Alienation, Engagement, and Welfare

James Fanciullo

Lingnan University

Forthcoming in the *Philosophical Quarterly*

Abstract: The alienation constraint on theories of well-being has been influentially expressed thus: 'what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive …. It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him' (Railton 1986: 9). Many agree this claim expresses something true, but there is little consensus on how exactly the constraint is to be understood. Here, I clarify the sense in which the quote offers a basic constraint on theories of well-being—a constraint that should be adopted by (e.g.) hedonists, desire satisfactionists, and objective list theorists alike. This constraint focuses on *affective engagement*, or positive affective stances in connection with a proposed good. I show that the constraint explains a near-universal intuition, and rules out a number of well-known theories of well-being.

Theories of well-being or welfare ultimately aim to tell us which things are basically good and bad for us—or, which things are the basic constituents of our well-being and ill-being. Generally speaking, attempts to explain the nature of well-being have fallen into three main camps: hedonist, or those taking well-being to consist in enjoyment or pleasure; desire satisfactionist, or those taking well-being to consist (somehow or other) in the satisfaction of desire; and objective list, or those taking well-being to consist in the possession of some set of 'objective goods' (e.g. friendship, knowledge, etc.). Despite the initial plausibility of each of these approaches, though, there of course remains little consensus over which particular theory of well-being we should accept.

Still, theorists of each of these three stripes (and others) seem uniformly motivated to satisfy a certain constraint on theories of well-being. While the exact content of this constraint is somewhat elusive, it has been convincingly and influentially expressed thus:

> 'what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him' (Railton 1986: 9)

To see some of the basic motivation for this, consider an individual whose job, or family, or religion leaves them cold. In fact, suppose that their job feels nothing but tedious and meaningless, their family is nothing but manipulative and toxic, and their religion only shames and demeans them. In this case, it would seem inappropriate to insist that their work, their family, or their religion was nevertheless intrinsically good for them. Indeed, to insist on this would be to in some sense *alienate* the subject from what is basically good for them, or from their personal good.

But in what sense, exactly, would theories of well-being 'alienate' subjects by making such insistences? To be sure, there seems to be something wrong with insisting, against clear evidence to the contrary, that something that is typically good for most people (such as family) is intrinsically good for even those individuals that seem to be exceptions to the general rule. Moreover, there seems to be something wrong with simply ignoring a given individual's experiences, and attitudes, in relation to a putative good. But there is nevertheless no clear understanding of what, exactly, goes wrong when we insist on a thing's intrinsic goodness for a person, such that this insistence amounts to alienating the subject from their own good.

This is a bit odd, I think, both because Railton's quote is so often cited in debates about well-being—I've found more examples than I can reasonably list here—specifically to avoid 'alienating' theories, and because theorists in this debate seem to think his quote expresses a constraint on these theories that is in some way *fundamental* or perhaps even *clearly* true.[1] Indeed, there does seem to be something true expressed in Railton's quote. But what exactly this something is seems elusive. Thus, getting a better grip on how Railton's quote expresses a basic constraint on theories of well-being may prove helpful in guiding our evaluation of the theories.

My aim in this paper is to do just that. In particular, I want to explore the discussion surrounding Railton's quote, and determine the precise sense in which it expresses a true constraint on theories of welfare. While these aims may, admittedly, seem relatively modest, I should note there will be some significant upshots from my discussion. Indeed, as I'll show, when the constraint is properly conceived, it rules out a number of well-known and seemingly plausible theories of well-being. It also constitutes a basic adequacy condition for any proposed theory of well-being, and so establishes a starting point for theorizing going forward. And besides, offering a plausible explanation of a near-universal intuition seems worthwhile in its own right.

Here, then, is how I'll proceed. In the following section, I'll address the leading way of construing Railton's quote in the literature, and suggest that this approach cannot capture what very many of us find plausible about his alluring thought. Then, in section 2, I'll offer my positive proposal. As I'll show, the allure of Railton's quote is best explained by appeal to a phenomenon I'll call *affective engagement*, or positive affective stances in connection with a given putative good. Thus, on the resulting constraint, what makes a theory of well-being unacceptably alienating is its implying that something may be basically good for a subject even if the subject displays no affective

[1] As Dale Dorsey puts it: 'References to this thesis are—or may as well be—countless' (Dorsey 2017: 685). Here are just six relevant such examples: Dorsey (2017), Fletcher (2013), Fletcher (2016), Heathwood (2019), Rosati (1996), and Sarch (2011).

engagement in relation to the thing. Finally, in section 3, I'll illustrate the constraint's promise by exploring its implications for the leading three approaches to welfare.

1. The resonance constraint

To get a grip on the popular way of developing Railton's alluring thought, let's return to our simple example of the subject whose job, religion, and family leave them cold. If a theory of well-being nevertheless insists that these things are intrinsically good for the subject, it will seem implausibly alienating. On the leading approach, what makes the theory implausibly alienating has to do with a failed connection of a very direct kind between our subject and the putative goods. In particular, it is the subject's lack of *valuing attitudes* (of the right sort) towards those very putative goods that makes the insistence upon their goodness for the subject problematically alienating.[2] It is only when the things are objects of the subject's relevant valuing attitudes that the things 'resonate with' or 'fit' the subject, and so it is only then that our claim of the thing's goodness for the subject cannot count as alienating.

This approach aligns closely with the first part of Railton's thought: that 'what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware' (Railton 1986: 9). Clearly, the thought is, we cannot insist that something is good for a subject if the subject lacks any positive stance toward the thing—that is, at least, so long as this is not attributable to the subject's irrationality or ignorance. Somewhat more precisely, the subject must have valuing attitudes of the right sort (where these may be hypothetical, or actual, or inconsistent with irrationality, or whatever) toward the thing, in order for it to be intrinsically good for them. Or, as Dale Dorsey puts it, according to the

[2] See especially Dorsey (2017), Rosati (1996), and Sarch (2011).

