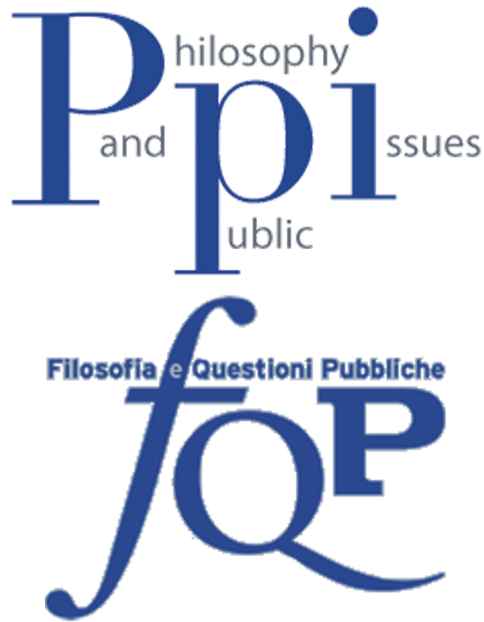
SYMPOSIUM

MORAL AND POLITICAL PHILOSOPHY FOR A BROKEN WORLD?



GOAL DIRECTNESS AND FUTURE-FOCUSED ACCOUNTS OF WELL-BEING

BY

AARON WILSON

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 125-156
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

Goal Directedness and Future-Focused Accounts of Well-Being

Aaron Wilson

Abstract. Well-being may not be the sole moral good, but a comprehensive ethics which did not take such considerations into account must surely strike us as odd, if not outright naive. The claim of this paper is first of all that the value of a theoretical approach to well-being lies not just in unifying familiar examples of our experience, but as an intellectual tool for application to new, unfamiliar cases. Such cases, where simple folk theories of well-being let us down, are a by-product of changes in technology and society. A theory of well-being, if at all possible, should be constructed in such a way that it could offer us insight into these puzzling cases as they occur. I then propose a sketch of how an explicitly future-focused account of well-being might look.

© 2014 – *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014): 125-156
Luiss University Press
E-ISSN 2240-7987 | P-ISSN 1591-0660

I
Introduction

The goal of a moral theory should ultimately be to assist us in the making of judgements concerning how to act. Many such judgements are made on the basis of the potential impact that our actions might have for well-being of ourselves and those around us. An adequate moral theory, therefore, requires an adequate theory of well-being. The notion of 'well-being' that I have in mind here is that which we talk about when we describe a certain happening as being 'good for' an individual, and can be likened to James Griffin's account of 'prudential value'[1]. In this paper I wish to make two claims about theories of well-being. The first is a methodological point, according to which a theory of well-being is of greatest value when it has potential applications to novel situations rather than just familiar ones. The second claim I wish to make is that if one accepts my proposal concerning the goal of a theory of well-being, then we should prefer to construct such a theory in a way that does not depend on features which are biased towards the 'average' human. My own preferred approach shares some similarities with Simon Keller's theory of well-being based on goals[2], however with some marked differences. Specifically, I reject his claim that goals are best thought of as only one source of well-being amongst several, and that it is the achievement of goals by individuals *whose goals they are* which is relevant to well-being.

[1] James Griffin, *Well-being: Its meaning, measurement and moral importance* (Clarendon Press 1986), pp. 3-4.
[2] Simon Keller, "Welfare and the achievement of goals" *Philosophical Studies* 121 (2004), pp. 27-41.

II

Focusing on the Future

The majority of moral judgements that we are called upon to make are simple and intuitive, and usually don't require a systematised theory of well-being in order for us to make a pronouncement on them. When someone loses their job or catches a cold, a folk theory of well-being is sufficient to tell us that such individuals have experienced harm. Such judgements come to us readily and often direct us well, however one area where folk theories of all kinds struggle is in their ability to provide consistency, or to deal with the unfamiliar. A philosophical account of well-being is of value not just in *explaining* the familiar cases for which the folk-theory is already an adequate guide, but because the philosophical account, if it is well-fashioned, should be able to guide our judgements when the folk theory lets us down. Following Feinberg I call these unfamiliar examples 'puzzling cases'[3], those in which our intuitions are either absent or else pull us in opposite directions. Such unfamiliar cases, while obviously rarer than the everyday cases, do nonetheless crop up more often than not, especially when one takes the point of view of human history as a whole.

As societies move from past to future, intuitions about things like well-being are liable to be disrupted by the new perspectives which greater access to information brings. The Athenians were convinced that some individuals were born slaves, and that the peoples to the north were mere barbarians to whom the goods accorded to citizens, such as freedom, self-determination and the right to property simply did not apply. Nearer our own time we have questioned the entrenched views about differences between

[3] Joel Feinberg, *The Moral Limits of the Criminal Law*, volume one (New York: Oxford University Press 1987), chapter 2.

genders, races, the insane, and for the first time seriously considered a moral duty of care towards animals. Questions are likewise beginning to be asked about whether environments, species, ecosystems and cultures can be harmed. The rise of secularism in parts of the world throws the previously widespread belief in an immortal soul into doubt, changing our ideas about how and when the subject of well-being persists. In almost all cases such judgements pull away from the folk theory of the time, often because they involve expanding existing concerns to beings previously deemed ineligible for one reason or another. Such changes in the scope of well-being are an inescapable consequence of human development. If we assume that our future as a society will be anything like our past, then our existing folk theories will steadily be shown to be inadequate, and new ethical challenges will present themselves. I propose that it is in dealing with an unknown future that a philosophical theory of well-being, one that goes beyond folk intuitions is required.

