

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Utilitas**.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/43124>

Published paper

Shemmer, Y. (2011) *Full Information, Well-Being, and Reasonable Desires*,
Utilitas, 23 (2), pp. 206-227

<http://dx.doi.org/10.1017/S0953820811000069>

Full Information, Well Being, and Reasonable Desires¹

Yonatan Shemmer – The University of Sheffield

Introduction

Griffin in his book ‘Well Being’ claims that we should see desires as normatively subordinate to what is good. It is only reasonable, Griffin says, to desire what is desirable, valuable or good². Since on such view what it is reasonable to do or to desire is determined by what is good, we must have an independent notion of what is good. But the history of philosophy is paved with failed attempts to find such independent notion of goodness. It is the despair generated by these failures that led major philosophers from Hobbes onward to invert the order of priority and base what is good on what we desire. According to the simple desire based view to say that something is good for someone is to say that the person desires it. One central problem with the simple desire based view is that we are sometimes uninformed and then our desires are based on a mistaken picture of the world and of ourselves in it. Railton tries to improve on the simple desire based view. He suggests that we should base our understanding of value only on well informed and rationally generated desires. Why? Because these are the desires that we should want ourselves to have. What is valuable, on this view, is what we could reasonably want ourselves to desire if we were fully informed.

Unfortunately, as I will try to show, what we could reasonably want ourselves to desire if we were fully informed does not match what we intuitively think is part of our well being. Either our intuitions or the desire based theory of well being must be abandoned.

In the first part of this paper I briefly trace the dialectic leading to the formulation of Railton’s view and present some famous criticisms made against his view. In the second part I consider four possible interpretations of his view and argue that only one of these can provide us with a coherent and determinate understanding of well being. In the

¹ Thanks to Jules Holroyd, Jimmy Lenman, Elijah Millgram, Shaul Smilanski, and John Skorupski for their helpful criticisms and suggestions.

² James Griffin, *Well Being*, OUP 1986 (p. 27)

third part I argue that this fourth interpretation clashes with ordinary intuitions about well being.

1. The Simple Desire Based View, Railton and Critics

Consider the following formulation of the simple desire based view as an analysis of the notion of well being. According to this analysis:

A: *'x is good for me (x is part of my well being) iff I desire x'.*

It is easy to see that this analysis fails. I may desire to take a very strong drug not knowing the negative long terms effects of the drug. In this case what I desire is certainly not what is good for me.

Full information theories try to provide a more sophisticated analysis of well being in terms of desires in order to deal with this type of counterexample. The basic thought behind all full information theories is that we have to understand well being not in terms of what the agent desires, but in terms of what she would desire if she had the relevant information. Thus we get:

B: *'x is good for me iff I would desire x were I to be presented with full information about x and the implications of bringing about x'.*

It is often claimed that as stated in 'B' full information views are incomplete. Information about the objects of one's desires is insufficient. One has to be informed about all possible alternatives in order for one's desires to be well grounded³. Furthermore being exposed to relevant information is not enough. I may be the kind of person who is not capable of understanding or of processing the information properly and thus I may not absorb the 'lesson' of the information to which I was exposed. Therefore full information theories ought to require that well being be determined by my desires only if I am properly informed and only if I am given the capacity to understand the information to which I am exposed⁴. However talking of 'the capacity to understand information' may not be sufficient either. In order to form a significant desire about which one of two movies to see, merely having information about the movies and being able to understand this information is not going to suffice either. I also need to know

³ I am not sure this criticism is cogent, but this is a topic for another paper.

⁴ For such a requirement see W. D. Falk, "On Learning About Reasons," in *Ought, Reasons, and Morality* (Ithaca, N.Y.: Cornell University Press, 1986).

what effect the two movies would have on me. Only someone who experiences the movies could know what effect the movies have on him. Finally I must be rational so that the desires that I form will be rationally grounded in my other desires and the information that is available to me.

Full information theories can accommodate these additional requirements. For example a full information theory can claim that:

*C: 'x is good for me iff I would desire x under the following conditions:
a. I am presented with full information about all the possible lives that are available to me, b. I have the capacity to understand the information I am presented with, c. I have experienced all the possible lives that are available to me, d. I am instrumentally rational'.*

Call the person I would be if I were fully informed, had increased intellectual capacities necessary to understand that new information, had experienced that additional information and were rational my fully informed self or FIS; call my actual self AS.

Many have argued that this extension of full information theories also runs into problems. For example, it has been argued that under this analysis what is good for me depends on the desires of a person (my FIS) that is in significant ways different from me. That person has intellectual capacities I do not have and experiences I do not have. What my FIS desires with regard to x, seems to have very little bearing on whether x is good for me given the vast differences between us. If, for example, I am an uneducated laborer and my FIS has learned math and experienced the joy of proving complicated theorems, then his desire to tackle a complicated mathematical question has little relevance to what constitutes a good life for me.

Peter Railton has suggested a formulation of the analysis that full information theories give to the concept of well being that would avoid this problem:

An individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and

*his circumstances, and entirely free of cognitive error or lapses
of instrumental rationality⁵.*

The suggestion slightly reformulated is this:

D: *'X is good for me iff my FIS while contemplating my situation would want me
to want x'.*

This version of the view is sometimes referred to as the 'advice model' or AM. Many objections have been raised against AM as well. I will not recount all of them. But I will briefly mention here two objections raised by Sobel and Rosati⁶, mainly in order to keep clear the distinction between their objections and mine.