> *Resonance Constraint*: an object, event, state of affairs, etc., φ is good for an agent x only if x takes a valuing attitude (of the right sort) towards φ. (Dorsey 2017: 687)[3]

It is thus a necessary—though not, of course, sufficient—condition of a thing's being (intrinsically) good for a subject that the subject takes (or would take) such a valuing attitude towards it. That, at least, is the idea behind this way of developing Railton's thought.

Plausibly, the Resonance Constraint will prevent cases of 'alienation' of the kind with which we began. Suppose then Ann's job feels nothing but tedious and meaningless, her family is nothing but manipulative and toxic, and her religion does nothing but shame and demean her. So long as this description implies that Ann lacks the relevant kind of valuing attitude toward these things, or that they do not 'fit' her evaluative profile, the Resonance Constraint will imply that her job, family, and religion cannot be good for her. The constraint will moreover offer an explanation as to why theories insisting that these things might nevertheless be good for her are intolerably alienating: they run afoul of the constraint, and so ignore the subject's evaluative profile regarding the things in determining what is good for them. This, at least, is the type of explanation the proponent of the constraint would presumably hope to offer.

But the Resonance Constraint is, if (intentionally) imprecise, also quite strong. It rules out from contributing to one's well-being anything that doesn't 'resonate with' or 'fit' one's evaluative profile. This means that, on many reasonable precisifications of the constraint, a subject who does not (or would not) take a valuing stance toward pleasure, or the satisfaction of desire, or friendship, or autonomy, etc., would show that these things cannot be basically good for all welfare subjects. In other words, such a subject would show that (many leading forms of) hedonism, desire

[3] Dorsey simply calls it 'The Constraint', though I add 'Resonance' to differentiate between this proposal and the one I'll offer below. I should also note that, while Dorsey goes on to offer likely the most plausible justification for the Resonance Constraint available, the constraint nevertheless retains the controversial implications I'll discuss below.

satisfactionism, and the objective list approach were false.[4] They would be false because, for as many subjects whose lives seem to be made better by pleasure, or desire satisfaction, etc., there is one whose evaluative profile these putative goods do not 'fit'. Since they do not 'resonate' with this subject, they cannot be intrinsically good for them—nor, contra these popular theories, can they be good for every subject who possesses them. In general, it seems we could for any theory find some relevantly quirky agent whose evaluative profile does not 'fit' the theory's proposed good—including, perhaps, the proposed good of valuing itself!—and who therefore disproves the theory.

Of course, you may at this point think: 'well *so what* if the Resonance Constraint disproves all these popular theories? So much the worse for them!' But the problem here really lies in the fact that these theories, as evidenced by their great enduringness, have a good deal of initial plausibility. As I mentioned, a subject who did not have the relevant valuing attitude toward (say) enjoyment may, on the Resonance Constraint, disprove hedonism. But this is not the whole story. Indeed, to further assess this, we can imagine this subject's life is chock-full of nothing but enjoyment. They are constantly and endlessly enjoying themselves—eating, dancing, reading, playing, etc.—but, as it happens, they never have the relevant valuing attitude toward these experiences. In this case, does it seem more plausible that the subject's life nevertheless contained at least *some* well-being, or that their lifetime level of well-being was zero? Does it seem more plausible, that is, that these very enjoyable experiences made the subject at least *somewhat* better off, or that, as the Resonance Constraint implies, the subject's life was no better than one spent doing nothing but emotionlessly staring at a wall? As I see it—and I suspect others will agree—it seems clearly more plausible to think the subject's life went at least somewhat better in virtue of these experiences, despite the

[4] The type of desire satisfactionism I have in mind here is the one on which the basic goods are complex states where one desires something and the thing obtains (or one believes it does). This is to be contrasted with the type on which the basic goods are the very things that one happens to desire and that obtain (or one believes obtain). Admittedly, if desiring is a valuing attitude of the relevant kind, the Resonance Constraint wouldn't rule out the latter kind of theory in the way I describe here. Still, this conception of valuing is controversial, and will arguably face opposition from desire satisfactionists and proponents of the Resonance Constraint alike.

subject's not happening to value them. This suggests that, for whatever amount of plausibility we attribute to the Resonance Constraint, it seems it will be defeated by the plausibility of a leading theory of well-being. And here, of course, we could easily replace enjoyment with desire satisfaction, or objective goods. Thus if my point here about enjoyment making the subject's life at least somewhat better for them does not move you, you can replace enjoyment with the purported goods of many leading theories of well-being. In each case, I think we'll find that the plausibility of many of those goods contributing at least *somewhat* to the subject's well-being ultimately seems more plausible than the subject's life being no better than one spent emotionlessly staring at a wall. So, it will not be sufficient for the proponent of the Resonance Constraint to simply endorse the fact that the constraint rules out many leading theories—they will also have to convince us that many of our basic, pre-theoretic intuitions about well-being (as evidenced by the enduring theories that represent them) are misguided.

Now, whether the constraint will indeed have any of these implications of course depends on the more precise definition its proponent offers. As it stands, the constraint doesn't necessarily rule these leading theories of well-being out. But, even so, it remains highly controversial. And another, more straightforward reason for this has to do with its implications for non-valuing welfare subjects, or subjects who are capable of welfare yet incapable of forming the relevant kind of valuing attitude.[5] Human infants, cats, and dogs, for instance, don't seem capable of forming any attitude we would plausibly call 'valuing'. Things cannot 'resonate' with these beings in the way the Resonance Constraint seems to require. In that case, the constraint seems to rule out that anything is good for these subjects. Yet that implication seems obviously wrong: surely an infant's happiness or health may be good for them, even if they cannot (and could not) form a valuing attitude toward these things. If that's right, it seems to suggest either that the Resonance Constraint is false, or that non-

[5] See Dorsey (2017) and Lin (2017).

valuing welfare subjects require a separate theory of welfare altogether. But do our pre-theoretic ideas of goodness really seem to require two separate theories for different welfare subjects? Do we seem to mean something radically different when we say an infant's happiness and health are good for them, and when we say our own happiness and health are good for us? It seems clear to me that we do not. Or, at least, it seems clear we have more reason to reject this than we have to accept the Resonance Constraint.[6] If that's right, it again suggests the Resonance Constraint faces a significant hurdle.