Of course, we cannot look directly into the future to see what challenges await us. The best we can do in the present is to try and account for those cases in own experience which are puzzling, in that they seem to conflict with some feature of our folk theory. It seems sensible to assume that in order for a theory of well-being to be 'future-proof', it must at least be able to be account for these present day puzzling cases, ideally whilst preserving as much of the folk theory as is salvageable. In the remainder of this article, I will attempt to show how my preferred theory of well-being based on goals is better able to deal with these puzzling cases than a number of the most popular alternative positions, particularly hedonist, desire-satisfaction and objective list accounts.

My use of these terms in the context of this discussion must necessarily be quite rough, as each of these labels in fact

encompasses a wide diversity of opinion, however in the interests of clarity it would be best to be explicit as to what exactly I take each of these approaches to well-being to amount to. By a 'hedonist' account, I mean any theory which takes facts about well-being to be ultimately reducible to facts about the balance of positive psychological experiences which an individual undergoes compared with negative ones. The paradigm cases of positive psychological experiences are pleasure and happiness, compared with pain and unhappiness on the other, although more sophisticated accounts may take into consideration more complex experiences such as love, satisfaction, frustration or tranquility. A desire-satisfaction account is one which takes an individual's well-being to be grounded in what that individual *wants*. A 'desire', on this account, is usually taken to be a specific kind of psychological state which disposes an individual to seek certain outcomes. Roughly speaking, when a state of affairs is an instance of benefit to an individual this is to be accounted for in terms of that individual's having a desiring attitude to that state of affairs. Most such theories resist the claim that a state of affair's being desired is both a necessary *and sufficient* condition of that state of affairs being a source of well-being for a person, however. Self-destructive or otherwise irrational desires are usually taken to be actively detrimental to a person's well-being rather than conducive to it. Finally, an objective list account is exactly what it sounds like: a list of things which are objectively good for the person who has them. Plausible candidates for such goods include virtue, liberty and happiness. Note that hybridised versions of these approaches are also common.

III

Puzzling Cases

In this section I wish to briefly outline two cases which cause problems for folk accounts of well-being. These are the issues of death and posthumous harm and the well-being of non-humans. Note that in choosing to focus on these two cases in particular I do not intend to suggest that these are the *only* puzzling cases deserving of scrutiny, only that focus on these cases is particularly instructive.

Death, it is widely supposed, is the cessation not just of life but also of the person itself. The folk theoretical approach to death is clear: life, except in some tragic cases, is a source of well-being for a person. Things which prolong one's life contribute to their well-being, and death, viewed as the cessation of life, is a great harm. Similarly the folk are quite clear that some sources of the things we take to be good for a person, like leaving a positive legacy or the health of their loved ones, relate to events which will occur after their own death. These attitudes are so firmly entrenched in the folk theory of well-being that many would consider it trivially true that death is harmful to the person that undergoes it, and that there is some value for them (even if we are initially reluctant to call it well-being) in events that occur after death. The suggestion that my own death, shortly followed by the death of all my acquaintances and the destruction of the Earth itself would not do me any harm just seems manifestly false. I will not seek to defend the claim that we should regard death and posthumous events as sources of harm, only that this is what the folk theory claims. If we regard a fit with the folk theory as desirable, then we should naturally prefer that a theory allow for the harmfulness of death and posthumous events, in the absence of compelling reason against. The puzzle in this case, however, is that well-being is value *to a subject*, the very subject

whose existence is assumed to end with death[4]. This too is a part of our folk theory, for the idea of a well-being in the abstract, devoid of a subject, simply seems like a contradiction in terms. Siding with the folk in cases revolving around death require that our theory of well-being depend on something that is not destroyed in the transition from living to dead.

Hedonistic theories, even those which are construed broadly enough to include complex or multiply realisable mental states such as 'happiness' or 'contentment' are ill-equipped to face the problem of death, for even if there were some doubt about personal survival post-mortem, anything like an experience or a mental state as we understand it must be rendered impossible once the mortal vessel has ceased to function. Capacity for pleasure is lost, as is the capacity for any feeling of satisfaction or dissatisfaction. Desire satisfaction accounts fare better, although they too experience problems of their own. Even though we might possess a desire for things which occur outside the span of our own lives, the problem remains of whom exactly is benefitted when that desire is fulfilled. It is possible that such a theorist might allow that the satisfactions of posthumous desires benefit deceased individuals retroactively, but this view too leads to some implausible consequences. If posthumous desire-satisfactions reach back in time, why not all desires? Perhaps I am unknowingly better off today thanks to the ice-cream I will eat twenty years from now. Also, suppose that I desire ice-cream today, but change my mind tomorrow. Would the satisfaction of this desire in the future still benefit me, even though I desired it for only a day? It seems odd to assume it would, but this exactly the assumption we make when we allow desires we had when we were alive to benefit us in death. We lose our desires when we die, just as surely as if we had changed our mind about them. This

[4] Thomas Nagel, *Mortal Questions* (Cambridge University Press 1979), chapter 1.

treatment of posthumous desire is of necessity somewhat rough, and it is possible a sufficiently sophisticated account of such might fare better faced with such cases. If this is to be achieved, however, then it seems that it must be achieved by bringing the desire-satisfaction account closer to the theory I intend to defend in this paper. Indeed, my own preferred approach may be viewed as a refinement of the desire-satisfaction approach rather than a member of a truly different species.

As regards those theories which take well-being to consist in a list of some objectively valuable things, the issue is simply what goods we may envisage that could both survive an individual and still realistically be described as *their* 'well-being'? Most of the objective goods countenanced by such theories are conditions like liberty, or the development of skills, but these seem to be just as rooted in the life of the subject as previous candidates.

I take it for granted that at least some non-humans experience harm, and so possess well-being. That said there a great many factors involved in human well-being which are indeed species-specific. Commonly cited components of well-being include development of virtue or the cultivation of abilities of appreciation of certain things (which is Raz's characterisation of what is 'good for' a being[5]), both of which we generally cannot attribute to non-humans.