Both Sobel and Rosati object that the notion of being fully informed is incoherent. One thing that I ought to be informed about in order to be fully informed is about how events would be experienced by people with different sorts of lives. In fact in order to be fully informed I would have to experience any event as it would be experienced by people with all possible sorts of lives⁷. However to decide what is the best way to experience an event I would have to compare the different experiences and thus inhabit all potentially conflicting points of view simultaneously. But experiencing an event from a point of view excludes experiencing it from a conflicting point of view⁸. Therefore, being fully informed requires one to be in an impossible position. I think this argument is successful against the view as currently stated. But in order to present my own argument I will temporarily put aside Sobel and Rosati's argument. I will do so in the hope of highlighting a different – and to my mind interesting – problem for fully informed desires views of well being⁹.

⁵ Peter Railton, "Facts and Values", *Philosophical Topics* XIV, No.2 (Fall 1986): 16. See also Peter Railton, "Moral Realism", *Philosophical Review* 95 (1986): 173-4.

⁶ Connie Rosati, "Persons, Perspectives, and Full Information Accounts of the Good," *Ethics* 105, no. 2 (1995): 296-325, David Sobel, "Full Information Accounts of Well-Being," *Ethics* 104, no. 4 (1994): 784-810.

⁷ At least with all the forms of life and personalities available to me.

⁸ Rosati and Sobel's objections are much more nuanced and detailed – but I take it that this is the gist of it.

⁹ What I will come to see as the only coherent version of Railton's view can avoid Sobel and Rosati's argument. I will not, however, discuss this in the current paper.

Rosati also presents a different objection to fully informed desires views. Her objection is based on the assumption that full information theories wish to respect the requirements set by motivational internalism. One demand set by motivational internalism, says Rosati, is that “The information that you would desire X for yourself under the specified conditions [the conditions of full information] will have pull for you, even if X itself still fails to move you¹⁰”, or maybe in a less stringent form that the “... actual person would care about the fact that something would prompt her concern under [the ideal] conditions¹¹”.

But, says Rosati, even if a person were to believe that her fully informed self desires x for her, she might not be motivated to get x herself, since she may “contend that the fully informed person [is] not really her¹²”.

Rosati herself is not sure about the force of her argument. She thinks that we should not require that our ideal self will be the same person as our actual self in the ‘ordinary sense’. “As long as the fully informed standpoint speaks to those concerns that led us to seek the bird’s-eye point of view in the first place, we can accord normative authority to the reactions of the person who occupies that standpoint¹³.” Such normative authority might be all the recommending force that a motivational internalist wishes to find in judgments about the good.¹⁴

Whether successful or not, Rosati’s argument depends on a prior internalist constraint on views of well being. The argument I will present in the remainder of this paper may appear to be similar to Rosati’s argument. But in contrast to Rosati’s arguments it does not presuppose internalism¹⁵.

¹⁰ Rosati, "Persons, Perspectives, and Full Information Accounts of the Good." p. 303.

¹¹ Ibid., p. 301.

¹² Ibid., p. 311.

¹³ Ibid., p. 311.

¹⁴ Rosati proceeds to argue that there are other problems with the normative authority of our fully informed self. The process that makes the actual self fully informed introduces so many changes to her personality that we cannot be sure that character traits that are normatively repugnant are not introduced into her personality. Ibid., p. 312.

¹⁵ For other criticisms of full information views see Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Harvard University Press and David Velleman, "Brandt's Definition Of "Good", "*The Philosophical Review* XCVII, no. 3 (1988): 353-371.

2. Four interpretations of Railton's view

The following example will help clarify the discussion. An FIS goes through a religious conversion. The conversion is rationally permissible but not required by the information available to him. From his point of view after the conversion it is desirable that I follow suit. From my perspective it is unreasonable to do so even if I am told that he would like me to. In cases in which the choices of my FIS are not rationally required by the information available to him, what is unreasonable for me to do constrains what he could justifiably want for me. The fourth interpretation I offer below is based on this thought.

According to Railton both the desires of the AS for herself and the desires of the FIS for herself are not indicative of the AS' non-moral good (or, well being). What constitutes the non-moral good of the AS are the desires of the FIS concerning the desires of the AS, when the FIS reflects on the AS' situation.

In this section I will assess this analysis of non-moral good. I will consider four interpretations of the view. On the first three interpretations the view must be rejected. Either it fails to explain how the desires of one's FIS can, on most but some rare occasions, be different from the desires of the actual self; or it ascribes unreasonable desires to the FIS, or it gives us an indeterminate account of what non-moral good is. The fourth and final interpretation is more successful on all these counts. It explains how the desires of one's fully informed self can significantly diverge from the desires of the actual self, it ascribes reasonable desires to the FIS, and it offers a determinate account of non-moral good. But this interpretation as well faces a serious problem. The notion of non-moral good that it delineates falls short of our intuitive notion of one's non-moral good.