Perhaps some properly formulated version of the Resonance Constraint can meet these challenges. Frankly, I have my doubts. Either way, though, we have some strong reasons for doubting that the constraint comes close to capturing what very many of us find plausible about Railton's alluring thought. When it is more fully thought through, it seems to have implications that few of us, at least pre-theoretically, would want to accept. And this casts doubt on the constraint's independent plausibility. Ultimately, the Resonance Constraint seems to focus both on too sophisticated an attitude to require of all instances of goodness for a subject, and too direct a connection between that attitude and a putative good. The former issue leads to problems regarding non-valuing welfare subjects, and the latter issue leads to problems regarding implausibly ruling out very many leading theories of well-being (since ruling them out requires just one quirky subject who lacks the relevant attitude toward a theory's proposed good). Addressing both of these issues in rethinking what we should take away from Railton's thought may lead to the constraint we're actually after.

---

[6] See especially Lin (2017) and Lin (2018). Admittedly, some theorists disagree: they offer views on which the same theory of welfare is not true of adult humans and infants. See e.g. Bruckner (2021) and Yelle (2016). Notably, though, these views are all sympathetic to something like the Resonance Constraint. So, while those who are already committed to something like this constraint may have reason to posit multiple theories of welfare, those of us without this commitment will presumably see little reason (and indeed, via Lin, see reason not) to do so. In the context of this paper, in any case, since Railton's thought remains attractive to theorists on both sides of this debate—this debate clearly does not mark the divide between those who accept Railton's thought and those who reject it—we, again, at least have reason to doubt that the Resonance Constraint captures what ultimately makes his thought plausible.

2. The engagement constraint

Despite its initial plausibility, the Resonance Constraint seems to have implications that few of us, at least pre-theoretically, would want to accept. Still, there seems to be something fundamentally true expressed in the thought, from Railton, that motivates it. Recall that this thought calls for 'a connection' between what is intrinsically good for a subject and what the subject finds 'in some degree compelling or attractive' (Railton 1986: 9). It then claims that what is intrinsically good for a subject must in some such way 'engage' them, and that theories running afoul of this constraint are 'intolerably alienating'. Clearly, there is a great deal of room between these claims and the Resonance Constraint. Indeed, for as strong as the Resonance Constraint may be, the thought that motivates it hardly commits one to much at all: just 'a connection' between what 'engages' a subject, or what the subject finds 'compelling or attractive', and what is intrinsically good for the subject. Given the issues surrounding the Resonance Constraint, then, what is the most plausible way of developing this thought so as to prevent intolerably alienating theories of welfare?

As I see it, two crucial steps leading from Railton's thought to the Resonance Constraint are responsible for the constraint's problems. The first is the significant step from talk of 'engagement' and what one finds 'compelling or attractive', to a focus on presumably more substantive valuing attitudes. Plausibly, any subject who is capable of welfare is also capable of being engaged by something, at least in the sense of finding the thing to some degree compelling or attractive. This includes welfare subjects such as human infants, and non-human animals like cats and dogs. What is far less plausible, we've seen, is that any subject who is capable of welfare is also capable of valuing something—and this is precisely where the Resonance Constraint runs into problems. To demand a connection of some sort between a subject's good and their finding things compelling or attractive at least does not force us to restrict our pre-theoretical idea of which subjects are capable of welfare;

to demand this between a subject's good and their valuing things, on the other hand, does force such a restriction. So, it seems the most plausible version of our desired constraint will focus on engagement or attraction, rather than (presumably more substantive) valuing attitudes.

The second step I have in mind is the one from talk of 'a connection' between a subject's good and their relevant attitudes, to conceiving of this connection as a very direct one, between the subject's attitudes and the specific things that are meant to be intrinsically good for them. Recall that the Resonance Constraint claims a thing is intrinsically good for one only if one has the relevant valuing attitude toward the thing. Presumably, this tight connection between good and attitude helps in preventing cases of 'alienation', as it ensures the thing that is said to be good for a subject is also directly endorsed or valued by the subject. This, however, doesn't seem to capture the truly 'intolerable' type of alienation with which our constraint should be concerned. To see this, imagine a quirky subject who doesn't value (say) enjoyment, or desire satisfaction, despite having a life that is chock-full of these putative goods. As we've seen, such a subject would (plausibly), on the Resonance Constraint, disprove theories like hedonism and desire satisfactionism. But would a hedonist or desire satisfactionist's response seem 'intolerably alienating', if they were to simply insist that the subject's experiences of enjoyment or desire satisfaction were nevertheless good for them? Would the truth of their theory seem to fly in the face of our pre-theoretical thoughts about a person's good, in virtue of this subject's quirky attitudes? Or does it instead seem entirely possible that a subject like this may, sadly, simply have attitudes that are at odds with what turns out to be basically good for them? As I see it, an insistence like this would not betray one's commitment to an 'intolerably alienating' theory of well-being. To see why, compare this case to Ann's, where her job feels nothing but tedious and meaningless, her family is nothing but manipulative and toxic, and her religion only shames and demeans her. A theory insisting that these things were nevertheless good for her would, at least as I see it, seem truly *intolerably* alienating. And it would seem this way

because, at least as I imagine the case, Ann is in no way attracted to or engaged by anything associated with the things. That is, it's not just that she lacks some positive stance toward the specific things ('work', 'family', 'religion') themselves, but that she lacks any positive stance in connection with the things at all. If a theory were to then insist that, despite Ann being a 'quirky' example, the things are nevertheless good for her, we would dismiss it out of hand. (Indeed, even describing her as 'quirky' would seem totally inapt, or to ignore that she lacks something of fundamental importance.) That, we would say, is truly *intolerably* alienating, as it *entirely* detaches the subject's good from their positive responses, or from what attracts or engages them to any degree. The hedonist and desire satisfactionist, on the other hand, at least seem to insist that what is good for the subject shares some necessary connection with what attracts or engages them—just not a connection, specifically, to the direct objects of the subject's particular values. As I see it, then, the constraint we're ultimately after will not demand that the subject is relevantly attracted to or engaged by the specific putative goods themselves. Instead, it will demand just that the subject displays relevant attraction or engagement somehow in connection with the putative goods. This, I think, is just what it is to demand that a theory not be '*intolerably alienating*'.

