractualism and *utilitarianism.* Utilitarianism is a particular view about how individual well-being should be taken into account in identifying the morally right action, a view that is arguably inconsistent with contractualism. Philosophical utilitarianism, in contrast, as Scanlon uses the

doi:10.1017/S0265052513000034
© 2013 Social Philosophy & Policy Foundation. Printed in the USA.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

Scanlon, however, rejects what he calls the "Complaint Model," which is the view that all personal reasons for rejecting principles are facts about well-being.[5] Scanlon holds instead that other kinds of facts provide good personal reasons for rejecting principles.[6] Specifically, he believes that facts about fairness provide good personal reasons. If one principle ensures the fair distribution of some opportunity or resource and another principle arbitrarily favors someone other than S, then, Scanlon holds, there is a good personal reason for S to reject the second principle, even if S's well-being would be no lower as a result of its general observance.

What is the difference, though, between well-being and other personal values? Well-being is commonly supposed to consist in what is intrinsically good for a person. Sometimes by "intrinsically good" we mean "good in itself": good in virtue of its intrinsic properties, the properties it has independent of its relation to anything else. More often we mean what is "good for its own sake," "good as an end," or noninstrumentally good.[7] A person's well-being is commonly supposed to consist in what is intrinsically good for him in this second, noninstrumental sense. This suggests that whereas well-being consists of what is intrinsically good for a person, other things are good only impersonally or instrumentally. If so, fairness is good only instrumentally or impersonally.

Scanlon, however, rejects this view. He does not regard fairness as having only impersonal value because he believes that facts about fairness provide personal reasons to reject principles; and his reasons for rejecting the Complaint Model imply that fairness is intrinsically valuable, something worth wanting for its own sake, apart from any beneficial consequences it might have in promoting well-being. Scanlon's position thus appears to be that our situations are intrinsically or noninstrumentally better in some way under a fair distribution than they are under an unfair distribution in which our well-being is no lower. From this it follows that fairness is, in some sense, "intrinsically good for us," although, according to Scanlon, it is not an aspect of well-being. This position makes sense only if well-being does not include everything that is intrinsically good for us. So one wonders what the difference is between well-being and other things that are intrinsically good for us.

A rational desire analysis of well-being cannot answer this question. According to the unrestricted version, facts about a person's well-being are facts about what this person would rationally desire for its own sake. This does not explain the difference between well-being and fairness

term, is the name of a general kind of moral theory, of which (aggregating, maximizing) utilitarianism is just one example. The point here is that contractualism would be another example of this general view if all personal reasons were reasons of well-being.

[5] Ibid., 229.

[6] Ibid., 229.

[7] For a similar distinction between two ideas of intrinsic goodness, see Christine Korsgaard, "Two Distinctions in Goodness," in *Creating the Kingdom of Ends* (New York: Cambridge University Press, 1996), 249–74.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

because we can rationally desire fairness for its own sake. This analysis is unsatisfactory in any case because we rationally desire many things that do not contribute to our well-being. Suppose I meet a stranger on a train and after chatting with him wish him well.[8] Although I rationally desire his future welfare for its own sake, his future welfare does not directly contribute to my well-being—not, at least, if I never think of him again or interact with him in any way.

A restricted rational desire analysis provides a solution to this problem. On this view, facts about a person's well-being are facts about what it would be rational to desire noninstrumentally *for* this person. Although it might be rational for me to desire noninstrumentally the future welfare of the stranger on the train, it does not follow that it is rational for me to desire this noninstrumentally *for myself*. I might desire his future happiness for him, but not for me, which would explain why his future happiness contributes to his welfare, and not to mine.[9] But this analysis also fails to explain Scanlon's distinction between well-being and other personal values. If fairness is of noninstrumental personal value, then it seems to follow that it is rational to desire noninstrumentally for a person that he have no less than a fair share. But then, according to this analysis, having no less than a fair share contributes directly or noninstrumentally

[8] This example is from Derek Parfit, *Reasons and Persons* (New York: Oxford University Press, 1984), 494. Parfit's stranger is not identified as being on a train. James Griffin puts him there in *Well-Being* (New York: Oxford University Press, 1986), 171, where he has been ever since.

[9] I believe that Sidgwick's formal analysis of well-being is a version of the restricted rational desire analysis. Sidgwick takes "ultimate good on the whole for me" to mean "what I should practically desire if my desires were in harmony with reason, assuming my own existence alone to be considered" (*The Methods of Ethics* [Macmillan, 1907; Indianapolis: Hackett Publishing Co., 1981], 112). This analysis addresses the stranger-on-the-train objection because it narrows the scope of the relevant objects of desire to what it would make sense to desire "assuming my own existence alone to be considered." But what is it to assume this? If it is to assume that only my own well-being or good is to be considered, then the analysis is circular. If it is to assume that only I exist, this would entail that a number of claims by ethical theorists about well-being are conceptually incoherent. It would entail, for example, that being a good parent to our children is not intrinsically good for us, since it is not rational to want to be a good parent to someone on the assumption that she does not exist. Some would deny that being a good parent is intrinsically good for us, thinking that it is good for us only as a means to the pleasure it gives us. But we should reject any analysis that renders conceptually incoherent the claim that being a good parent contributes (directly) to our well-being, because any such analysis would fail to illuminate the concept that ethical theorists are using when they claim (perhaps mistakenly) that being a good parent does contribute (directly) to our well-being. (See, for example, *Reasons and Persons*, 495, where Parfit identifies being a successful parent as something that would contribute directly to our well-being if the "Success Theory" were true.) There is, however, another interpretation of Sidgwick that avoids these difficulties. On this interpretation "ultimate good for me" means "what I should practically desire *for myself* if my desires were in harmony with reason." In other words, a person's well-being consists only in what he ought to desire, or would rationally desire, for himself. If this is the correct interpretation of Sidgwick's view, then his formal analysis of well-being is a version of the restricted rational desire analysis. His *substantive* theory of well-being, in contrast, is a form of hedonism, because he believes that the only things a person actually ought to desire for himself are "desirable states of consciousness" (*The Methods of Ethics*, 397).