Some constraints on any interpretation of the advice model

The advice model does not describe the criteria that the FIS is supposed to use in order to determine what to want for his actual counterpart. Such criteria are clearly required. The advice model appears to rely in its determination of non-moral good on the natural psychological mechanism of desire formation and desire change in the FIS. But

the advice model does not determine the well being of the AS by appeal to the new desires of the fully informed self for himself— as we have seen these might not be suitable for the actual self – rather the advice model claims that the AS’s well being is determined by those desires of the FIS that the FIS deems would be *appropriate* for the AS. The FIS must apply some standards in order to decide what, given his desires for himself, would be desirable for his AS. This fact is masked by the language of the advice model. To see that consider the relation between two different people. John desires that his neighbor clean their common path more often. Here John is merely hoping that his neighbor will act in ways that will promote his own (John’s) goals. Since John and his neighbor do not have the same goals what John desires for his neighbor is not be what is good for the neighbor. The situation might not be different for the FIS and his AS. Since they do not have the same goals merely asking the FIS what he desires that the AS desires will only reveal something about the FIS’ desiderative system. If the FIS wants a life of leisure for himself and wants a life of hard work for the AS – we should not conclude that hard work is good for the AS. Rather we should conclude that the FIS has a perverse desire to see his AS suffer. Those who suggest that we should appeal to the desire of the FIS for the AS hope that since the FIS and the AS are (in some sense) the same person, the FIS’s preferences will reveal what is good for the AS. But we already know that this is not the case, since we are asking the FIS not what he desires for himself but rather what he desires that the AS desires. Thus the fact that the FIS and the AS are the same person is relied on not because the desires of these two counterparts of the self are identical but rather because the proponent of the model hope that a person will have some good will towards his hypothetical counterpart. What we are really asking the FIS is to take his own desires for himself as a base, and perform the appropriate adjustments to fit the circumstances of the AS given his good will towards him. But this exercise requires the application of some standards in order to determine what is the right adjustment to perform, and the advice model does not supply them¹⁶.

The fact that we are asking the FIS to apply certain standards or perform an adjustment as part of the procedure that determines what is good for the AS is itself not

¹⁶ The worry here is similar to a worry discussed by Rosati as part of her argument from internalism. See: Rosati, "Persons, Perspectives, and Full Information Accounts of the Good.", p. 312.

unproblematic,¹⁷ but I want to focus on a more basic problem. The problem is that it is just unclear what kind of adjustment the FIS is supposed to make in order to move from his own desires to the desire he would like his AS to have.

One seemingly appealing answer to this problem is that the FIS is supposed to want his AS to want everything that he wants for himself minus those things that the AS is not going to enjoy. For example, the FIS wants fancy wine for himself, because he has experienced many good wines and developed a fine pallet, but he knows that the pleasure that the AS (whose pallet is not as sophisticated) is going to get from that wine is not worth its price. So he does not want the AS to want the fancy wine. The problem is that this way of explaining what the ‘appropriate adjustment’ is, relies on the very concepts we are trying to analyze – those of pleasure and well-being.

What could be the standards that should govern the advice of the FIS?

We should certainly require that the advice of the FIS would be reasonable. The desires the FIS should want the AS to have should be desires she has good reason to want the AS to have. Since the FIS is supposed to be fully rational you might think that this requirement will already be trivially satisfied. This is doubtful since the FIS is supposed to be instrumentally rational and as we have seen instrumental rationality will only ensure the satisfaction of the goals of the FIS, goals that may not match those of the AS. In any case, we shall make the requirement explicit.

1. The reasonability constraint: Any interpretation of the advice model should be constrained by the requirement that the advice of the FIS would be reasonable.

This constraint does not, however seem sufficient. It is not clear, at least not on the face of it, whether merely in virtue of being fully informed and fully reasonable the FIS will be benevolent or will care about the actual self’s autonomy. Therefore in order to get a satisfying theory of well being we might need to impose further desiderata on the nature of the FIS’ desires. We might want to add, for example that the FIS will be benevolent. But to add this constraint will make our analysis circular. Being benevolent is

¹⁷ The problem is that the initial impetus for a fully informed desires view of well being was the thought that by appealing to the desires of my FIS we are replicating the desires that someone like me would have with regard to a certain object but under improved epistemic conditions. It now turns out that we are appealing to a different process altogether – one that presupposes independent normative standards.

intending to promote in one's actions the well being of the person towards whom one is benevolent. If we allow the notion of benevolence into our understanding of the nature of the FIS we thereby include the notion of 'well being', the very notion we set out to analyze in the first place.

I will, therefore, focus here on a different desideratum. We must assume that the FIS will not want to rob the actual self of her autonomy. The FIS must want for the AS only desires that the AS could come to have without losing her autonomy.

There are two problems with this suggestion. First what autonomy is, is under great dispute. And second, one might think, that autonomy is just one among many goods that the FIS should want for the actual self and that the FIS should at least balance the goal of ensuring the agent's autonomy with the a concern for the maximizations of these other goods¹⁸.

To answer the first problem we have to find a minimal constraint that will be acceptable to most views of autonomy. I suggest the following constraint:

The advice of the FIS should be in principle justifiable to the actual self. If the FIS cannot explain why it makes sense for the AS to adopt the FIS' advice then it shouldn't be given. Of course, the actual self may not be in a position to understand the justification, but the justification must be such that it could be in principle intelligible to others who the AS has good reason to trust. Advice that cannot be justified even in principle to someone with sufficient information and who the AS would be rational to trust, loses its status as advice and becomes an imposition. And for that to be the case, the FIS should at the very least, be able to justify the advice to herself. Advice that she cannot justify even to herself is advice that she could not, even in principle, justify to the AS. Thus the autonomy constraint and the reasonability constraint seem to partially converge. The reasonability constraint is a pre-condition for the satisfaction of the autonomy constraint.

2. The minimal autonomy constraint: Any interpretation of the advice model should be constrained by the requirement that the FIS should not give the AS any advice which the AS could not even in principle reasonably accept. Where 'an

¹⁸ Thanks to Elijah Millgram for pressing me on that point.

advice can be reasonably accepted in principle' if there could be someone that the AS reasonably trusts that could see the justification for accepting the advice.