This may be slightly unclear, so an example might help. Suppose a relevantly quirky subject doesn't have the relevant valuing attitude toward (say) friendship. They simply do not value friendship and so, on the Resonance Constraint, friendship cannot be intrinsically good for them. Now, as we've seen, it may not seem *intolerably* alienating to insist that friendship is nevertheless good for this subject. After all, we can suppose that the subject in fact has experienced a great deal of joy, camaraderie, and affection as a result of friendships they've had. If these features of friendship, and indeed friendship in general, are nevertheless not things that the subject values, we will not seem *intolerably* obstinate, it seems to me, in insisting that the friendships may still be good for them. At the very least, we will not here entirely estrange the subject's good from what attracts

or engages them; it's instead just that, as it happens, the subject fails to value something that entails their being attracted or engaged. In support of this, we can, again, contrast this case with Ann's. In her case, it *would* seem intolerably alienating to insist that her job, family, and religion were all good for her. And this, it seems, is precisely because she displays no attraction or engagement in connection with these things at all. To insist that these things were nevertheless good for her *would* be to entirely estrange her good from what attracts or engages her, and this, it seems, would be truly intolerably alienating. In general, then, what we want the constraint to prevent is not cases where something is intrinsically good for a subject despite the subject failing to have the relevant valuing attitude directly toward it, but rather, we want it to prevent cases where something is intrinsically good for a subject despite the subject failing to display any attraction or engagement in connection with it at all. This is how we can rule out theories insisting that things like Ann's family or religion can be intrinsically good for her, without ruling out theories insisting that, as it happens, it's possible for a subject not to value something (like friendship) that is intrinsically good for them. That is, it's how we can rule out truly *intolerably* alienating theories—or those *entirely* detaching a subject's positive responses or attractions from their good—without ruling out theories that, if they are alienating at all, are tolerably so—or those allowing a subject's good to come apart from the direct objects of their particular values.

Note that this is not to say that our constraint will rule out theories that do not allow this, or those that remain committed to the Resonance Constraint. After all, to require that a subject display attraction or engagement in connection with a thing in order for it to be good for them may be consistent with requiring a subject to have relevant valuing attitudes directly toward the thing in order for it to be good for them. As we've seen, though, we have reason to think that the latter requirement is too strong—or, at least, that it implausibly rules out too many contending theories, and faces too many problems, for it to capture what very many of us find plausible about Railton's

alluring thought. What we need, then, is a formulation with a different scope: both in terms of which responses we require of subjects, and in terms of the connection we require between these attitudes and the putative goods. Let's take these points in turn.

First, recall the types of attitudes Railton calls for theories of well-being to require in order for things to be intrinsically good for subjects: the subject must be in some degree compelled, attracted, or engaged. Clearly, one can be compelled, attracted, or engaged without (substantively) valuing anything: think of enjoying reality tv, yearning for just one more cigarette, or even, more generally, being attracted to things like food or sensations in the way beings like infants, cats, and dogs often are as well. In these cases, we seem to have *positive psychological responses* in connection with things, without these responses rising to the level of valuing. But which psychological responses, in particular, are we picking out when we demand that a subject be compelled, attracted, or engaged?

A first, natural suggestion might be: 'pro-', *conative* attitudes like desire and preference. These are the psychologically familiar attitudes that, perhaps among other things, essentially *motivate*: they involve a tendency to bring about or promote their objects. There are a few reasons this proposal may seem attractive. For one, on some views, valuing is also a form of conative attitude. And, if that's right, then this proposal may explain why a focus on valuing attitudes alone was able to explain perhaps some, but not all, cases of a subject's seeming attracted or engaged. After all, to explain all such cases, on this proposal, we need other conative attitudes like desire and preference too. For another, conative attitudes are, as I say, closely linked to motivation. And, you might wonder: is there any time we seem more clearly compelled or attracted, than when we are motivated to bring something about? Indeed, experience suggests that perhaps the most paradigmatic examples of being compelled, attracted, or engaged are just cases where we're motivated. Take, for example, cases where we *want*—to eat when we're hungry, to see our partners after a long time apart, to be offered a job, and so on. In light of all this, it might seem that the positive psychic responses we're

after—or those capturing the idea of a subject's being compelled, attracted, or engaged—are just those in the domain of motivation, or those made up of our positive conative attitudes.

As I see it, this categorization will not do. Whereas the focus on valuing attitudes seemed to count too few subjects as being attracted or engaged, a focus on conative attitudes more generally would seem to me to count too many subjects as attracted or engaged. The problem relates to motivation. In particular, notice that these attitudes 'motivate' in the following sense: they involve a tendency to bring about or promote their objects. Such motivation, though, seems to be involved in all intentional action. Whenever we intentionally pursue a goal, in other words, it seems we must have been motivated—we must have been disposed to bring about or promote some end. In that case, conative attitudes are involved in just any intentional action we perform. And, in that case, the current proposal suggests that attraction or engagement is involved in just any intentional action we perform. This, though, seems clearly wrong: not all intentional, goal-directed behavior involves a subject's being compelled, attracted, or engaged—at least, not in the sense that Railton and we have in mind.[7] Cases illustrating this point are well known, and often focus in particular on the conative attitude of desire. Consider, for instance, Warren Quinn's *Radio Man*, who is 'in a strange functional state that disposes [him] to turn on radios that [he sees] to be turned off' (Quinn 1993: 32). This state, which is said to be a desire, is meant to be *merely* functional, in that it is no more than a bare disposition to behave. There is no further purpose for which he turns on the radios, and indeed nothing favorable he sees in anything relating to turning them on (not hearing music, flipping switches, or anything else). He is in a motivational state that is, in effect, stipulated *not* to involve any attraction or engagement. The case is thus precisely one where motivation and engagement seem to come apart. And while there may, of course, still be one sense in which turning on radios 'attracts' or 'engages' this person—perhaps similar to the sense in which electrons are 'attracted' to protons—it

[7] See Heathwood (2019) and indeed Railton (2012).

seems clear enough to me that these are not the senses of 'attracted' or 'engaged' that we are concerned with when thinking of someone's good.[8] This is only further illustrated by the case of Ann.