to a person's well-being. So this analysis also cannot make sense of Scanlon's distinction between well-being and other personal values.

A possible response is to challenge the assumption that fairness is rational to desire *for* anyone. Fairness is an impersonal value, one might think, not a personal one. The fact that one distribution of opportunities and resources is fair and another one is unfair is a good reason to prefer a world that contains the first distribution, but it is not a reason for anyone to prefer his own situation in this world. Scanlon himself seems to have regarded fairness as an impersonal value before developing his version of contractualism.[10] Against this, however, it seems that if a distribution arbitrarily favors me over you, this is something that you can object to on your own behalf, and not only on behalf of the world. If a distribution arbitrarily favors me over you, it seems there is a way in which *your situation* is worse than it would be in a world where no one is arbitrarily disfavored. If so, fairness has personal value. If it has personal value, then it is something that is rational to want *for* someone, that is to say, on his or her behalf. It is tempting to say here that although fairness is something that is rational to want for someone in *some* sense, this is not the sense of wanting *for* that is relevant to well-being. But then this difference in kinds of wanting *for* is something we would like to understand, and the restricted rational desire analysis, as stated above, cannot explain it.

Another possible solution is to deny that fairness has intrinsic value. Fairness is valuable, one might think, only because unfairness reduces well-being. When we think we are being treated unfairly, we feel put out, and, even if we do not know we are being treated unfairly, having less than a fair share of opportunities and resources might prevent us from being as happy as we would otherwise be. Unfairness seems objectionable, though, even when it would increase everyone's well-being. Scanlon imagines a situation in which Dudley can make everyone happier by surreptitiously taking more than his fair share of rationed water from the village well and using it to water the flowers in the public garden, which everyone will enjoy.[11] If there is something bad about Dudley taking more than his fair share even when he uses it to increase everyone's well-being (above what it would be when no one takes more than his fair share), the value of fairness cannot be entirely reduced to well-being.

Another possible solution is to reject the starting assumption that a person's well-being consists in everything that is intrinsically good for him. Only good or positive experiences, such as pleasure, delight, enjoyment, contentment, and peace of mind, increase one's well-being. Only bad or negative experiences, such as pain, displeasure, suffering, and anxiety decrease it. This provides a clear explanation of why fairness is

[10] Scanlon, *The Difficulty of Tolerance,* 31–32.
[11] Ibid., 33.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

not an element of well-being, because fairness is not an experience. It fails, however, to explain what ethical theorists are disagreeing about when they disagree about well-being. In the standard imaginary case, Jones and Smith are equally happy because they both believe their families love them and that their colleagues admire and respect them. But whereas Jones's beliefs are true, Smith's beliefs are false. Smith's family secretly hates him and his colleagues secretly have nothing but contempt for him. Are Jones and Smith equal in terms of well-being if Smith will never know the truth? Some say yes, others say no. Perhaps those who say "yes" are right, but an analysis of the concept of well-being should not entail this. A good analysis should be able to identify what these ethical theorists are intelligibly disagreeing about, and the experiential analysis of well-being renders conceptually incoherent the position that Jones's well-being is greater.

Shelly Kagan's intrinsic property proposal is open to a similar objection. The well-being of the stranger on the train makes no direct contribution to my well-being, Kagan suggests, because "changes in well-being must involve changes in the person."[12] A change is "in the person," in the relevant sense, only if it involves changes in intrinsic or nonrelational properties. So, Kagan concludes, "if something constitutes an (ultimate) benefit to a person, it must involve a person's intrinsic properties."[13] This would provide a good explanation of why fairness is not an element of well-being, which is that fairness is a relational property. But it also makes nonsense of standard discussions of well-being, including Kagan's own. Kagan identifies three different kinds of theory of well-being: hedonistic theories, desire-satisfaction theories, and objective list theories.[14] Hedonistic theories hold that only a person's experiences of pleasure contribute to his well-being. Desire-satisfaction theories hold that only states of affairs that a person desires noninstrumentally contribute to his well-being. Objective list theories hold that having certain substantive goods other than pleasure, such as friendship and knowledge, contribute to a person's well-being, and contribute independently of whether this person desires them. If a person's well-being consists only of his intrinsic properties, then it makes no sense to characterize desire-satisfaction and objective list views as theories of well-being, because they entail that some relational properties contribute to well-being. Desire-satisfaction theories, for example, entail that what is good for a person is a relation between his attitudes and the world: that the world is the way that he wants it noninstrumentally to be. And standard objective list views hold that a person is made intrinsically better off by having friends and knowledge, both of which involve relational properties. So if well-being consists only of intrinsic properties, as Kagan sug-

[12] Shelly Kagan, "The Limits of Well-Being," *Social Philosophy and Policy* 9, no. 2 (1992): 181.