As for the second problem those who raise it have a point. It is often worth it to give up some of your autonomy in order to achieve other goals or other goods. We can however accept the central insight behind this objection while maintaining our minimal autonomy constraint: we say that it is legitimate for the FIS to advise the AS to act in ways that will harm her autonomy as long as the FIS can justify that loss of autonomy.

Finally any interpretation of the advice model should give us a determinate understanding of the agent's well being. To see the importance of this third constraint it is important to remember that on Railton's view we do not have direct access to one's non-moral good. We do not have a characterization of non-moral good in terms of either the state of the world that constitutes such good or the mental states that constitute non-moral good. The way to the content of one's non-moral good is oblique. We describe a procedure whose outcome, whatever it ends up being, determines one's non-moral well being. But for such oblique characterization to work the procedure must be determinate. Unless we have a determinate verdict about the desires of the fully informed agent, we do not have a clear analysis of the concept of well being.

2. Our analysis of the point of view of the fully informed self should give us – in principle at least – a determinate answer to the question: what would the fully informed person desire that the actual self desires¹⁹.

First interpretation

I will start by considering the view as stated above in **D**. What is good for the actual self according to this formulation is determined by desires of her fully informed counterpart. Which desires? Those desires of her fully informed self about the desires of the actual self in the circumstances that the actual agent finds herself in.

An example will help: I hold a glass of transparent liquid in my hand. I am about to drink it thinking that it is gin. In fact it is some kind of bitter lemonade²⁰. I dislike

¹⁹ There may be some cases in which there is more than one state of affairs that could be part of one's well being. There is no reason to expect the theory to select between these good states of affairs.

bitter lemonade. Fully informed desire views hope to explain why it is good for me not to drink the content of my glass by appealing to the desires that my FIS has that I desire not to drink the content of my glass. But we should consider the desires of my FIS more carefully.

My FIS, we are told, wants me to want not to drink the content of my glass. But how exactly does my FIS want me to come to have the desire not to drink the content of my glass? Up to a second ago I either had the desire to drink the content of my glass or I was in a psychological state from which I would naturally and rationally develop the desire to drink the content of my glass. So which exact changes in my psychology does my FIS hope for, changes that will ensure that I have the desires she wishes me to have? Since my FIS cannot change the past it must be that she wants me to have a change of desires now. But it's crazy for my FIS to want that I have such an unexplainable psychological leap; that I find myself with a desire that has nothing to do with what I wanted before, what my plans were and what information was available to me. Wanting me to be afflicted by such psychological rift is unreasonable and certainly incompatible with the autonomy constraint. So Railton's view *as phrased* fails straightaway.

We can certainly imagine extreme cases in which the consequence of a failure to change my desires is worse than the psychological rift caused by that change. In these extreme cases it would be reasonable for my FIS to want me to change my desires. But in most cases what my FIS could reasonably desire on my behalf – on the first interpretation of Railton's view - will not diverge from what I desire for myself and will greatly diverge from what we intuitively think is good for me.

You might think that my requirement that we assess the desires that the FIS has in light of the psychological state that I will find myself in if his recommendations about my desires be implemented, is one requirement too many. Why? Because, so you may think, the recommendations of my FIS about what desire I should have need not really be implemented. Talking about which desires my fully informed self would want me to have (when he considers my circumstances) is just another way of asking what is good for me. So whether I can realistically come to have these desires or not is quite irrelevant. It is

²⁰ This is of course a less dramatic version of Williams' gin and petrol example in "Internal and External Reasons," reprinted in *Moral Luck*, Cambridge University Press 1981, p.102.

quite possible, you might think, that I cannot come to desire what is good for me, at least not without loss of autonomy. It would be unfortunate but should not change our view about what is and what is not good for me.

This objection misunderstands the nature of Railton's proposal. As opposed to other views about the good, full information theories despair of the possibility of providing us with a list of states of affairs that are good for us, or a direct characterization of these states of affairs. As we noted before, their path is oblique. They claim that the way to determine what is good for a person is to describe a natural psychological process and claim that whatever desires emerge out of this process determine what is good for one. But on this approach the natural psychological process should make sense in its own right. We cannot – when we fail to make sense of the purported natural process – appeal to claims whose truth is supposed to be determined by that very process. Whether something is good for me or not is supposed to be determined by whether my fully informed self could desire it if she thought about my circumstances and at the very least cared about my autonomy. We can not turn around and explain how she can want me to want certain things, under these constraints, because these things are good for me. Doing so would be putting the cart before the horse. A full information view that would rely on judgments about what is good for me in order to determine the desires of my FIS would thus fail the determinacy constraint. It would fail to provide us with a determinate procedure for deciding what constitutes my well being.

My FIS could want me to want to be better informed, or could want me to want to be told what my FIS wants for me – and so we could conclude that it is generally good for me to be better informed or to know what my FIS wants me to want. But in order to reach these general conclusions we do not need a full information theory. I already want to be better informed.

Second interpretation

There may be a way of amending the view and thus saving it from this last criticism. Maybe my FIS wants me to have certain desires after she reveals to me all relevant information, or on the assumption that I have all relevant information.

E: *'X is good for me iff my FIS while considering my circumstances and imagining that I have all the information available to her would want me to want x'.*

Thus my FIS is supposed to assume that I know that the liquid in the glass is bitter lemonade and not gin and then ask herself what she desires that I desire. But this approach is also bound to fail. Presuming that I have all the information that my FIS has or giving me that entire information means treating me as identical to my FIS and thus finding my well being to be identical to that of my FIS. But the whole point of the Advice Model was to separate what is good for my fully informed self from what is good for me. What is good for my fully informed self is often good for her only because she has capacities, knowledge and experiences that I do not have. However on the second interpretation the two will coincide. We must therefore also reject this interpretation.