Most attempts at species-independent theories of well-being have focused on commonalities between human and animal well-being, arguing that animal well-being amounts to human well-being less those goods requiring distinctly human faculties to realise. We all pursue pain avoidance behaviours, for instance, and exhibit distress when our basic needs are not met. Even

[5] Joseph Raz, "The Role of Well-being," in *Philosophical Perspectives* 18 (2004), pp. 269-94.

theorists normally at odds with hedonistic approaches to human well-being have taken an approach to the well-being of animals that differs very little from what a hedonist would endorse. Walker in particular proposes an Aristotelian account of 'flourishing' which acknowledges factors such as health and comfort as necessities and therefore constitutive of the good life.[6] Although animals are incapable of the ultimate aims of a human *eudaimon*, these environmental factors form a kind of 'base' for flourishing which both man and animal can achieve, thereby providing a basis for the ethical treatment of animals.

The key assumption at work in a theory like Walker's is that all beings capable of experiencing harm or benefit are appropriately similar to typical humans for comparison to be made. Instead of trying to find a truly adaptable, neurology-independent basis for well-being that could be applied to hypothetical aliens or artificial intelligences, the attempt is once again to hope that all judgements about well-being can be guided by experience of the most familiar, cosy cases of human experience. With regards to the well-being of animals this tendency is especially troubling. Although we seemingly cannot conceive of a subject of well-being that is not a human agent in disguise, our attempts to explain animal behaviour have little in common with the way we interpret human action.

Sheep, for instance, spend most of their time grazing, and this is variously taken as evidence that either the sheep experiences pleasure from eating grass, or that it must have a deeply held desire to graze. Either way we interpret this as evidence that 'sheep well-being' must consist in adequate grazing, in this case a

[6] Rebecca Walker, "The good life for non-human animals: What virtue requires of humans," in Walker and Ivanhoe (eds.) *Working Virtue* (Oxford: Oxford University Press, 2007).

simple connection is made between behaviour and well-being. Explaining human action is rarely amenable to such simple assessment however. When a human eats, we do not always assume that this means eating is good for them in every case. A human might eat because they're bored or unhappy, or as the result of a psychological compulsion that they would do better to excise. Human well-being takes the form it does because of the way it is rooted in the human condition as a whole, in which explicitly human faculties of self-reflection shape the landscape of what is valuable, including those kinds of value which depend primarily on faculties we share with animals. I do not wish to deny that human and animal motivations are never importantly similar, only that we are unjustified in assuming that the difference between human and animal motivation is simply that humans have some additional sources of well-being which animals lack. There seems a contradiction here, in that we assume outright that animals are not so complex that their motivations can be more convoluted than they appear at face value, but then attempt to jury-rig a logic to their well-being designed for a species with a far more intricate set of goals. Our usual candidates for theories of well-being, as we have already seen, require assumptions of precisely this kind in order to apply. The difference between these cases and the ones involving death is that whereas in the former it seemed as though the necessary states of being were notably absent, in these cases the issue is that the necessary states might be altogether too alien to fit within the paradigm designed for a single species.

IV

Distinguishing the 'Subject' and 'Vessel' of Well-Being

Perhaps the most seemingly intractable obstacle thrown up by the puzzling cases is that there are times when we want to allow that a person's well-being can be affected even when the subject of that well-being cannot be easily established. In the earlier discussion of death for instance, the concern was raised that dead persons cannot have well-being because they had ceased to exist as persons, and as such we were unable to reconcile our intuitive notion that death is usually harmful to us with the equally intuitive notion that harm requires a subject. One option we have already explored is to claim that certain events that occur after death retroactively affect the value of a person's life. For example, Bertrand Russell's life might have been more valuable had his efforts campaigning for nuclear disarmament actually hastened the abolition of atomic weaponry. Although this is quite a plausible assessment of how posthumous events can affect the value of a life, it is unclear why we should regard this value as 'well-being' rather than some other kind, such as moral value[7]. Well-being must amount to something which is good *for* the person to whom it applies, rather than something which is merely good *about* them, which is what the Russell example appears to suggest. Compare this with the tragic case of the victims of the Dunblane massacre in 1996. It seems plausible to assert that the lives of the victims gained some value retroactively over the following weeks when the shocking events of the shooting precipitated the banning of handguns in the UK, potentially saving many lives. Few would argue that this enhanced the *well-being* of the victims, though. In order to explain why well-being, as opposed to moral value or 'significance' of a life might be able to increase or decrease posthumously, we need to give a

[7] James Griffin, *Well-being*, pp. 2-4.

characterisation to the relationship between a person and their well-being which supports that assessment.

The seeming difficulty here, I suggest, is a consequence of mistaking questions of the *subject* or owner of well-being for questions of the *vessel*, or what one might call the bearer of that value. We have established that for some value to be a case of well-being there must be an individual whose well-being it is. We have yet to establish that such value *inheres* in that person, that this person is both the subject *and* the vessel for well-being, and in fact we have reason to believe it does not. The most promising strategy for reconciling death as the termination of the subject with death as a source of harm for the subject is to take the vessel of well-being to be something which *does indeed* persist beyond death, or at least remained relevant to the establishment of harm and benefit. Whatever this relevant feature is, however, *cannot* be a property of a person, because objects which do not exist cannot instantiate properties, and the death cases we examined assume the termination of the person's existence. The person's *body* does of course persist, but this is now simply inanimate matter, no more a likely vessel for well-being than sand from the beach. If it is indeed sensible to talk about well-being with reference to death and the dead, then despite the fact that well-being can only be understood with reference to a person by definition, it cannot be a condition which inheres in persons, even the person to whom well-being applies. Contrast this with other inhering states or conditions of objects, for example being a certain temperature. If an object ceases to be then it can no longer possess temperature, because that condition of being a certain temperature has no subject in which to inhere. Our best option for reconciling death with a theory of well-being, therefore, is to reject the view that well-being is a persisting state of the person altogether. It is perfectly simple to make use of concepts like 'harm' and 'well-being' without assuming that such prudential value must

correspond with such a state. Prudential value may require a subject to which it applies, but we are under no obligation to take the subject of application to likewise be the vessel of that value also. In terms of David Velleman's contrasting of 'momentary' well-being with well-being over an extended period of time[8], I can be understood as claiming that the best prospects for a theory of well-being which preserve intuitions about the harmfulness of death involve rejecting the latter notion in favour of the former.