[13] Ibid., 186.

[14] Kagan, *Normative Ethics* (Boulder, CO: Westview, 1998), 29–41.

gests, he should not have presented desire-satisfaction and the standard objective list views as genuine alternatives to the hedonistic view. He should have said that as theories of well-being they are both nonstarters, because as theories of well-being they are conceptually incoherent, and that if they are theories of anything, they are theories of something else entirely.[15]

If we assume, then, that fairness is a personal value, and that it is an intrinsic value that cannot be reduced to well-being, and that neither an experiential nor an intrinsic property analysis of the concept of well-being is correct, then we face the challenge of trying to explain the difference between well-being and other personal values. The restricted rational desire analysis cannot explain this for the reasons given above, but two other analyses might be able to do so: a rational care analysis and a rational sympathy analysis.

II. The Rational Care Analysis

Stephen Darwall has proposed a rational care analysis of welfare. "Welfare," as he uses the term, is synonymous with "well-being," as I am using the term here. Darwall's analysis is that p is good for S =df. one should want p insofar as one cares about S.[16] This analysis seems vulnerable to the following objection. According to this analysis, p is good for S if the following conditional is true: if one cares for S, one should want p. It seems to follow as a matter of logic that if one does not care for S, then regardless of whether one should want p, p is good for S, including, for example, excruciating pain. Darwall, however, makes it clear that what he means by "insofar as one cares for S" is "on a hypothesis one accepts or is committed to in caring, namely that the cared for is *worth* caring for."[17] So Darwall's analysis is more accurately stated this way: p is good for S =df. one should, if S is worthy of care, want p. Assuming, then, that S is worthy of care and it is not the case that one should want

[15] Kagan might in fact believe that desire-satisfaction and objective list views are not theories of well-being, but of something else. In "Me and My Life," *Proceedings of the Aristotelian Society* New Series 94 (1994): 318, he makes a distinction between a person being well off and a person's life going well. Perhaps Kagan believes that the intrinsic property analysis is the correct understanding of what is good for a person, but not of what makes a person's life go well. He might think, for example, that although having true friends would not in itself be better for a person than having false ones, it might still make this person's life go better. Perhaps Kagan believes that when philosophers use the term "well-being" they are confusing two different ideas. Desire-satisfaction theories and objective list theories are not, in fact, theories of well-being properly understood; they are theories of something else: how well a person's life goes (impersonally, and not for him). It remains the case that the intrinsic property analysis cannot possibly be an analysis of what other leading theorists mean when they use the term "well-being," because they clearly understand it in a way that allows relational properties to contribute.

[16] Stephen Darwall, *Welfare and Rational Care* (Princeton: Princeton University Press, 2002), 4, 7–9. Note to nonspecialist readers: "=df." means that the expression to the left of the symbol is being defined by the expression to the right of the symbol.

[17] Ibid., 8. See, also, 38.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

S to suffer excruciating pain, excruciating pain does not contribute to S's welfare.

Like the unrestricted rational desire analysis, however, this analysis also faces a scope problem. This is because there are many things that it is rational to want, including things that have nothing to do with S's welfare. It is rational for me to want my children to be happy, but it does not follow that the happiness of my children directly contributes to S's welfare. S might be entirely unrelated to my children and entirely unconcerned with them. If, however, S is worthy of care, it follows from this analysis that the welfare of my children is good for S, because it is rational to want. In fact, assuming that S is worthy of care, it follows from this analysis that every good thing contributes to S's welfare, because every good thing is rational to want.

A possible solution is to restate the analysis as follows: p is good for S =df. if S is worthy of care, one should want p for S. This in fact is how Darwall reformulates his view in response to some objections by Feldman.[18] This analysis, however, cannot go beyond the restricted rational desire analysis in explaining the distinction between well-being and other personal values. Suppose that p is the following state of affairs: not having less than a fair share of some good. Assume that p is something one should want for S, because one should want him to be treated fairly. Then it follows that p is good for S on this analysis. But if fairness is an intrinsic value that cannot be reduced to well-being, it would be wrong to conclude that p contributes to S's well-being. So this formulation of Darwall's analysis does not constitute an advance over the restricted rational desire analysis.

Darwall's core idea, however, suggests an analysis that might explain the difference between well-being and other personal values. The core idea is that what is good for a person is a matter of what it would be rational to want on the hypothesis that he is worthy of care. This suggests the following analysis: p is good for S =df. one should want p, and it would not be the case that one should want p if S were unworthy of care. As it stands, this formulation is open to another objection, which is that it might be false that one should want p even if p contributes to S's well-being. This would be the case, for example, if p involves an injustice that one should not want performed. This objection, however, might be addressed by the following reformulation: p is good for S =df. there is a way in which p is good and p would not be good in this way if S were unworthy of care.[19] This seems to address the injustice objection because,

[18] Stephen Darwall, "Reply to Feldman, Hurka, and Rosati," *Philosophical Studies* 130, no. 3 (2006): 642.