Third interpretation

A different approach must be sought. Maybe my FIS is not *just* supposed to want me to have certain desires but rather is supposed to want me to have these desires after I am told that my FIS wants me to have them. What my FIS wants is that I would want to have certain desires when it is disclosed to me that these are the desires he wants me to have. Call this view the 'Disclosure Advice Model' or for short 'DAM'.

F: *x is good for me iff my FIS while considering my circumstances and assuming that I know that he wants me to desire x, would want me to desire x.'*

This suggestion does solve some of the problems mentioned above. If I believe that my FIS wants me to have certain desires, and assuming that I want to have whatever desires my FIS wants me to have, and that I am reasonable, it is quite possible that I will naturally come to have these desires. So it's not crazy, even for an FIS who does not want to impose on me a psychological rift, to want me to have them under these conditions.

However we now see that this interpretation as well as the first two does not satisfy constraint number 1, in fact these interpretations do not address in any way the concern that led us to accept the first constraint. All three interpretations appeal to the desires of the FIS without specifying in any way the standards under which these desires

should be adopted. As I claimed above such standards are clearly needed. The FIS must ask herself not only what she desires for herself or for the actual self, but also what she *should* want the actual self to want given what she desires. To answer this question she must appeal to some justifying standards that will determine whether or not it is reasonable for her to desire that the actual self have certain desires.

Fourth interpretation

When we combine the lesson that pushed us toward DAM and our reasonability and minimal autonomy constraint we get the following condition:

An FIS must only desire for his AS desires that it would be rational for the AS to adopt if he learns that his FIS wants him to adopt them.

Whether it is rational for the AS to adopt a desire when he learns that his purported²¹ FIS wants him to adopt it, depends at least on whether he can understand the justification for adopting the desire, or at the very least on whether there could be anyone that the AS might reasonably trust that could understand that justification.

From the point of view of the FIS things look thus: we are asking my FIS to think what desires she could want me to want if a. she thought of herself as being in my shoes and if b. the list of desires she wanted me to adopt were constrained by the condition that it would be rational for me to adopt them if I learnt that she wants me to have them. Thus when my FIS deliberates whether to desire that I desire x or not, she cannot just stipulate that I will adopt any desire she wants me to adopt. If, concerning some desire for x that my FIS wants me to have, it is irrational for me²² to come to have this desire, then she should not want it for me. My FIS must ask herself what desires I could rationally come to have before she can decide what to desire to want for me.

DAM must then be interpreted along the following lines

G: *'x is good for me iff my FIS while contemplating my circumstances, and assuming that I am informed that she wants me to desire x, could reasonably want me to desire x, which depends on whether she could have good reasons to think that it would be*

²¹ Whenever I talk of the FIS as seen from the perspective of the AS I will qualify the term with the word 'purported'. As we will shortly see unless the AS has good reason to think that a piece of advice he gets is indeed reasonable he has reason to doubt that this advice originates from the FIS.

²² In the sense described above.

rational for me to adopt the desire for x if I thought that she wanted me to have it, which in turn depends on whether she could provide me or someone I could trust with a justification for adopting the desire for x.'

We should now consider our last interpretation of DAM (**G**).

Considering G

The first question we should answer is on what basis could my FIS have good reasons for thinking that it would be reasonable for me to adopt a desire she wants me to want. Suppose that I get a message informing me that my purported FIS wants me to want certain states of affairs. This of course would be quite a suspicious message. To believe that it is authentic I would need to believe four facts. First, that this message conveys to me information about someone who is me at a different time (that is, someone who developed from me or that I developed from). Second, I would have to believe that this person wants me to have the desires I am told she wants me to have. And third I would have to believe that she is fully informed and instrumentally rational.

Furthermore to have a good reason to act in accordance with my belief about the desires of my purported FIS I would have to think that she has a justification for wanting me to have the desire in question and she takes my autonomy into account when she decides which desires she wants me to have.

For the sake of argument let us assume that I can have good reasons to believe that the message is from someone who is me at a different time and that this person is fully informed and instrumentally rational. We are still left with the question of whether I have good reasons to believe that she has the desires in question and that she is justified in wanting me to have whatever desires she wants me to have. Could, then, I have good reasons to believe these two facts? The answer, I think, is mixed. There are some desires that I may have good reasons for thinking that my FIS wants me to have. And there are some circumstances in which I may have good reasons for thinking that she is justified in wanting me to have certain desires. There are however circumstances in which I could not have good reasons for believing in any of these purported facts.

Let us consider non-problematic circumstances first; circumstances in which I could have a good reason for believing that my FIS wants me to have certain preferences. Consider the situation I am in when I am about to drink a glass of petrol thinking it is gin. My FIS knows I have petrol in my glass and therefore wants me to lose the desire to drink the contents of the glass. Could I have good reasons to think that this is what she wants and that she is justified in so wanting? The answer is yes. I could in principle be presented with evidence convincing me that I have petrol in my glass. And I therefore could in principle set up a mechanism to verify such evidence and would then be justified in acting on it. Furthermore it makes sense for my FIS to want me not to drink petrol, since he knows that I would rather not die and so he knows that it would be reasonable for me to accept his advice. It is, therefore, in principle possible for me to come to believe that my purported FIS does not want me to drink the content of my glass and that his desire meets the reasonability and autonomy constraints.