Recall one of the features of Ann's case: her work feels nothing but tedious and meaningless. Suppose this includes going to meetings (or filling in reports, etc.) that are boring, repetitive, and pointless. As long as Ann goes to the meeting, she must have been motivated to do so. And as long as Ann was motivated to go to the meeting, she must have in one sense been 'attracted' to going. Yet if this is the sense of 'attracted' that we are concerned with when thinking of someone's good, then it seems that Ann's work is not in fact something we should be concerned to rule out as good for her. Moreover, it seems theories insisting that her work is good for her are not ones we should be concerned to rule out as implausibly alienating. After all, she performs some action in working, and so must in this sense be 'attracted' to (some aspect of) working. We are supposing, though, that the meetings (and all other aspects of her work) are nothing but boring, repetitive, and pointless—there is nothing positive she sees in them, no mode of presentation under which they are appealing or interesting, nothing about them she regards in a positive way. In short, she is merely *behaving*. And while mere behavior may be enough for *motivation*, it does not seem to be enough for proper *attraction*.

I take all of this to suggest that motivation is not the core of the kind of attraction and engagement we are concerned with when thinking of someone's good. When Radio Man turns on a radio, he may thereby perform a motivated act, but he is not thereby made better off. This accords with the fact that he finds nothing compelling or attractive about turning them on. It also accords with the fact that, even though he displays motivation, he is not protected from intolerable alienation: his act would remain motivated, after all, even if he loathed everything about turning on

[8] I borrow this analogy from Heathwood (2019: 683).

radios. How, then, should we categorize the kind of attraction or engagement we are concerned to ensure here?

As I see it, the cases we've considered thus far suggest an answer. A focus on conation seems attractive precisely in those cases where the motivating attitude is *more* than just a behavioral disposition. Recall a few such cases: yearning for a cigarette, wanting to see one's partner, desiring to be offered a job. In each of these cases, there is not just a disposition to perform some act, but also an implicit *psychological feeling*—an affectively valenced representation of the object.[9] The same is true when we enjoy reality tv, or think of delicious food when we're hungry. Importantly, these psychological feelings are also precisely what Ann and Radio Man lack: they are disposed to act, but lack any feeling, emotion, or affective investment. It's not just that they fail to be overcome by *intense* passion or emotion, but that there is no psychic feeling or engagement with what they do at all: nothing appealing they see in what they are doing, no degree of enthusiasm, excitement, or gusto. More generally, then, the cases where conative attitudes indeed seem to involve engagement or attraction on the part of their subjects are those where the attitudes come with a positively valenced psychological feeling, or a positive affective representation of the attitude's object. As I'll put it, these are cases where the subjects display *affective engagement*.

Of course, it's not just conative attitudes that can involve such engagement. Consider some of our other simple examples: enjoying reality tv, liking pizza, or experiencing sensory pleasure. These cases, too, all involve some positively valenced affective response, or some degree of phenomenologically salient affective appeal. In categorizing these responses, then, I mean to be quite inclusive: they may include desires, likes, pleasures, enjoyments, certain emotions or moods, or other positively valenced affective states.[10] As long as the responses can plausibly be construed as

[9] See Heathwood (2019) and Railton (2012).
[10] See e.g. Haybron (2008) for some of the emotions or moods that may be included here. While Haybron also emphasizes the importance of affect in theorizing about a subject's good, the kinds of affective responses his theory

involved in a subject's being compelled, attracted, or engaged—in the genuinely favoring, positively regarding, phenomenologically salient, *affective* sense I've outlined here—we can include them in this characterization. Ultimately, these responses are unified by their involving some degree of affective interest (however slight). Thus, again as I'll put it, when one has a positively valenced response of this sort—or one is properly compelled, attracted, or engaged—one displays *affective engagement*.

As I see it, it is this phenomenon that the constraint we're looking for should focus on. This is because, unlike valuing or conation, affective engagement seems to capture just the type of phenomenon that Railton was (and we are) particularly concerned to ensure when theorizing about a subject's good.[11] It focuses on the very terms—'compelling', 'attractive', and 'engage'—that seem to make his thought particularly alluring, and generalizes to include all related parts of one's affective life. It thus captures the forms of attraction and engagement displayed by happy and healthy infants, dogs, cats, and other non-valuing welfare subjects, whose responses seem to ensure they are properly connected to their good. And it excludes the bare dispositions to act displayed by Radio Man and Ann, whose acts seem insufficient for ensuring they are properly attracted, engaged, or connected to their good. Our aim of developing Railton's thought, then, seems to lead us to affective engagement. This seems like the phenomenon we want to require.

Second, we must address the type of connection we want to ensure between affective engagement and the putative goods, on the constraint, in order for the goods to count as intrinsically good for one. As I see it—and as some of my earlier claims have alluded to—we should again be quite inclusive in characterizing this connection. In particular, we should avoid demanding that

---

requires are considerably more limited than those captured by my notion of affective engagement, which includes many additional (non-'emotional nature-fulfilling') desires, pleasures, enjoyments, etc.