[19] If one wonders what it means to say that p is good in some way only because S is worthy of care, here is one interpretation: there is a reason to want p that is a good reason only because S is worthy of care. So: p is good for S =df. there is a good reason to want p that is a good reason only because S is worthy of care.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

even if p involves an unjust act that should not be performed, p might still be good in some way, and it might be good in this way only because S is worthy of care.

Another kind of objection to the rational care analysis is that it is circular. Darwall maintains that the best explanation of why S's welfare is normative—why facts about things being good or bad for S give us reasons to act or to want things to occur—is that S is worthy of care.[20] But one might think that the best explanation of why someone is worthy of care is that things are good or bad for him. If so, it seems the explanation of why the fact is normative that things are good or bad for S is, ultimately, that things are good or bad for S, which is not much of an explanation.[21] One might, however, also reasonably deny that the best explanation of S being worthy of care is that things are good or bad for him. One might think that the best explanation of S being worthy of care is instead that he is capable (or once was capable or will be capable) of having certain affective mental states, the kind that warrant sympathy. If so, then one might say that although the best explanation of why facts about S's welfare are normative is that S is worthy of care, the best explanation of why S is worthy of care is that S is capable of these affective mental states.

Here, then, is how I propose to understand the rational care analysis of well-being: p increases S's well-being =df. p is good in a way that it would not be if S were unworthy of care (independent of any consequence of p). S is worthy of care if and only if S is capable of certain affective mental states (yet to be identified). Darwall does not state his view this way. This analysis, however, does make central use of the idea of being worthy of care, and one might arrive at this analysis by reformulating Darwall's stated view in response to objections. Moreover, this analysis offers a possible explanation of the difference between well-being and other personal values.

Personal reasons of well-being for S to want p are good reasons for S to prefer his own situation if p that would not be good reasons if S were unworthy of care. Personal reasons for S to want p are not reasons of well-being if they would be good reasons for S to prefer his situation even if S were not worthy of care. Imagine here that although S is not worthy of care, because he does not have the kind of affective mental states that warrant sympathy, he is still worthy of respect, because he has the capacity for practical reasoning.[22] In this situation there might be personal reasons for S to want p that are not reasons of well-being because there might be good reasons for S to prefer his situation if p, even if S is

[20] Darwall, *Welfare and Rational Care,* 6.

[21] For a different circularity objection, see Connie Rosati, "Darwall on Welfare and Rational Care," *Philosophical Studies* 130, no. 3 (2006): 619–635.

[22] I imagine someone like this in "An Objection to Attitudinal Hedonism," *Philosophical Studies* 115, no. 2 (August 2003): 197–200.

unworthy of care, because S is nonetheless worthy of respect. For example, the fact that a principle will arbitrarily favor someone other than S might be a good reason for S to reject this principle because S is worthy of respect, and it might be a good reason even if S is not worthy of care, because he lacks the capacity to suffer.

III. The Rational Sympathy Analysis

The rational care analysis, however, is not the only analysis of well-being that might explain the difference between well-being and other personal values. There is also a rational sympathy analysis.

Brad Hooker suggests that a state of affairs is bad for a person if we would feel sorry for him in believing that this state of affairs occurs.[23] We might add that a state of affairs is good for a person if we would feel glad for him in believing that this state of affairs occurs. As it stands this is unsatisfactory. Even if something very bad happens to someone that makes him suffer terribly for the rest of his life, we might not feel sorry for him because we don't like him or because we just don't care about him.

A natural modification of Hooker's proposal is that a state of affairs contributes to S's well-being if someone with the normal amount of concern for people who does not have any negative attitudes toward S would feel glad for S in consciously believing p. But this does not guarantee the desired result either. Perhaps, although you have the normal amount of concern for people, you are too tired to feel glad for anyone right now. Under these conditions you might *say* upon hearing about p that you feel glad for S but really you might not feel anything at all.

Perhaps, in order to provide the proper test of whether a state of affairs contributes to a person's well-being, it is possible to come up with a longer list of conditions that a person must meet. If a person who meets all these conditions would feel glad for S in consciously believing p, p would contribute to S's well-being. At this point, however, we should wonder why *these* are the relevant conditions. Why does feeling glad or sorry for a person under *these* conditions provide the proper test of welfare or ill-fare? The best answer, it seems to me, is that they are the conditions under which it is most likely that a person will respond properly to the relevant considerations: under these conditions a person is most likely to recognize and respond properly to the factors that are genuinely good reasons to feel glad or sorry for S. But if this is the correct explanation, we might simplify the account as follows: p increases S's well-being =df. p is a good reason to feel glad for S (independent of any consequence of p); p decreases S's well-being =df. p is a good reason to feel sorry for S (independent of any consequence of p). I call this the

[23] Brad Hooker, "Does Moral Virtue Constitute a Benefit to the Agent?" in Roger Crisp, ed., *How Should One Live?* (New York: Oxford University Press, 1996), 149, 155.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

rational sympathy analysis of well-being because it holds that whether something increases or decreases a person's well-being depends on whether it is a good reason to feel a certain kind of sympathy.