Consider a problematic case in turn. Could I be sure that my purported FIS desires me to desire some ice cream that she thinks is wonderful? Imagine that I have only had chocolate ice cream before and my FIS after tasting vanilla ice cream wants me to have vanilla ice cream. It is *not* clear on what grounds I could rest my confidence that she wants me to have this ice cream or that she is justified in wanting me to have it. It is not clear what facts could provide me with sufficient evidence for the claim that my purported FIS is justified in wanting me to have vanilla ice-cream.

There are three limitations that prevent the FIS from determining what to want for me on the basis of her experiences.

i. Lack of instrumental relation

Assume for now that we do not doubt that the FIS had the experience she claims to have had when she tasted the vanilla ice-cream. The question is whether that experience can serve as the basis for justifying her desire that I desire to have the same ice cream?

The relation between one's experiences and one's preferences for oneself is not a normative relation. This relation becomes normative only when certain information links

it to other existing preferences of the self²³. When this is not the case the relation is merely a causal relation²⁴. And the relation between one's preferences and one's own well being, for a fully informed person, is a constitutive relation²⁵. But both these relations cannot ground the normative relation that should guide the FIS in determining what she would want me to want. The experiences of the FIS can be the basis of her desires for the desires of the AS only when they can be linked to a means-ends inference that is based on the preferences of the AS. In these cases there may be a normative link between the preferences of the FIS and her advice for the AS. If the FIS knows that studying for the exam in a certain way will lead me to fail and she knows that I want to succeed she may want me to avoid those choices that would lead to a failure. But the ice cream case is not of that sort. The FIS may taste two flavors of ice cream and prefer one over the other for herself. But unless these flavors have a characteristic that is somehow linked to what I desire (or should desire on the basis of other advice of the FIS) then the FIS' experience cannot help her determine what flavor she should want me to want.

Here we reach the Humean origins of the desire based view. Facts about the experience of the FIS, fundamentally (that is in cases that do not involve a means-end relation) do not have normative authority. For the FIS herself these facts have a causal connection with her preferences and then a constitutive connection with her well being. But the FIS must seek a normative justification before she gives advice to the AS – and therefore cannot ground her advice in her experiences.

An objector might argue against this conclusion thus: if the vanilla ice-cream is tastier to the FIS than the chocolate one, then it would be tastier to the AS if she had tried it. And if the FIS prefers it over chocolate that means that the AS would have preferred it over chocolate if she would have tasted both. What more does one need in order to show that the AS would be reasonable in adopting the advice of the FIS to want the vanilla over the chocolate?

This simple argument masks the important assumptions upon which it is founded.

²³ When I want A and my experience teaches me that the best way to achieve A is by doing B then my experience (given my previous desires) exerts normative pressure on me.

²⁴ Or so we would think if we adopted a desire based view of well being.

²⁵ See footnote 24.

On one interpretation the argument attempts to forge normative connections between the experiences of the FIS and his recommendations for the AS by an implicit reliance on a constant desire of the AS for pleasure. On that interpretation the experiences of the FIS reveal that this desire for pleasure can be satisfied. But part of the appeal of a desire based view of well being is its ability to give content to the slippery notion of pleasure of hedonistic or utilitarian views of well being. Indeed what the desire based view attempts to clarify is what it is for something to be part of one's pleasure or well being. But then a reliance on the notion of pleasure in the analysis would render the view circular.

Alternatively the objector might try to rely, not on the experiences or the pleasure of the FIS, but rather on the fact that the AS herself would have preferred vanilla to chocolate if she had had the same experiences. However in doing so he fails to recognize the fact that since the relation between the FIS' preferences and her own well being is a constitutive relation it cannot exist when the preferences are not present. Since the AS does not form similar preferences her well being cannot be similarly constituted. The situation is different in cases in which the information available to the FIS can be linked via a means-ends relation to the preferences of the AS. In these cases one can rely on the preferences of the AS to constitute her own well being, and the information available to the FIS only helps specify preferences that exist independently.

We can now return to the question with which we started this section. Could I have good reasons for thinking that my purported FIS is justified when she desires me to have a certain kind of ice-cream? The answer is that since my FIS can have no justification for wanting me to have vanilla over chocolate, I could have no reason for believing that she is justified in wanting me to have the preference for vanilla over chocolate.

The first limitation, then, is on preferences whose basis is in experiences that cannot be linked to the existing goals of the AS or with goals the AS should have. Only experiences that are relevant by way of means-ends reasoning to the goals of the AS (or the goals that the AS should have) could ground the justification of the advice of the FIS.

ii. Some experiences are not publicly verifiable

Sometimes even though the experiences of the FIS are linked via means-ends relations to the existing preferences of the AS (or to those preferences that the AS should have) it is nevertheless impossible for the FIS to justify his desires about the preferences of the AS. Imagine that I feel a great delight after tasting a certain kind of ice cream - call it delight of type A - and want to repeat that experience. Now imagine that my FIS experiences what he claims is a qualitatively similar delight – delight of type A - when he tastes a mango. We now have a case in which the experiences of the FIS seem to be linked in the appropriate way via means-ends relations to the existing preferences of the AS. According to the information available to him eating mango will give me delight of type A – which I already want.