In taking this step we would be giving up on a deeply entrenched way of thinking about well-being, although I suggest abandoning our intuitions about death would be the larger sacrifice. I suggest that this entrenchment arises from the fact that in our everyday talk it is natural to take stock of our well-being by reflecting on our psychological attitudes, which are indeed persisting conditions. Additionally we can talk about our mental states in terms of being happy, happier than we were, even the happiest we've ever been, and that we are mostly, completely or not at all content. As such the natural assumption that we are liable to make is that well-being is a *state we are in* at a given time that represents a quantity of value, much like a bank account, from which external events can add or subtract amounts. I previously claimed that we should respect the folk theory of well-being insofar as that is possible, but it appears that abandoning the bank account view will ultimately be a price worth paying.

The motivation for the bank account approach to well-being does not, it seems, stem from any inherent attractiveness to a folk theory of well-being. In fact, I suggest that there is nothing about the way we talk of well-being to suggest that *value for* a person must be *value in* a person at all. Rather, if there is any advantage to speaking of well-being in this way, then this advantage is in its

[8] David Velleman "Well-Being and Time," in *Pacific Philosophical Quarterly* 72 (1991), 48-77.

amenability to the making of social policy decisions rather than being in accordance with our most natural ways of thinking. If well-being is quantitative then it is (theoretically) comparable between people, and allows for easier judgements about the distribution of society's goods. Rawls' minmax principle requires the identification, at least in principle, of the person in a society with the lowest welfare, and insists that inequality in distribution be tolerated only if this person's welfare would be higher under such a distribution than under a more equal one. Utilitarian theories have a similar need for this quantitative account of welfare. Utilitarianism is committed to the rectitude of the greatest aggregate welfare, and as such requires that there is something which can, at least in principle, be aggregated. Even amongst these theorists however, there is a certain amount of wariness against taking the concept of comparable, quantitative welfare too literally. Rawls is keen to make clear that relative welfare cannot be easily calculated from the original position, and so brings in a notion of 'primary goods' to stand in for welfare in the judgements regarding fair distribution instead, whereas the various difficulties involved in measuring utility hardly need repeating here.

In fact, we observe that writers less concerned with the problems of aggregation already tend to bypass the bank-account approach to well-being altogether. Feinberg in his study of harm and benefit is concerned not with utility maximisation but with the practical application of the harm principle, and as such concerns himself only with occurrences of harm and not with the relation between incidents of harm and a person's overall well-being, yet his project does not suffer as a result of this difference of focus.

The precise articulation of harm being a 'setback to an interest' is also attributable to Feinberg[9], although he is more concerned to differentiate types of interests that can be harmed than explaining in detail what an interest actually is and how it is constituted. In claiming harm to be the setting back of someone's interests, Feinberg is attempting to provide the same theoretical resource that we are striving for: the facts about a person that explain why something has an effect on their well-being. The fact that he additionally considers moral wrong to be something which is to be viewed from an individual perspective is an early point of divergence between what he and I propose, according to which Feinberg's account is similar to that implied by the private ownership theory of moral value proposed by Moore[10]. It is worth noting that Feinberg's implied endorsement of this approach to moral value may simply be a consequence of his focus on establishing the boundaries of the harm principle rather than developing a comprehensive theory of welfare, a fact which is also likely responsible for his lack of detail on the nature of interests. He instead seems to take a person's interests for granted and focus on which harms to them are allowable. Nonetheless, Feinberg's notion of well-being as connected with interests suggests a promising avenue by which to develop a 'future-focused' account of well-being. Interests, I propose, should be our preferred candidate for the *vessel* of well-being.

Let us say then that the vessel of well-being is indeed an interest. We might then say that harm to a person consists of harm done to, or perhaps we might more accurately say damage or a setback to, something which is an interest of that person. Benefit, conversely, is to be thought of as promotion of an

[9] Joel Feinberg , *The Moral Limits of the Criminal Law*, volume one, chapter 1.
[10] L.W. Sumner, *Welfare, Happiness and Ethics*, (Oxford: Oxford University Press 1996), section 3.1.

interest. What I have in mind when we speak of 'promoting' an interest is simply making that state of affairs more probable, or working to bring it about in some way. This, I suggest, is an acceptably natural way of thinking about well-being that accords suitably well with our folk theory. If a consequence of adopting such a view is that we must take talk of well-being in the aggregate, or over time, to be nothing more than a useful heuristic for assessing the best, most equitable disbursement of resources then so be it.

I suggest the following terminology for interests. An interest of an individual is a *state of affairs* which has some value, either positive or negative, for that individual. For something to be 'in a person's interests', is for that state of affairs to come about or to persist. Suitable candidates for interests are not restricted to psychological states, or to states of affairs which involve persons themselves in some way. In this way, the account of well-being I offer distinguishes itself from desire-satisfaction and hedonist accounts. Positive psychological states or the satisfaction of desires may well contribute to well-being some, or indeed most of the time on the account I propose, however neither one of these is a requirement of a state of affair's having value for a person. Leaving aside the question of what makes a state of affairs an interest for the moment, taking states of affairs to be the vessels of well-being rather than some feature of persons or persons themselves has some immediate advantages for a theory of well-being.