[11] Indeed, this is supported by Railton's more recent arguments regarding normativity and desire. Railton suggests that desire's (putative) role in various normative domains can be explained provided we adopt a particular conception of desire, according to which desire is (in part), 'underneath, a compound rather than simple state, with two distinctive aspects: a degree of *positive affective attraction* and a degree of *focused appetitive striving*' (Railton 2012: 31). Thus Railton, too, thinks an affective component is necessary.

subjects affectively engage directly with a thing in order for it to be intrinsically good for them. And this is because, it seems, many theories of welfare imply that things can be intrinsically good for a subject even if the subject does not affectively engage directly with the thing, and yet, importantly, the theories accord with the spirit of our desired constraint. To see this, take hedonism. If enjoyment is the sole basic good, and our constraint requires direct affective engagement with a thing in order for it to be intrinsically good for a subject, it seems our constraint requires a subject to affectively engage with enjoyment itself in order for it to be intrinsically good for them. This, though, would be a weird thing to demand. Clearly, what matters isn't that the subject affectively engages with enjoyment, or the thing that's said to be intrinsically good; rather, what matters is that enjoyment, or the thing that's said to be intrinsically good, itself *involves* affective engagement. Similarly, it would be silly to insist that desire satisfaction, or friendship, etc., is good for one only if one affectively engages with the satisfaction of desire, or friendship, *itself*. Instead, what we want our constraint to demand is again just that desire satisfaction, or friendship, itself *involves* affective engagement (more on this below). So, we should formulate our constraint in terms of a putative intrinsic good's involving or necessitating affective engagement, rather than in terms of a subject's being affectively engaged directly with the good.

On that note, we're now in position to offer a more precise formulation of our constraint. On what we may call the

> *Engagement Constraint*: a thing φ is intrinsically good for an agent x only if x's having φ involves x's displaying affective engagement (or, involves x's possession of positive affective stances in connection with φ).

This constraint captures both the type of responses, and the connection between these responses and a theory's proposed goods, that we want to require of subjects in theorizing about their good. It implies of Ann, for instance, that since she lacks any positive affective responses at all in connection with her job, family, and religion, these things cannot be intrinsically good for her. It thus protects Ann from theories insisting that while her job, family, or religion are in no way sources of affective engagement, they are nevertheless good for her. These theories, we've seen, would be truly *intolerably* alienating, as they would *entirely* separate what attracts or engages a subject from their good. Still, the constraint stops short of implying that those lacking certain valuing attitudes toward just anything can show that the thing is not intrinsically good for them. Indeed, so long as possession of the thing involves affective engagement on the part of the subject, the subject's lack of valuing attitude toward the thing doesn't rule it out as being intrinsically good for them. Thus, proposed goods like enjoyment, desire satisfaction, friendship, and the like, will not necessarily be ruled out by the Engagement Constraint, and so we get the restriction we were after without implausibly ruling out any leading approaches to welfare in their entirety. It seems, then, that the Engagement Constraint protects against just the type of alienation that those moved by Railton's thought are most concerned to preclude.

This is not the only issue for the Resonance Constraint that the Engagement Constraint avoids. Specifically, the Engagement Constraint also doesn't rule subjects who are incapable of valuing out from being welfare subjects. Since this constraint requires only affective engagement in connection with a thing in order for it to be good for one, it implies only that subjects who are incapable of affective engagement are incapable of welfare. And, of course, the capacity for affective engagement is not unique to valuers, some proper subset of human beings, or even human beings in general: it is a capacity we share with infants, cats, dogs, and many other beings. In fact, it is plausibly a capacity we share with all welfare subjects—if something is incapable of affective

engagement, in the broad sense we've specified here, it is hard to see how things could be good or bad for the thing. Without any affective life, after all, it is unclear how something could even be a welfare subject: the thing would be incapable of finding anything attractive or engaging (or aversive or unpleasant) to any degree at all.[12] While some may, admittedly, find even this controversial, it is at least far more plausible than the restriction on welfare subjects implied by the Resonance Constraint. With no affective life whatever—no capacity for finding things compelling or engaging to any degree—nothing can be good for one. As I see it, this implication is, far from hard to accept, in fact hard to resist.[13]

While I think all of this suffices to show that the Engagement Constraint offers the best expression of Railton's alluring thought, and so the best explanation of a near-universal intuition, whether we should ultimately *accept* the constraint is a separate matter. Needless to say, I cannot fully explore or defend all of the constraint's implications here. At most, I can offer a tentative case for accepting the constraint, by illustrating some of its promise. To do this, then, it may be helpful to focus on an implication I briefly mentioned earlier. I said that one attractive feature of the Engagement Constraint—and a feature that separates it from the Resonance Constraint—is its remaining compatible with each of the three leading approaches to well-being. Really, though, its implications for these theories are not so straightforward. Exploring and defending these implications is the topic of the next section.

---

[12] At the very least, it seems clear to me that things could not be good for such a thing in the same way they are good for us. Thus even if it is insisted that some beings without affective capacities are indeed welfare subjects—say, plants, or beings that are capable only of motivational states like Radio Man's—it seems plausible these affective capacities mark a divide between two radically different kinds of 'welfare subject'. After all, as we've seen, mere functioning or behaving seems insufficient for proper engagement, yet proper engagement is precisely what we, following Railton, have identified as key to our conception of a subject's good. Thus unlike in the case of valuing, it seems to me, we do mean something radically different when we say things are good for subjects, like plants, that lack affective capacities, and when we say things are good for subjects, like us, who have them.

[13] Cf. Chalmers on 'philosophical Vulcans', which lack affective capacities but may nevertheless have 'serious intellectual and moral goals' (Chalmers 2022: 343-4). Chalmers appeals to these goals in arguing that such beings would have, specifically, moral status. Again, though, as soon as we recognize that their 'serious goals' would be no different from Radio Man's goal of turning on radios, the thought that they would be proper welfare subjects, at least, seems far less plausible—see fn. 12.

3. Some implications

Admittedly, I cannot explore or defend all of the Engagement Constraint's implications, even for the leading three approaches to welfare alone, in their entirety here. Still, it will be helpful to see some of the more important implications for these approaches, even if they must be presented in broad strokes.

Consider first then hedonism. As we've seen, hedonist theories—taking welfare to consist in enjoyment or pleasure—all seem straightforwardly compatible with the Engagement Constraint. This is because these theories all view well-being as a certain kind of affective state or response. So, the constraint's implications for this leading approach to welfare, at least, seem particularly straightforward.