The rational sympathy analysis is better than the unrestricted rational desire analysis because it avoids the scope problem. Although it might be rational for me to desire the welfare of the stranger on the train, the fact that he fares poorly would not be any reason to feel sorry *for me.* The rational sympathy analysis is also better than the restricted rational desire analysis because it offers an explanation of the difference between well-being and other personal values, such as fairness. To see this, consider two principles. One arbitrarily disfavors you; the other arbitrarily disfavors no one. This is a good reason for you to reject the first principle. Is it a reason of well-being? Assume that although the first principle arbitrarily disfavors you, you do not know this and you will know of no action that this principle unfairly permits and you will not lose any substantive good as a result of an action that this principle unfairly permits. On these assumptions, it makes sense to hold that this reason of fairness for you to reject this principle is not a fact about your well-being and that it cannot be reduced to such facts. We might now explain this in the following way: the fact that you will be arbitrarily disfavored by this principle is not, on these assumptions, a good reason to feel sorry for you. Scanlon's rejection of the Complaint Model can then be understood as follows: there are good personal reasons to reject principles that are not good reasons to feel sorry for anyone. Contractualism is not a form of philosophical utilitarianism because there are personal reasons to reject principles that are given by states of affairs, such as facts about fairness, that are not good reasons to feel sorry for anyone.

Another reason to like the rational sympathy analysis is that it provides a plausible explanation of what hedonists and non-hedonists about well-being are disagreeing about and why their disagreements are so intractable. Is Jones better off than Smith, even if Smith will never know that his family secretly hates him? Traditional hedonists say that Smith is no worse off than Jones, because his experience of life is just as pleasant. Theorists who reject traditional hedonism say that Jones is better off than Smith.[24] The rational sympathy analysis enables us to understand what this disagreement is about. These theorists are disagreeing about whether the fact that Smith's family secretly hates him is a good reason to *feel sorry* for him, given that he is unaware of this. This disagreement is intractable because both positions make sense. To some it seems perfectly obvious that there is a good reason to feel sorry for Smith: his family secretly hates him. Since, however, this state of affairs involves no *suffering* on his part,

[24] But for a nontraditional hedonist view that allows one to say that Jones's life goes better for him than Smith's life goes for him, see Fred Feldman, *Pleasure and the Good Life* (New York: Oxford University Press, 2004), 112.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

it seems just as obvious to others that this is no reason at all to feel sorry for him.

A predictable objection to the rational sympathy analysis is that sometimes there is no good reason to feel sorry for someone even if he is suffering. Suppose a vicious murderer is suffering while awaiting execution. Clearly this suffering is bad for him, but, one might argue, there is no good reason to feel sorry for him. After all, he is a vicious murderer who deserves to suffer. This might lead some to reject the rational sympathy analysis. This, however, would rest on the mistaken assumption that there is no reason to feel sorry for someone who deserves to suffer. Even if a vicious murderer is getting what he deserves, his suffering is still a good reason to feel sorry for him. Perhaps no one owes it to him to feel sorry for him. Perhaps no one is open to criticism for not feeling sorry for him. Perhaps there is no good reason for his victims to feel sorry for him. Perhaps the fact that he has victimized them "silences," "cancels," or "blocks" any reason there might have been *for them* to feel sorry for him. It remains the case that his suffering is a consideration that counts in favor of someone suitably situated feeling sorry for him, a reason that would make this attitude appropriate in the absence of countervailing reasons. It would be a good reason for his mother or Jesus to feel sorry for him.[25]

Another predictable objection is that the rational sympathy analysis gets the order of explanation wrong. "Surely," someone might say, "whether a state of affairs is a good reason to feel sorry for a person depends on whether it decreases his well-being, not the other way around." For the reasons given in the first section, however, we need an analysis of well-being in order to understand the difference between well-being and other personal values, such as fairness. Because the rational sympathy analysis is not formally circular, and because it addresses the scope problem, and because it is not open to any obvious counterexample, and because it illuminates the difference between well-being and other personal values, this analysis of well-being might be both correct and informative.

IV. Comparison

Both the rational care and the rational sympathy analyses offer possible explanations of the difference between well-being and other personal values, which should support the sense that the distinction between them

[25] Note, too, that this objection to the rational sympathy view does not seem to provide a decisive reason to favor the rational care view. The rational care view says that p is bad for S if and only if p is bad in a way that it would not be bad if S were unworthy of care. So if S is worthy of care, then p can be bad for S even if we should not feel sorry for S because he is so evil. But if S is worthy of care, then it seems we should feel sorry for S even if he is evil. And if he is not worthy of care, then the rational care analysis does not support the conclusion that his punishment is bad for him, since now there is no way in which it is bad that it would not be bad were S unworthy of care. S's suffering might still be impersonally bad, but it would be bad in this way even if S were unworthy of care.

is genuine. Is there any reason, though, to think that one account is true and the other is false?