For one thing, in disentangling the subject of well-being from the vessel of well-being, we gain the ability to talk about value for a person after that person has ceased to exist. Suppose we wish to say that it is in my interests that I have great-great-grandchildren. It is no more a problem to say that my great-great-grandchildren stand in a relation of 'valuable for' with me than it is to say that

they stand in a relation of 'being the descendants of' with me. The 'interest' relation makes no implicit reference to my being alive at the time, or to my having certain properties any more than the 'descendant of' relation does. This is not the only advantage to such an approach however.

If the vessels of well-being are thought to be unrestricted states of affairs rather than residing in a quality of persons, then we may expect an easier time speaking of the well-being of a non-human entity. Recall that many of the existing accounts struggled to accommodate the possibility that novel kinds of entity might be possessed of a capacity for well-being just as potentially rich as that of a human. The account of interests as states of affairs does not have that problem, for there are no entities which are excluded *in principle* from taking certain states of affairs as sources of value. Consider a simple, single-celled organism such as an amoeba. It is generally speaking difficult to speak of such a creature as having well-being precisely because it is assumed that the vessels of well-being are things which the amoeba cannot possess, such as psychological states. As previously noted, however, it is a virtue of a theory of well-being that it is adaptable to novel situations. While it is entirely possible that the amoeba does in fact lack the capacity to have well-being, we should consider it an advantage of a theory that it does not exclude any living thing from even the *possibility* of having well-being. We must be able to ask questions about amoeboid well-being, even if we ultimately conclude that it has none, in the same way that we may one day be called upon to ask questions concerning the well-being of machines or ecosystems.

V

The Interest-Making Relation

If the vessel of well-being is the interest conceived of as a state of affairs, we then require an account of what grounds the relationship between an individual and their interests.

The proposal I have in mind for this is in many respects similar to the theory of achievement of goals put forward by Keller, most notably in taking those things which are actively sought by an individual to be a source of value for them[11], and in advocating for an 'Unrestricted' view of such sources of well-being. While he terms these sought-out states of affairs 'goals' I prefer to retain the term 'interests', as I will ultimately seek to claim that some things can be sought after implicitly, or more or less passively in the absence of explicit adoption of that thing as a goal. Furthermore, my proposal concerning interests is in fact intended to be broader than Keller's theory of goals, as he considers the achievement of goals to only be one among many possible sources of well-being, rather than a theory of well-being itself. Additionally I reject Keller's view that it is the personal achievement of goals by the person who has them which is always a necessary component of their value to well-being.

The Unrestricted View of well-being holds that the achievement of goals, or the coming to pass of interested states of affairs in my terminology, is intrinsically good for an individual irrespective of what those goals actually are. If a state of affairs is an interest of an individual then it will be a source of some well-being to them irrespective of whether or not such goals are those that would be adopted by a rational being or whether the promotion of such a goal might in fact be a setback to some more important interest. Although it is intuitively plausible to

[11] Simon Keller, "Welfare and the achievement of goals," p. 32

hold that at least some goals make this contribution to well-being, it is less apparent that this is true of all the goals that people set themselves. To use Keller's own example, eating a handful of gravel does not seem to advance one's well-being, even if one has fervently wanted to do so their entire life[12].

The objection to the Unrestricted View on this basis is that interests themselves can seemingly be assessed as valuable independently of their status as interests in the first place. Unfortunately it is beyond the scope of this paper to present a complete defence of this position here, so I must ask the reader to simply assume it in the discussion that follows. I do however hold that the criteria I present for a state of affairs' being an interest is sufficiently sophisticated to rule out bizarre goals such as eating gravel from being true interests in most cases, although it is likely that there will be at least some instances where bizarre interests are indeed permitted by my account. In actual fact, I consider that this additional flexibility might be a valuable feature of a future-focused account of well-being. If our theory were such that all unusual or novel interests were to be excluded in principle, then we would fail in our goal of providing a theoretical account suitable to be applied to unfamiliar cases. Provided that there are adequate restrictions in place against making our account of well-being overly permissive, we should not be concerned that it allows for at least some unexpected consequences concerning the interests of entities.

In simplistic terms, I propose that an individual's interests are the states of affairs which that individual is disposed to bring about. To be the object of some kind of directed effort on the part of an agent is all that there is to being an interest. In basing my account of interests on dispositional states it could be argued that this approach is more of a refinement to the desire-

[12] Ibid., p. 28.

satisfaction based account than an outright replacement for it. I would not resist this comparison, however I take the notion of a 'desire' to be too narrow to capture what I intend when I speak of 'seeking to bring about' a state of affairs.

We do not come to understand what an individual's interests are simply by observing their behaviour and 'reading off' their dispositional states, as we would come to understand their desires. For one thing, it can be assumed that at least some behaviour is a product of coercion or of false beliefs, which do not represent the underlying attitudes of the individual in question. This is not the criticism I have in mind when I say that simply taking behaviour to indicate dispositional attitudes is too simplistic however. The desire-satisfaction theorist can account for such cases easily enough simply by insisting that only action undertaken in the absence of coercion and in the absence of false beliefs should be taken to indicate the presence of a desire. At the very least we should adhere to such a restriction in our own project, but it is not the case that in inventorying all the desires of a person we arrive at a list of their interests also. An individual's interests, rather, are calculated based on the best possible interpretation we can give to their behaviour conceived of as goal-directed, and viewed as a whole. I should make it clear at this juncture that I should not be read as claiming that interpretation is simply the means by which we *discover* or attribute interests to individuals, but the stronger claim that a person's interests just are those states of affairs which the best possible interpretation of their behaviour would indicate their interests to be. What I propose might be described as a kind of *interest functionalism*, analogous to functional accounts of mental content. Interests, as I have already argued, do not reside in the mind in anything like the manner that it is assumed that mental content is bound to the mind, but we may nonetheless take a functionalist stance on how states of affairs come to be the interests of persons. One