Its implications for the other two leading approaches, however, are not so simple. Consider next desire satisfactionism. As we've seen, desire satisfactionist views take well-being consist, somehow or other, in the satisfaction of desire. Clearly, there are many possible versions of this view. And, perhaps unsurprisingly, it seems many such versions will in fact conflict with the Engagement Constraint. This is because, as we've seen, desire is typically characterized as a state that essentially *motivates*, and not one that is essentially *affectively laden*. As a result, we've seen cases where desire, as a purely motivating state, can come apart from proper attraction or engagement. In these cases—such as Radio Man's—desires may thus be satisfied without this involving any positive affective response on the part of the subject. On many versions of desire satisfactionism, however—such as 'unrestricted' versions, which count the satisfaction of just any desire as contributing to one's welfare—desire satisfaction of this sort, too, makes one better off. It then follows that any such version of desire satisfactionism will be ruled out by the Engagement Constraint, as the theory implies that something is good for a subject despite the thing involving no affective engagement on

the part of the subject. Consequently, it seems, the Engagement Constraint may indeed rule out a good deal of desire satisfactionist views.

This is a problem for the Engagement Constraint, I think, only if these versions of desire satisfactionism remain independently plausible, and only if their independent plausibility cannot be explained in terms of the Engagement Constraint. In that case, though, this isn't really a problem for the constraint. After all, notice that the Engagement Constraint was partly motivated by the fact that subjects like Radio Man don't seem to be made any better off when their desires, which are no more than bare dispositions to act, are satisfied. Indeed, cases like Radio Man are precisely the ones that make certain versions of desire satisfactionism, like unrestricted versions, seem implausible as theories of what makes a subject's life go well.[14] They are precisely the cases that motivate the widely-held thought that, in order to capture the *actual* appeal of desire satisfactionism, the view must be restricted. And, plausibly, the actual appeal of desire satisfactionism—suggested by the many cases where desire satisfaction does intuitively seem to make one better off—can be captured by the Engagement Constraint. As we've seen, for example, provided Radio Man has no affective investment in turning on radios—nothing good he sees in turning them on, nothing regarded in a positive way—it simply seems implausible to insist that doing so is good for him. On the other hand, as soon as we imagine that Radio Man's desires are *affectively laden*—that he is thrilled at the prospect of turning on radios, that he yearns to turn them on—it suddenly seems plausible to think that their satisfaction may make him better off. It seems, then, that our constraint may in fact plausibly capture the instances of desire satisfaction that make desire satisfactionism attractive as an approach.

In general, then, while our constraint may rule out certain versions of desire satisfactionism, it seems to do so for precisely the reasons that these versions of the approach seem implausible.

---

[14] See Heathwood (2019).

Moreover, it instructs us as to which versions of the approach ultimately *are* plausible: namely, those focusing on the satisfaction of affective desire. Of course, we've only considered a few cases that support this conclusion here. Notably, though, at least one desire satisfactionist, Chris Heathwood, has convincingly and thoroughly argued for the same conclusion.[15] And another theorist, James Fanciullo, has defended an analogous point regarding preference.[16] That these people offer independent reasons for agreeing with our constraint's implications here only adds to the constraint's plausibility.

Consider next the objective list approach. On paradigmatic versions of this view, very roughly, there are multiple things that are basically good for us (hence the 'list'), and these things are 'brutely' good for us, or good for us even if we do not happen to desire or enjoy them (hence the 'objective').[17] It is perhaps unsurprising, then, that these theories are commonly objected to on the grounds that they alienate subjects from their good—and, indeed, that the Engagement Constraint will rule a number of them out. This is perhaps unsurprising because, typically, objective lists aim to expand the set of basic goods beyond those directly depending on our affective stances and lives, making them more likely candidates for clashing with our constraint. While particular objective lists of course differ in their details, there is a good deal of overlap in the goods they typically propose. Thus, on what I'll call the *paradigmatic objective list theory*, well-being consists in the possession of: life, knowledge, play, aesthetic experience, friendship, rational activity, religion, and happiness.[18] Focusing on this list will then be instructive in illustrating the Engagement Constraint's implications for paradigmatic versions of this approach.

---

[15] Heathwood (2019).
[16] Fanciullo (2022).
[17] This definition is, as I say, very rough. For complications here, see Fletcher (2016).
[18] This list is inspired by (the overlap between) several representative lists discussed in Fletcher (2016).

Predictably, of course, several of the list's proposed goods seem to involve affective engagement, or seem to be such that their possession by a subject requires the subject's having positive affective stances in connection with them. For instance, play, friendship, and happiness all seem to be things that, just in virtue of what they are, require that those who have them also experience affective engagement as a result. So, at least as far as these goods are involved, the lists seem consistent with our constraint. Problems arise, however, when we look to many of the other proposed goods. Now, things here may not be entirely straightforward, since what some of these things require will depend on how exactly we're conceiving of them. When it comes to 'aesthetic experience', for instance, it may be not unreasonable to think that this is meant to require some affective engagement on the part of its subject. Perhaps, these theorists will claim, the type of experience they have in mind involves not just encountering or seeing things of aesthetic worth, but also in some way engaging with, appreciating, or positively regarding these aesthetically valuable things. But even granting such admittedly borderline cases, there seem to be others where problems more clearly arise. As we've seen in the case of Ann, for instance, it seems religion can in some cases be part of a subject's life without this involving any affective engagement on the part of the subject. As we've imagined her, Ann has no positive affective stances whatever in connection with religion. And, as we've also seen, theories insisting that her religion—which merely shames and demeans her—is nevertheless good for her should be rejected as intolerably alienating. Similarly, in the case of knowledge, it seems we can easily imagine a subject whose possession of some inane piece of knowledge does nothing to engage them—or perhaps even drives them to madness—leaving us with no reason to think it's basically good for them.[19] And something similar can be said, it seems, for many of the paradigmatic objective list theory's proposed goods. Accordingly, this theory, and several of the goods it proposes, run afoul of the Engagement Constraint, and should thus be

[19] See Fletcher (2016: 156-7).

abandoned. Moreover, since this theory is not a single, idiosyncratic version of the objective list approach, but instead represents a wide variety of proposed objective lists, it seems many of these proposed lists, too, should be abandoned. The Engagement Constraint secures this plausible result.