A possible objection to the rational sympathy analysis is that it is open to the "wrong kind of reasons problem" that all "fitting-attitude" accounts of evaluative properties face. A fitting attitude account of an evaluative property maintains that a state of affairs has this evaluative property if and only if it is fitting to have some attitude toward this state of affairs. A state of affairs is enviable or shameful, for example, if and only if it is fitting to feel envious or ashamed of it. Fitting attitude accounts of this kind have been criticized on the ground that sometimes there is good reason to have an attitude toward a state of affairs that does not warrant ascribing the relevant evaluative property to it.[26] Imagine that an evil demon will torture you unless you start feeling ashamed of your admirable and praiseworthy accomplishments. This seems like a good reason for you to feel ashamed of your accomplishments, but this reason does not warrant the judgment that these achievements are shameful because it is the *wrong kind* of reason. So, it seems, either this account of shamefulness must be supplemented with an account of the *right* kind of reasons or we should reject this kind of account of shamefulness altogether.

If the rational sympathy analysis of well-being is vulnerable to this kind of objection, this is a good reason to prefer the rational care analysis. According to the rational care analysis, p contributes to S's well-being if and only if p is good in a way that it would not be good if S were unworthy of care. Because this analysis does not include any claims about fitting attitudes, it is immune to the wrong kind of reasons problem.[27]

Perhaps, though, the wrong kind of reasons problem is a bogus problem. Consider again the possibility that someone will be tortured unless he feels ashamed of his admirable and praiseworthy accomplishments. This is a good reason for him to *want* to feel ashamed of his praiseworthy and admirable accomplishments, but, one might say, it is not a good reason for him to actually feel ashamed of them.[28] Likewise, the fact that

[26] See, for example, Wlodek Rabinowicz and Toni Ronnow-Rasmussen, "The Strike of the Demon: On Fitting Pro-attitudes and Value," *Ethics* 114, no. 3 (2004): 391–423.

[27] One might argue that the rational care analysis has its own wrong kind of reasons problem. What is it, after all, for x to be good in a way that it would not be good if S were unworthy of care? One might answer that it is for there to be a reason to want x that is a good reason only because S is worthy of care. If so, then the rational care view seems to face a wrong kind of reasons problem of its own because it makes implicit use of the notion of "reasons to want." The way in which it makes use of this notion, however, blocks the objection. The fact that it is rational to want S to suffer if one will be tortured unless one wants S to suffer does not render it true, on this analysis, that suffering is good for S, because it is not the case that this is a good reason *because* S is worthy of care. One might now puzzle over what it is for a reason to be a good reason because S is worthy of care, but this is another issue. So the rational care analysis is not vulnerable to the wrong kind of reasons problem in the way that the rational sympathy analysis appears to be.

[28] See Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990), 37, and Derek Parfit, *On What Matters,* vol. 1 (New York: Oxford University Press, 2011), 50–51, 420–32.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

I will be tortured unless I feel glad that someone else is suffering might be a good reason for me to *want* to feel glad that he is suffering, but it would not be a good reason for me to actually feel glad *for him* that he is suffering.[29]

If the wrong kind of reasons problem is bogus, we might prefer the rational sympathy analysis because it seems to offer a more illuminating explanation of the truths that the rational care analysis generates than the other way around. When I ask myself why a state of affairs would be bad in a way that it would not be bad if S were unworthy of care, it seems that a good answer is that this state of affairs provides a good reason to feel sorry for S. If, on the other hand, I ask why something is a good reason to feel sorry for S, the answer that this thing is bad in a way that it would not be bad if S were unworthy of care seems unilluminating.

According to the rational sympathy analysis, p is good for S if p is a good reason to feel glad for S. All such reasons presuppose, or entail, that S is worthy of care. So the notion of being worthy of care is implicit in the rational sympathy view. The fact that someone is worthy of care, however, does not seem to play an important role in explaining *why* p is a good reason to feel glad for S. If we think that a person is worthy of care in virtue of certain affective mental states, then the fact that a person is capable of these mental states explains why there are *any* reasons to feel glad or sorry for this person. It does not, however, tell us much about why a *particular* reason is or is not a good reason to feel glad or sorry for someone.

Consider, for example, the question of whether the fact, p, that Smith's happiness is based on false beliefs about his family's love for him, is a good reason to feel sorry for Smith. Suppose p is a good reason. Is this *explained* by the fact that p is bad in a way that it would not be bad if Smith were unworthy of care? It is true that if Smith were unworthy of care, p would provide no good reason to feel sorry for him, and so, presumably, there would be one less way in which p is bad. On the other hand, it is arguable that the best explanation of why there is one less way in which p is bad is that, if Smith is unworthy of care, p provides no reason to feel sorry for him. Hence it seems that facts about reasons for sympathy provide a better explanation for why p would be bad in one less way on the hypothesis that a person is unworthy of care than the other way around.