immediate objection that one might raise to such a view is that it appears to entail that in cases where two equally good interpretations of an agent's behaviour are possible, their interests will be indeterminate. To this objection I would first of all caution that such cases are likely to occur relatively infrequently in practice. The interpretive project I have in mind is intended to draw upon all available data in attributing interests to individuals, including (where such things are possible) their own self-reports. While self-reports are of course not foolproof guides to the best interests of a person, we should only expect indeterminacy of interests to the extent that there is ever a serious discrepancy between the behaviour of a person and *their own* sincere interpretation of the goals underlying their behaviour. If a person is indeed in such a sorry state, then why would we expect their goals to be amenable to a single coherent interpretation in the first place? Such cases may well exist, but if so, then all we need conclude is that in these rare cases, a person's interests are under no obligation to be any more settled than their owner is. This certainly need not disadvantage us much from the point of view of deciding how we should act towards such people. Even if the totality of the interests attributable to such individuals are indeterminate, chances are that many interests within that totality can indeed be attributed to them unproblematically. As to the remainder, we may simply wish to withhold judgement for the time being in the hope that new information will reveal dispositional states that were previously hidden.

The fact that understanding the interests of an individual is an inherently interpretative project is crucial to the account I propose. In our everyday dealings with people, we seldom arrive at decisions about their personalities or values on the basis of isolated incidents. Instead we observe them over time, noting their tendencies towards certain ways of behaving and approaching life in general. This was amongst MacIntyre's

insights when he claimed that human action is intelligible only through an understanding of the narrative histories surrounding those actions[13]. Whether MacIntyre was correct that narratives are the *only* way in which human lives and action can be sensibly considered remains a contentious issue (see Strawson[14] and Woods[15]), but what is of greatest importance for our project is the relationship he proposed between interpretation of any kind and the concept of a virtue. Virtues, he claimed, were not comparable to mere dispositions to act a certain way at a certain time, or even a disjunction of such dispositions. Rather they were present in different ways throughout a person's life. For example we come to recognise courage not by recognising the set of dispositions which together suffice for courage, but by locating a person's actions within a narrative that renders them intelligible. This is the proper way of understanding interests, I propose. Just as the virtue of courage cannot be straightforwardly identified with the tendency to stand one's ground in battle, an interest cannot be identified with a single set of dispositions, self-reports or desires, but must be interpreted from them when viewed within the proper context. This form of interpretation need not be strictly a narrative one, as I believe it is possible to accept my proposal concerning interests as constructed in this way without subscribing to such a view of the self or action in general. Also, in practice narrative accounts of the person commonly rely on stereotypes which impose value based on how a person is to be categorised rather than their idiosyncratic approach to life, for example narrative histories of what constitutes a good life for a woman, or a member of a slave class. On this basis, we may be

13 Alasdair MacIntyre, *After Virtue* (Gerald Duckworth & Co. Ltd. 1981), chapter 15.
14 Galen Strawson, "Against Narrativity," *Ratio* 17 (2004), pp. 428-452, at 428.
15 Angela Woods, "The Limits of Narrative: provocation for the medical humanities," *Med Humanities* 37 (2011), pp. 73-78, at 73-74.

inclined to reject a specifically *narrative* kind of interpretive project, but the importance of interpretation itself cannot be denied.

We engage in a similar interpretive activity in the process of reflection on our own interests as we do with those of others. Even if we assume that we have a special access to many of our own desires and preferences as they occur, often we will need to consider critically whether these desires are merely fleeting or evidence of a more deeply held aspiration. Our interests are quite literally a *product of* this critical reflection on behaviour and mental states. As such, we have no special access to them the way we have access to some of the features of our mental lives. This, I suggest, is quite intuitive. No child is born knowing what they want to be when they grow up, but rather must work hard to reveal their own attitudes over the course of their lives. A fleeting desire or passing fancy is not a sound basis for determining the course of one's life. Instead one must come to an understanding of oneself in order to make such a decision. Interests, as viewed by the person who possesses them, are those states of affairs which they *judge* to be good for them, rather than those which they simply feel a desire for or approval towards. Let us suppose that on a particularly trying day Van Gogh formed a strong desire to give up his artistic career and become a businessman. He probably realised at the time that this was an unusual desire for him, that yesterday he wanted to be a painter and tomorrow he would once again. This desire would only be evidence of an *interest* in becoming a businessman, rather than a fleeting desire, if it was plausibly corroborated by everything else Van Gogh knew about himself and his goals.

The distinction between narrow 'desires' and broad 'interests' in place, we can see how the interpretive project can allow us to calculate a person's sources of well-being in a satisfying way when

their attitudes and behaviour are more stable. If an athlete spends enough time training, one possible explanation is that there is value for him in one day competing in a prestigious sporting event. In that case, we would be entitled to view a knee injury that ended his athletic career as a setback to that interest and therefore harmful to him. Another, equally possible explanation however is that his overbearing parents pressured him into a life that he finds tedious and unrewarding, however after years of brainwashing he has come to internalise his parents' wishes and comes to believe in the desirability of his training, even though he gets no more satisfaction from it than he used to. In the latter case it appears that there would be no value from his point of view in his success as an athlete, so a knee injury would not be harmful to him, or at least not harmful for the same reasons. It may actually be beneficial to him, in that it enables him to leave the career that he found unrewarding with impunity. In either instance we could imagine that his beliefs and desires about what held value for him were the same, and yet we arrived at differing conclusions about what was harmful to him. Something makes the difference in this example, and that something is the broader context in which the person's attitudes and behaviour is understood, and that could only be arrived at through the process of interpretation I have described. This is the strength of an account based on 'interests' as I have formulated them, over narrower dispositional attitudes.