This, though, is certainly not to say that all versions of the objective list approach will be ruled out by the Engagement Constraint. This is because, as we've seen, the connection we require between affective engagement and a subject in order for a thing to be intrinsically good for them is just that the thing involves affective engagement, or that the subject has positive affective stances in connection with the thing. And, as we've also just seen, there are many things that involve affective engagement, despite being the type of 'objective' goods that the objective list theorist seems attracted to. Interestingly, at least one objective list theorist, Guy Fletcher, seems to have noticed just this. On his view, the objective goods are: 'Achievement, Friendship, Happiness, Pleasure, Self-Respect, [and] Virtue' (Fletcher 2013: 214). As he notes, 'everything on the list involves the person's engagement through her holding various kinds of pro-attitudes such as endorsement, desires and affection', and as a result, his theory is able 'to respect Railton's point' (Fletcher 2013: 216). Plausibly, then, and as Fletcher himself suggests, his theory satisfies the basic constraint suggested by Railton's quote—and it seems to do so precisely because it satisfies the Engagement Constraint.[20] Friendship, for instance, necessarily involves affective engagement in connection with one's relationship with another person. And something similar can arguably be said of each of the theory's proposed goods. As this illustrates, demanding of a theory of welfare that its proposed goods involve affective engagement, or that one's possession of the goods necessarily involves one's having positive affective stances in connection with them, doesn't force us to abandon the objective list approach. It only forces us to restrict our conception of which things can plausibly be included

[20] While Fletcher emphasizes that his proposed goods all involve, specifically, 'pro-attitudes', his conception of these attitudes, given his examples, in fact seems closely aligned with our notion of affective engagement.

on such a list—and in a way, we might add, that is both plausible and desirable. After all, as illustrated by cases like Ann's, avoiding proposed goods that don't ensure affective engagement is precisely our goal in offering the constraint. So, if anything, our reflections on the leading approaches to well-being support our endorsement of the constraint we've developed.

Of course, we haven't seen anything like a complete defense of the Engagement Constraint's implications here. But, insofar as the implications we've considered seem plausible, I think they at least provide a proof of concept. As the above examples illustrate, the constraint seems to rule out just the kinds of theories that we, following Railton, were concerned to rule out in the first place. Moreover, the constraint plausibly points us toward the versions of the leading three approaches to welfare that seem most defensible. Indeed, it is telling that theorists who otherwise disagree deeply about the nature of well-being—hedonists, desire satisfactionists, and objective list theorists alike—have been moved to offer theories that satisfy the constraint, and that these theories seem to have considerable advantages over the competition. If nothing else, then, our exploration of the constraint's implications here at least seems to establish its significant promise.

Again, I of course cannot cover every proposed good or theory the Engagement Constraint rules out. But the ones we've considered so far at least suggest a kind of generalized test. On the

> *Engagement Test*: for any proposed basic good φ, if there is some possible welfare subject x whose possession of φ does not involve x's displaying affective engagement (or, does not involve x's possession of positive affective stances in connection with φ), then φ cannot be basically good for x.

This test is an upshot of the Engagement Constraint, and is illustrated by the proposed goods we've just considered. Whether you're friendly or hostile to the constraint, this test may help in

determining the constraint's implications. If what I've argued here is correct, though—and Railton's thought is indeed properly captured by the Engagement Constraint—then whether we accept this test, and constraint, will ultimately determine whether we accept Railton's thought at all.

4. Conclusion

I conclude, then, that Railton's alluring thought in fact leads us to the Engagement Constraint. This constraint best explains what it is for a theory of welfare to be intolerably alienating, as well as the near-universal intuition that such alienation must be avoided. Ultimately, it is the basic constraint on theories of well-being that, when fully developed, Railton's thought expresses. It is the source of what many of us have long intuitively sensed.

Still, there remain many issues for further thought. As I've said, we've only been able to consider a few of the proposed goods and theories ruled out by the Engagement Constraint, and this discussion has mostly been restricted to the leading three approaches to welfare. And since these are, of course, hardly the only approaches to well-being available, there may be others for which the constraint has important implications too. Nonetheless, exploring these implications must be left for another time. For now, it's enough that we've secured the constraint, test, and framework needed for this exploration.[21]

[21] Many thanks to two anonymous referees for exceptionally helpful comments. And special thanks to Dan Pallies for extremely helpful comments at a crucial time in this paper's development.

References

Bruckner, D.W. (2021) 'Human and Animal Well-Being', *Pacific Philosophical Quarterly*, 102: 393-412.

Chalmers, D. (2022) *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: W.W. Norton.

Dorsey, D. (2017) 'Why Should Welfare "Fit"?', *Philosophical Quarterly*, 67: 685-708.

Fanciullo, J. (2022) 'On Sense and Preference', *Journal of Moral Philosophy*, 19: 280-302.

Fletcher, G. (2013) 'A Fresh Start for the Objective-List Theory of Well-Being', *Utilitas*, 25: 206-20.

— (2016) 'Objective List Theories', in G. Fletcher (ed.) *The Routledge Handbook of the Philosophy of Well-Being*. New York: Routledge.

Haybron, D. (2008) *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: OUP.

Heathwood, C. (2019) 'Which Desires Are Relevant to Well-Being?', *Noûs*, 53: 664-88.

Lin, E. (2017) 'Against Welfare Subjectivism', *Noûs*, 51: 354-77.

— (2018) 'Welfare Invariabilism', *Ethics*, 128: 320-45.

Quinn, W. (1993) 'Putting Rationality in its Place', in R. Frey and C. Morris (eds.) *Value, Welfare and Morality*. Cambridge: CUP.

Railton, P. (1986) 'Facts and Values', *Philosophical Topics*, 14: 5-31.

— (2012) 'That Obscure Object, Desire', *Proceedings and Addresses of the American Philosophical Association*, 86: 22-46.

Rosati, C. (1996) 'Internalism and the Good for a Person', *Ethics*, 107: 297-326.

Sarch, A. (2011) 'Internalism About a Person's Good: Don't Believe It', *Philosophical Studies*, 154: 161-84.

Yelle, B. (2016) 'In Defense of Sophisticated Theories of Welfare', *Philosophia*, 44: 1409-18.