For this reason I like the rational sympathy account better, but I see no decisive objection to the rational care account, and I conclude that both

[29] One might offer a similar defense of the restricted rational desire analysis of well-being. Some might argue that if I will be tortured unless I want someone else to suffer, then I ought to want this person to suffer. Because this does not make it true that this person's suffering is good for him, one might conclude that the restricted rational desire analysis is wrong. A reasonable response is that although the fact that I will be tortured unless I want someone else to suffer might be a good reason for me to want this person to be tortured, it is not a good reason for me to want this *for him,* and so does not present a counterexample to the restricted rational desire analysis.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

are live options. Both offer explanations of the difference between well-being and other personal values, and both explanations employ, in one way or another, the distinction between two ways in which something can be good for a person. Something can be good for a person in virtue of his having positive and negative affective states and so in being the kind of being who is worthy of sympathetic concern. Something can also be good for a person in virtue of his having the capacity for practical reason and so in being the kind of being who is worthy of respect. Although neither the rational care nor the rational sympathy analysis is fully worked out here, this similarity nonetheless indicates what the fundamental difference between well-being and other personal values might ultimately consist in.

V. The Concept of Well-Being

How well, though, does either analysis capture the intuitive idea of well-being and identify what is important about it? Each of us can ask, "What is the best kind of life for me to have?" When we ask this question without assuming the truth of hedonism and without assuming that the best life for us involves always doing the morally right thing, we are asking about what ethical theorists have called *well-being.* The rational care and rational sympathy views, however, both seem strangely disconnected from this question. Does the fact that something is good in a way that it would not be if I were unworthy of care give me any insight into what the best kind of life is for me? Is this question fully answered by citing all the good reasons there are to feel glad or sorry for me?

The philosophical idea of well-being is odd, though. Scanlon characterizes it this way: "The intuitive notion of well-being . . . is an idea of the quality of life for the person who lives it that is broader than material or social conditions, at least potentially broader than experiential quality, different from worthiness or value, and narrower than choiceworthiness all things considered."[30] But what lies between experiential quality and choiceworthiness? If I ask myself what the best kind of life is for me, and I am thinking of experiential quality, then the answer is obvious: the most pleasant or enjoyable life. If I am not thinking of experiential quality, it seems I am asking what kind of life will give me least reason for regret, which is the question of choiceworthiness. What do we miss, then, if we set aside the philosophical notion of well-being and confine our attention to experiential quality and choiceworthiness?

The philosophical notion of well-being was invented, I believe, to perform three distinct, but interconnected tasks. First, to identify what we are wondering about in our ordinary lives when we ask ourselves, what is the best kind of life for me to have? Second, to identify what hedonists,

[30] Scanlon, *What We Owe to Each Other,* 112–13.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

desire-satisfaction theorists, and objective list theorists are disagreeing about. Third, to identify what the best utilitarian moral theory would direct us to maximize. When I reflect on what the best kind of life is for me, it seems that I am asking something different from what kind of life will be the most fun and also different from what kind of life would be the most choiceworthy. I can meaningfully ask whether my life would be missing something important if I just had fun. I can also meaningfully ask whether the most choiceworthy life would be the best one *for me.* Perhaps the most choiceworthy life is one devoted to helping others or to developing my talents as fully as possible. But if a life devoted to helping others would not develop my talents as much as one devoted to artistic or scientific pursuits, or if a life devoted to developing my talents would not be much fun, then it makes sense to question whether the most choiceworthy life is the best one *for me.*

It also seems that hedonist, desire-satisfaction, and objective list theorists are genuinely disagreeing about something, and not merely misunderstanding each other. When Nozick challenges hedonism by asking whether we would choose to lead our entire lives hooked up to an imaginary experience machine, he is not denying that a life attached to the experience machine would have the greatest positive experiential quality. He is questioning whether a life with the greatest positive experiential quality is really the best kind of life for us to have.[31] And when Feldman defends hedonism against this kind of challenge, he is claiming more than that the most pleasant lives have the greatest positive experiential quality; he is making the substantive claim that, contrary to what desire-satisfaction and objective list theorists hold, the goodness of a life for the person who leads it is entirely a function of the moments of pleasure and pain it contains.[32] If Feldman were claiming only that the experientially best life for us is the most pleasant one, Nozick would no doubt agree. But there seems to be a genuine disagreement here. *Well-being* is the name that ethical theorists have given to what this disagreement is about.

Finally, it makes sense for a utilitarian to hold that the right action is the one that maximizes well-being without holding either that well-being consists solely of pleasure or that the most choiceworthy life necessarily contains the most well-being. The most choiceworthy life according to utilitarianism is the one that *results* in the most well-being overall, not the one that *contains* the most well-being, and one need not be a hedonist to believe that ultimately all that matters morally is the impact of our actions on well-being. So when we consider the various roles that the notion of well-being was invented to play, it seems we would be missing something if we set this notion aside and confined our attention solely to experiential quality and choiceworthiness.

[31] Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books: 1974), 42–45.
[32] See Feldman, *Pleasure and the Good Life*, 9–20.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

The challenge I identified at the outset is to explain the distinction between this notion of well-being and other personal values, such as fairness. This distinction is not well-explained, I argued, either by the intuitive idea of what is intrinsically good for a person or by the more theoretical idea of what it is rational to want noninstrumentally for a person. So if neither the rational care analysis nor the rational sympathy analysis provides a good explanation, we should keep looking. It is possible, though, that after looking further we will conclude that the notion of well-being is inherently unclear, even confused. Perhaps there is in reality no clear notion of the best life that is different from both the notions of the experientially best life and the most choiceworthy life. Perhaps there is in reality no clearly identifiable fact that hedonists, desire-satisfaction theorists, and objective list theorists are disagreeing about. Perhaps utilitarians should abandon welfarism the moment they reject hedonism and adopt some non-welfarist consequentialism instead, one that directs us to act so as to bring about the most valuable world, as opposed to the world with the optimal amount of well-being. Perhaps there is no satisfactory analysis of the concept of well-being because the notion is inherently unclear or confused. How serious a problem would this be?