VI

Issues for Well-Being Based on Interests

Based on this analysis of what interests actually are, we can point to a number of consequences such an account has for the

kinds of things that can be sources of well-being. I shall now deal with a few of what I consider the most important of these.

For one, the only entities which have the capacity for well-being are those capable of behaviour sophisticated enough for the interpretive project to get off the ground. This rules out simple deterministic devices such as thermometers from having interests, since the behaviour of a thermometer cannot be viewed as directed towards any specific end. More complex non-human entities like ecosystems, however, can be interpreted in this way. Free of interference, ecosystems generally move towards the greatest possible degree of biodiversity, with a web of predatory behaviours engineered to support this goal. It is then sensible to speak of an ecosystem being harmed by any action which would decrease biodiversity within it. Secondly, as interests require goal-directedness, only contingent states of affairs can be sources of harm or benefit. States of affairs which are necessary or impossible cannot be related to well-being. It might be the case that everything I undertake to bring about in my life is dependent on the truth of some necessary state of affairs, for instance [Pv¬P], where P might stand for literally any contingent state of affairs. It seems incorrect to say however that because [Pv¬P]'s holding is a necessary condition of any interest I have being fulfilled, I derive some benefit from it, except in the most trivial sense.

It might potentially be questioned whether or not all interests that a person has actually require the process of interpretation I have described in order to be revealed. To what extent, if any, are some of the interests of persons universally possessed, at least amongst persons of a certain species? It is strongly intuitive that ensuring a balance of positive experiences in one's life over negative is an interest which all sane humans have (at least when it is not overruled by some more valued interest of theirs) simply

by virtue of their being humans and possessing the appropriate capacities for happiness and sadness. Also, all living things are programmed to promote the survival of their genes, so could this be an interest universal to all living things? I would be inclined to resist this line of argument as far as possible. In any population of sufficient size there will be outliers and exceptions to any norm. Buddhist monks or medieval flagellants had behaviour patterns that did not suggest either of these supposedly universal interests were held by them, and interpretation of a person's behaviour, in the broadest sense, is how we a person's interests are established in the first place. I am not claiming that these admittedly unusual individuals lacked any inclination towards seeking pleasure over pain, or were never even tempted to act in accordance with the biological urge to procreate. Nor is it necessary for me to deny that giving in to such temptation would even have been a source of happiness for them. What I deny is simply that mental states which are pleasurable are conducive to well-being in virtue of any intrinsic value of the state itself, in the absence of a revealed interest with which it is in accordance. In order for such states to be interests of a person, their behaviour and attitudes must direct them towards pursuit of such states. No experiences, not even those which are usually considered desirable by default receive automatic elevation to the status of 'interests' without going through the same process of interpretation as any other state of affairs.

If interests are constituted by interpreting how people act and self-report over time, then we would expect many of our interests to be quite broadly defined. Even so, there is scope in this approach for more passively held interests that are not *explicitly* sought out by an individual. No one would deny that I have an interest in a meteorite not striking the Earth and wiping out all life on the planet, despite my not being able to act in such a way as to affect the probability of this occurrence. I may never

outwardly act as though I believe such an event might be bad for me, or ever articulate it, but if one observes my behaviour and vocalisations then one will realise that many of the interests which I can be attributed as having depend for their success on all life on Earth not experiencing fiery death in the near future, which would be sufficient grounds for ruling that such a thing was against my interests.

Once we allow that some interests can be held passively, however, it could be objected that this implies that we have infinite ranges of interests at any given moment, since many such interests are merely implied by other interests of ours. At this moment in time I am writing a sentence, but precisely which interests of mine am I evidencing in doing so? It is maybe correct to say that I have an interest in finishing typing this paragraph, and an interest in finishing this paper, and an interest in a successful career as a philosopher which is best served by work on this and other papers in the future. I might even have an interest in one day being able to reflect on my life's work with satisfaction, any one of which interpretations are supported by my writing that sentence a little while ago. I do not however consider this kind of infinity a particularly troubling possibility however. Many states of affairs which have value for me will be valuable for me under more than one description. The writing of a sentence may have value for me because it is a part of a paper, the completion of which I value independently, at the same time as I value it as a part of a larger body of work, or as a job well done in its own right. Interests, it seems, may 'nest' one within the other. Some interest of mine may be promoted by the coming to pass of some other state of affairs, just as my writing of a sentence promotes my satisfaction at a body of work that I can one day feel proud of. We should not resist this conclusion, but rather embrace it. If we were unwilling to allow certain states of affairs to be good for us because their value to us were subsumed

into some other, more general goal, then our theory of well-being would be immediately crippled. Our interest in maximising our own personal well-being would be the only possible interest that could not be so subsumed. This would be an undesirable consequence of our theory, and so I propose that we allow that sources of well-being sometimes have value as constituent parts of more general goals, even if this does lead to our having infinitely many ways that we might be benefited at any time.

The existence of passive interests draws attention to an important distinction between *value for* a person and *value to* a person. When we speak of something as being *valuable to* a person, it seems that what we have in mind is that such a thing is looked upon in an approving way by that person. Quite literally, it is valued *by* them in virtue of some attitude they have to that thing. When we speak of something as being valuable for a person though, we should not be interpreted as claiming that this person necessarily has some kind of valuing attitude. Looking upon a state of affairs as good or valuable is neither necessary nor sufficient for something's being *valuable for* that person. Four-year-olds rarely have valuing attitudes towards fresh vegetables, yet obtaining sufficient vitamins from such foods are undoubtedly valuable for anyone for whom health is a source of well-being. Similarly, not everything towards which someone has a valuing attitude is necessarily relevant to their well-being. One may consider talent at playing the banjo to be valuable without in fact enjoying the music from this particular instrument. A valuing attitude may often be indicative of an interest, but is no more a guarantee that a state of affairs will be a source of well-being, taken in isolation, than a desire is.