Or consider another example. I am a liberal homosexual. My purported FIS is a redneck homophobe that can't even think about spending time with people like me. This is not for him a moral issue, rather a choice backed by deep emotional and desiderative support. Since he knows that my desires are different than his he does not simply want me to have the object of his desires. For himself, he wants a heterosexual marriage but he doesn't want me to want it. He knows that in my current state such marriage will not improve my well being. On the other hand he cannot think of me continuing to have what he considers to be my current 'disgusting homosexual desires'. He wants me at the very least to start initiating a change in my preferences. His preferences for my desires are communicated to me. Now assume that his desires about homosexuality are the result of experiences he had after trying for the first time a heterosexual relation. His heterosexual experience leads him to develop a new appreciation of heterosexuality and a feeling of disgust towards homosexual relations. Further imagine that he claims that the experiences he has are of a similar character to ones I had in the past. He claims to know that the delight he experiences now is one I experienced before and one I would very much like to repeat, and the disgust he experiences now when thinking about homosexuality is one he knows I have experienced before (in other circumstances) and one I very much want to avoid. In this case as well the experiences of the FIS seem to be appropriately connected via a means-ends inference to the existing desires of the AS.

But there is no way, even in principle, for me or for anyone I trust to verify that the experiences of my FIS have the same phenomenology as the experiences I liked in the

past. Therefore the relevant means-ends connection cannot be utilized to justify the FIS' recommendation to me, and I could not have a good reason for believing that he would desire me to change my life if he considered things from my point of view.

The full information theory under consideration tries to cash out intuitions about well being in terms of what a rational FIS would want me to want. But a rational FIS cannot want me to want what is unreasonable for me to want – and it is unreasonable for me to adjust my desires on the basis of information that is in principle unverifiable.

I will now consider a third limitation on the desires of my FIS concerning my desires.

iii. Changes in the preferences of the FIS

A desire based theory of well being is necessarily a dynamic theory of well being. What is good for me now may not be good for me tomorrow. One might think that the same is not true for the FIS. I think it is: what is good for the FIS now may not be good for her tomorrow; what she desires now may not be what she desires tomorrow. Two factors explain the possibility of a change of preferences of the FIS over time. First, as we have seen, the experiences of the FIS cannot normatively determine her preferences and therefore even if we assume that her experiences are complete, this fact does not fully determine what she should desire on the basis of these experiences. Second, even if we assume that given a fixed set of experiences one could not legitimately have two different set of preferences, the FIS could nevertheless not be held to one fixed set of preferences. This is because there is a kind of information that one can never be in full possession of; this is the kind of information that one gets as one experiences one additional moment of life.

So the FIS' desires for himself change overtime. But what does this tell us about his desires for his AS? There are two options to consider.

1. The FIS' preferences for the AS²⁶ change.
2. The FIS' preferences for the AS stay fixed.

²⁶ For simplicity sake let us assume that we consider the preferences of the FIS for his AS at a specific point of the AS' life.

If, or to the extent that, the FIS' preferences for the AS change we do not have a determinate account of the AS' well being. The determinacy constraints thus limits the kind of desires that the FIS may want the AS to have. The FIS cannot ground his wants concerning the AS' desires in his own changing desires, since if he did his wants for the desires of the AS would have to change as well.

If, on the other hand, the FIS' preferences for the AS stay fixed, they do not vary with his own preferences. In itself this is not a problem since we have been assuming all along that not everything that the FIS wants for himself he will want his AS to want. However once we realize that the preferences of the FIS for his AS cannot vary with his own preferences and cannot even be based on preferences of the FIS that may vary over time, we must revisit the basis for the FIS' advice to the AS.

There is only one type of preferences that the FIS could have for the AS that will not change as his own preferences for himself change. Preferences whose grounds are the combination of the actual preferences of the AS (or the preferences he should have) and publicly available information about the means for satisfying these preferences.

The third limitation thus points in the same direction as the previous two limitations. Some experiences – experiences that do not provide my FIS information that can be verified and experiences that cannot be linked via means ends reasoning to the goals of the AS (or the goals that the AS should have)– cannot be the grounds for desires of my FIS concerning the AS' desires.

Our investigation is not over though. We have considered four interpretations of Railton's view. We rejected the first three. The fourth interpretation **G** (the second variant of the disclosure advice model) has been partially promising. This interpretation successfully explains how my FIS could reasonably want me to have quite a few desires that would diverge from the desires that I actually have and whose objects are states of affairs that we naturally consider part of my non-moral good. We also saw however that on this interpretation there is a range of desires that my FIS could not reasonably want me to have. These are desires he would want me to have on the basis of experiences a. he could not link to desires I do have or should have or b. experiences that are not even in principle publicly accessible. What should we then think about the fourth interpretation?

The fourth interpretation offers us a coherent and determinate account of one's non-moral good. But as I will argue in the remaining section, the account of well being offered by the fourth interpretation fails to match our intuitions about one's non-moral good. As an analysis of our pre theoretical understanding of non-moral good it may need to be rejected.

3. A clash with intuitions

Our starting point was the realization that many of one's actual desires cannot determine what is good for her. We were looking for a theory that would deliver an amended set of desires hoping that this amended set would capture our intuitions about the notion of well being. We have investigated Railton's view, considered how best to interpret it, and discovered what states of affairs it deems part of one's well being. Call the set of the desires a person would have, had she adopted the desires her FIS wanted her to have under the fourth interpretation (G) of Railton's view S^* . What we want to know is whether the states of affairs desired by the desires in S^* constitute a person's well being? One way to go about answering that question is to ask whether S^* captures our pre-theoretical intuitions about well being.