At the outset I observed that Scanlon's contrast between contractualism and philosophical utilitarianism is obscure if well-being consists of everything that is intrinsically good for us. But the obscurity of this distinction might not matter very much. One might see the notion of well-being, as used by ethical theorists, as a theoretical construct, one intended to refer to whatever is noninstrumentally good for a person.[33] If fairness is one of those things, so be it. In that case, there is one less reason to reject the Complaint Model. Rejection of the Complaint Model is not essential to contractualism, however. Scanlon rejects it partly because he believes that fairness is not an element of well-being and cannot be reduced to any element of well-being. If he is mistaken about this, Scanlon could still hold that fairness is a personal value that cannot be reduced to other elements of well-being.

Scanlon's contractualism also does not make essential use of the notion of well-being, and it is possible to present and defend this moral theory, and to employ it in our reasoning about right and wrong, without using this notion at all. In our ordinary nontheoretical reasoning about right and wrong, we consider individuals' interests and claims and make judgments about their relative weight, typically without appealing to the notion of well-being at all. We also do not make much use of this notion in our first person deliberations about what to do or how to live.[34] So there is no need for a contractualist theory of morality to give the notion

[33] Thanks to Doug Portmore for suggesting this in discussion.
[34] See Scanlon, *What We Owe to Each Other,* 126–33.

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press

of well-being an important role to play or to make the precise content of this notion clear.

The difference between contractualism and other theories of morality can also be understood and appreciated without placing any weight on the distinction between contractualism and philosophical utilitarianism. Even if there is no clear distinction between contractualism and what Scanlon calls *philosophical* utilitarianism, there is still a clear distinction between contractualism and *act* utilitarianism. Act utilitarianism directs us always to act so as to bring about the greatest net happiness aggregated over individuals, regardless of how these actions impact specific individuals and regardless of their objections to our acting in this way. Contractualism, in contrast, holds that if an individual's reason to prefer his own situation when no one is permitted to perform an action has greater weight than anyone's reason to prefer his own situation when he or others are permitted to perform it, then the action is impermissible, regardless of how much aggregate happiness it would produce.

Finally, contractualism offers a distinctive view of the importance of morality. Contractualism identifies the importance of doing the right thing with the importance of being able to justify our conduct to each other. This differs from consequentialist views that identify the importance of doing the right thing with the instrumental value of acting so as bring about the most valuable world. Scanlon's main point in contrasting contractualism with philosophical utilitarianism is that contractualism offers a more plausible account of the importance of morality. Failing to act so as to bring about the most well-being is not a serious failing, whereas treating each other unjustifiably is. In failing to act so as to bring about the most well-being, we fail to produce something of value, but in failing to act in ways that we can justify to each other we fail to value each other properly as rational beings, as beings who can recognize and assess reasons. This changes the nature of our relations with each other and warrants the painful self-reproach of guilt, in a way that merely failing to produce something of value does not. If an action is permitted by a principle that no one can reasonably reject on his own behalf, then we can justify this action to him, by explaining why he could not reasonably reject every principle that permits it. If our conduct adversely affects someone and is permitted only by principles that he could reasonably reject, then we cannot justify ourselves to him in this way. Because acting only in ways that we can justify to each other is a condition of valuing each other properly as rational beings, doing the right thing is a condition of mutual respect. Acting so as to bring about the most valuable world, in contrast, is not. This is not to deny the core consequentialist commitment that there is always sufficient reason to act so as to bring about the most valuable world. It is to say that moral wrongness cannot be plausibly explained as a failure to do so, because whereas wronging others is a serious failure, failing to act so as to bring about the most valuable world

is not. Scanlon's contractualism thus offers a distinctive view of the importance of morality regardless of whether the concept of well-being can be elucidated so as to produce a clear and interesting contrast between contractualism and philosophical utilitarianism.

We might still be curious, though, about the concept of well-being itself. What, exactly, are we asking when we ask what the best life is for us (and we are not asking about experiential quality or choiceworthiness)? What, exactly, are hedonists, desire-satisfaction theorists, and objective list theorists disagreeing about? Any satisfactory answer to this question must allow that all three views are live options, epistemically possible, and thus conceptually coherent, which neither the experiential nor the intrinsic property analysis allows. A restricted rational desire analysis can explain what these theorists are intelligibly disagreeing about, but it does not adequately explain the difference between well-being and other personal values such as fairness. The rational sympathy and rational care analyses might be able to explain this difference, but both seem strangely disconnected from the central question we are interested in, which is what the best life is for us. Can any analysis of well-being do everything we want from it? This remains to be seen.

Philosophy, Arizona State University

https://doi.org/10.1017/S0265052513000034 Published online by Cambridge University Press