VII

Revisiting Death and Non-Humans

I suggest that thinking of interests in this way, as states of affairs which agents undertake to bring about, can assist us in a satisfactory approach to the puzzling cases I drew attention to earlier in this paper. The issue with such cases, we established, was that sometimes the *vessel* of well-being appeared to be destroyed but the capacity for well-being was not, while at other times certain entities which appeared able to possess well-being did not meet the standards necessary for being such vessels, at least not without some pretty large assumptions being made.

The theory of interests as states of affairs aims to solve that problem by making interests themselves the vessels of well-being. According to such an account, when we speak of well-being as valuable we are in actual fact speaking of the value of certain states of affairs *for* some individual. If I claim that someone lives a charmed life of great well-being, I am claiming that their life is full of things which are valuable for them. Although I may speak in a loose sense as though their well-being is a thing which somehow resides within that person, we must avoid taking such talk too literally. The way we should understand talk of having well-being, I have argued, is analogous to the way we should talk of someone as having many admirers. The admiration directed towards someone is evidently a fact concerning them, but admiration does not *inhere* in its object. It if it is sensible to speak of such a thing as inhering at all it must surely inhere in this person's admirers, who are the vessels of that admiration. Admiration is a thing which may plausibly survive the destruction of the thing which is admired, but not the destruction of the things which do the admiring. Well-being, I have claimed, is something like this.

When I die, the facts about which states of affairs I undertook to bring about in life will remain unchanged. While I was capable of behaviour, I adopted goals, sometimes explicitly and sometimes implicitly. A corpse is incapable of such goal-directed behaviour. In any case, if we assume that death involves the termination of a person's existence altogether, it seems inappropriate by definition to take any activity we do observe from the decomposing mortal coil to be indication of the *person's* interests. Our interpretive project is therefore limited, as it should be, to the events of that person's life. Based on this, we can come to a conclusion about what states of affairs had value for that person, and acknowledge that even post-mortem, such occurrences still have this feature which qualifies them for consideration as harms and benefits.

As to the question of whether or not death itself is a harm, given what we have said so far about interests, the answer is 'sometimes'. Death itself is harmful only to the extent that it is a setback to one's interests. Quite possibly this would be true of most deaths, although by no means all. The harmfulness of death, now that it can be established to be potentially harmful at all, depends entirely on whether or not it would adversely affect any of the things which I cared about. Suppose I wished my children to continue to live long and happy lives, or that a charitable foundation I set up continues to do its work into the future, then these interests could not be harmed by my dying. Other interests of mine, for instance those interpreted from my fondness for good coffee and good books, will indeed be harmed in as much as I will never experience any of those things again.

The understanding we can offer to non-human well-being is similarly advantaged. The account of well-being I offer is based on interpretations of behaviour rather than any specific

psychological or somatic capacity. The benefit to the issue of non-human well-being is twofold.

Firstly, no entity capable of interpretable behaviour is excluded from the domain of creatures with well-being. The presence or absence of neurological features indicative of pleasure, pain, rationality, or a theory of mind or self are not relevant to determining whether or not a given entity might be deserving of consideration as having well-being. Of course, whatever knowledge we have about the capacities of such creatures must be brought to bear in carrying out the necessary interpretive project, but by taking interests as the vessels of well-being rather than psychological capacities we do not exclude any entities capable of behaviour from consideration until after their behaviour has been investigated. It might be that we ultimately deem the evidence for their having interests to be insufficient, but at least we have the ability to even ask such questions of these beings.

Secondly, a view of well-being as based on interests does not make the assumption that the human experience is the paradigm of how things can be 'good for' an individual. Much discussion of the well-being of animals, as we have previously noted, makes the assumption that the well-being of animals is like that of humans, but lacking access to those sources of harm and benefit which their more limited faculties deny them. Well-being based on goal-directed behaviour allows for the possibility that some animals may have sources of well-being that humans lack, or that both humans and animals lack some sources of well-being applicable to different, novel entities.

VIII

A Future-Focused Account of Well-Being

As useful as it is to be able to apply our theory of well-being to the puzzling cases, such theories, we have decided, are to earn their keep by their applicability to novel situations, involving circumstances and entities not currently included within the existing sphere of concern. Here, too, I propose we may be optimistic concerning the prospects of the interest based approach I describe. The theory of interests I describe is far less limited in terms of what kinds of things it can take to be sources of well-being, and what kinds of beings have the capacity for such value. Historical change has always brought with it changes of this kind, and we have no reason to doubt that it will continue to do so. My proposal is simply this: that instead of dismantling our systematised theory of well-being every time our folk theory receives a shock, we should aim to construct a theoretical account that is capable of accommodating these changes in our intuitions. As new entities or new situations begin to demand our attention, these intuitions about what can experience harm and under what circumstances are of course bound to change. This approach suggests that such change can be viewed as a change in our evidence, demanding revision of how we *apply* our theory, rather than the theory itself. If such flexibility in a theory is possible without sacrificing fit with our existing folk conceptions, we should probably pursue such a theory above others.

University of Manchester

If you need to cite this article, please use the following format:

Wilson, Aaron, “Goal Directedness and Future-Focused Accounts of Well-Being,” *Philosophy and Public Issues (New Series)*, Vol. 4, No. 2 (2014), 125-156, edited by S. Maffettone, G. Pellegrino and M. Bocchiola