Let us consider an example: I am in a mood for an action movie. I want to see the most recent Bond movie. My FIS saw the movie. He had a terrible time. I ended up seeing it too and had a terrible time. I got out and wished I saw a good drama instead. Indeed if the inner workings of my mind were slightly different and I would have wanted to see a drama, and I went to see one, I would have had a great time. But things were not different and I went and saw an action movie that as it turned out I didn't enjoy even one little bit. Not, mind you, because it was a bad action movie, as far as action movies go it was perfectly all right – rather because the experience of watching it made me – as I was watching it - want to watch something more sophisticated, more subtle, less violent. Going to see that action movie was not, we tend to think, part of the best life I could have had. Seeing a good drama would have been much better. My FIS has exactly the same thought. He says to himself 'the life my AS had was not the best life he could have had'. But could he reasonably have wanted me to want to see the drama.

Could my FIS reasonably believe me to have any good reason for thinking that he wants me to watch a drama and not an action movie – as he should believe in order to want me to watch the drama according to the fourth interpretation of Railton’s view?

The answer is no! There is no information that I could have gotten that would have made me believe that my FIS had the experiences that would justify the desires he claims to have had for my desires. I can believe that my FIS had changed his desires (from the desires we both had prior to him becoming fully informed to his current fully informed desires) but I could have no reason for thinking that he was rationally required to do so. The change in his desires is not irrational but is in no way rationally required or predictable given the information I could reasonably, even in principle, have access to and given what I knew or could have known about my own goals or the goals I should be having. Given my desire to see a good action movie, no information, short of information that revealed that the Bond movie was not an action movie or was not a good one – information that was not available since it was an action movie and a good one - could have required a change in his desires and therefore justified a change in my desires.

This example shows that the only coherent version of Railton’s view cannot explain our intuitions about well being. We intuitively think that my life would have gone better if I watched a drama and not the Bond movie. Railton’s view – on our fourth interpretation – cannot explain that result.

Consider another example. This is an inverted version of my previous example of a homophobic and homosexual self. I (my AS) am a redneck homophobe. I hate homosexuals, liberals, intellectuals and in general any person displaying too much sensitivity, too much intelligence or too little patriotism. My attitude is not a matter of belief (or not a matter of beliefs alone) and certainly not a matter of mere moral beliefs; it reflects my deepest emotional reactions and my most central desires. My FIS had the misfortune (so I see it) to be stuck on an island with a homosexual person for two whole weeks. The experience changed the life of that poor soul (my FIS). He now respects homosexuals and wants to move to San Francisco. He would have wanted me to have a complete change of desires. From his perspective my happiness is being held back by my prejudices, by my bigoted upbringing, by a distorted view of what is attractive, virtuous,

joyful, moving and worthy. I, so he thinks, do not see the beauty and joy in the multiplicity of human experiences and forms of being.

We (the readers of this story) and my FIS think that my well being would increase significantly were I to adopt the preferences of my FIS. But can Railton's view on our fourth interpretation deliver a similar conclusion? I think the answer is no. My FIS could not reasonably think that I have good reasons for believing that his new desires with respect to homosexuality are the desires of a reasonable and fully informed person with my constitution. The only beliefs I have grounds for forming are that he was bewitched, that the isolation on the island got to him, that he is confused, and that he is generally out of his mind. Of course, I know that others have similar desires to the ones now acquired by my FIS. But as far as I can tell their desires could not possibly justify a change in my desires: they are either lying about their experiences, or utterly different than me, or possessed by perverse passions in ways that hampers them from forming rational desires. In any case, I have no grounds for believing that my purported FIS is justified in wanting me to change my homophobic desires. My FIS could therefore not reasonably desire that I change my desires from what they currently are.

Conclusions

1. Full information views of well being analyze well being in terms of the desires of an FIS. To see whether any such analysis is acceptable or not one has to carefully examine the exact content of the desires that the FIS is expected to have. In particular one has to ensure that the desires the view attributes to the FIS are reasonable. If these are desires about the desires that the actual self should adopt, the reasonableness of the desires of the FIS must be tested bearing in mind what desires it would be reasonable for the actual self to adopt while maintaining her autonomy. Many seemingly attractive versions of full information views will be found unsatisfactory in light of such examination.
2. According to the best interpretation of Railton's view well being is determined by the desires a fully informed person could justifiably want his actual self to have, given the standards mentioned in the first conclusion above. When we consider examples we see that there is a gap between what we intuitively think is good for

one and what one's FIS could justifiably desire for one. We are therefore bound to decide whether we want to reject the analysis of well being in terms of fully informed desires or whether we want to question our intuitions²⁷.

3. If we choose to reject our current intuitions concerning well being and to revise our conception of well being in light of the fourth interpretation of Railton's view offered above we impose the following limitation on the concept of well being: what is good for us must be determined either by our current desires or by the set of desires we have good reason to come to have. What desires we have good reason to come to have will be determined by the combination of our current stable desires and policies, and information that is, at least in principle, publicly available.²⁸

²⁷ I take it that Railton's view is the best developed analysis of well being in terms of fully informed desires currently on the market. I therefore tentatively suggest that our investigation teaches us something general about the relation between well being and what could reasonably be desired by a fully informed agent.

²⁸ The question of how should an agents change intrinsic desires in light of pressure from other intrinsic desire and new information is an extremely complicated question and deserves a discussion of its own. My formulation in the text is not intended to support any specific view about norms for changing intrinsic desires. It is merely intended to indicate that some such changes could be reasonably demanded.

For examples of cases in which changes in intrinsic desires would be justified see my "Desiring at Will and Humeanism in Practical Reason", *Philosophical Studies* 119 (3):265-294.