

L.W. SUMNER'S 'WELFARE, HAPPINESS, AND ETHICS' ('WHE') SUMMARISED

ANDREW SINSTEAD-REID

Research Fellow, Tax and Transfer Policy Institute
Crawford School of Public Policy, Australian National University

[Page numbers are those in the 1996 hardback edition of the book (which are the same in the 2003 paperback reprint); 'Macquarie Dictionary' spellings of words have been employed throughout.]

0. Preface (pp vii-x)

[0.1.] "The project of this book [*WHE*] is a philosophical inquiry into the nature and value of welfare. Its main conclusions can be easily summarised. The first six chapters seek the best theory about the nature of welfare. ... The conclusion drawn from [the critiques of alternative theories of welfare in Chapter 3 (objective theories), Chapter 4 (hedonism) and Chapter 5 (desire theories)] is that philosophical theories about the nature of welfare have so far been too remote from our everyday experience of it. Chapter 6 sets out to remedy this deficiency by developing a better version of a subjective theory, one which connects welfare with happiness or life satisfaction. The final chapter then defends welfarism: the thesis that individual welfare (so understood) is the only thing with final or ultimate ethical value, the only state of affairs we have a moral reason to promote for its own sake" (*WHE*, viii).

[0.2.] "The unequal division of labour between the two major concerns of the book – six chapters on the nature of welfare to one on its value – reflects my (possibly jaundiced) assessment of the current state of the argument concerning welfarism. ... [Welfarism's] rejection has too often been based on a misrepresentation of the nature of welfare. ... [T]he only way to counteract it is to show why the established theories about the nature of welfare – the ones typically taken for granted in discussions of welfarism – are inadequate, and to construct a better theory. Welfarism deserves to be tested on the basis of the best theory about what welfare is [after which] making a plausible case for welfarism is then a comparatively straightforward task" (*WHE*, viii).

[0.3.] "Although welfarism tells us that only welfare matters in its own right, it dictates no particular way of using it to ground our familiar moral categories. ... It does require holding, contrary to some non-consequentialist moral structures, that the good is [both logically and ethically] prior to the right. ... If it turns out that a good case can be made for welfarism, and for a subjective conception of welfare, that will be the time to ask how we should go on from these results to build a moral theory" (*WHE*, viii-ix).

1. The Concept of Welfare (pp 1-25)

[1.1.] We can all agree that welfare (whether our own or that of others) matters (not just as a practical concern, but also as a prominent feature of our 'common-sense' morality). But: why?

[1.2.] Well, to even *be* a 'subject-of-a-life', a person must have *some* view of what will make their life go better or worse for them (likely accompanied by a settled preference for the former over the latter condition), and either make their major life decisions accordingly, or fail to do so (and regret this as a failing later). A 'threshold' degree of 'self-centredness' is therefore required for any person to be a 'subject-of-a-life' which is truly *theirs* (as "[f]alling below this minimum, having too little regard for one's own good, is not a virtue but a pathology, not altruism or saintliness but debasement or servility", *WHE*, 1-2).

[1.3.] With the exception perhaps of psychopaths, our (typical) condition as social beings renders impractical the pursuit of pure egoism (as we can scarcely avoid extending concern for the welfare of others to those to whom we are closely connected; but most of us also manage to extend this concern to those to whom we are not closely connected, or connected merely by a shared 'species-identity'). "... [A]s would-be egoists, we are dismal failures" (*WHE*, 2).

[1.4.] Our 'common-sense' morality therefore requires us to show due concern or respect for others (as a general constraint), "and for ourselves as well, in case we are not already naturally inclined in that direction" (*WHE*, 2). More specific constraints stipulate regard for the 'best interests' of those entrusting their well-being to (an)other(s).

Traditional virtues are also ‘thickened’ by welfare-oriented concepts, such as harm and benefit, without which there would be little left in our ‘common-sense’ morality: “not only little substantive content, but few of our thick ethical concepts as well” (*WHE*, 3).

[1.5.] “The centrality of welfare in ethics has long been recognised by moral philosophers. ... Welfarism is one possible answer to the question of how much welfare matters for ethics: it counts for everything (*WHE*, 3). ‘Welfarism’ is a theory of the good (i.e. that the ‘good’ to be promoted by an ethical system is the general ‘welfare’ of ‘welfare subjects’). In answering questions about the nature and ethical value of welfare, it makes sense to begin with its nature (“since in general we expect the value of a thing to depend on its nature, rather than vice versa”; *WHE*, 4). “Most of us manage our ordinary lives on the unreflective assumption that well-being is worth pursuing or promoting and that ill-being ... is equally worth avoiding or preventing, that it is a good thing for lives to go well and a bad thing for them to go badly, that it counts in favour of some activity or condition that it is beneficial and against it that it is harmful, and so on” (*WHE*, 4).

[1.6.] So: “... issues [regarding the value of welfare] can readily be connected to the concerns of our everyday lives. In this respect, though intricate and even abstract, [these issues] remain intelligible and accessible [to all] ... [as] surely we must already know what welfare is, since we all seek it in our everyday lives. How, then, could a philosophical inquiry into the nature of welfare ever yield any novel results? Indeed, if we do not already agree on what counts as welfare, how could it ever yield any results at all?” (*WHE*, 4-5).

1.1. The Nature of Things (pp 5-10)

[1.1.1.] Like theories of causation, theories of welfare “seem to occupy a middle ground between the merely conceptual and the fully empirical” (*WHE*, 6), based on the presupposition of a ‘shared network of concepts’ “without which we would lack prereflective agreement concerning what is to count” as welfare, “generalising over all such [notions of welfare] in all domains, telling us what must be true of any states of the world for them to be [*about* welfare]” (*WHE*, 6). The aim of an inquiry into the nature of welfare “is not to tell us what is good or bad for us, or to advise us on how to attain the former and avoid the latter. ... [A]ny prudential advice presupposes some general account of what it is for something to benefit or harm us, thus what it is for our lives to go well or badly. ... [A] theory of welfare ... provide[s] this constitutive account. ... [I]t yield[s] truth conditions for claims about, or assessments of, our interest or well-being [or existence conditions for relations between *sources* of intrinsic prudential value and whether such sources are prudentially valuable in our lives]” (*WHE*, 7). Such an inquiry “presupposes that we share a concept of welfare which is rich and stable enough to support a high degree of preanalytic agreement on what is to count as faring well or badly. ... [A] theory of welfare ... will [also] tell us what the world must be like for [a particular conception of welfare] to apply to it, what must be the case in order for any welfare assessments to be true” (*WHE*, 7).

[1.1.2.] But: welfare differs from other metaphysical concepts (like causation, or substance, or numbers) in that “it may be doubted whether its nature really can be settled independently of assumptions about its value”, in which case: “What are appropriate standards for adjudicating among rival theories [of welfare]?” (*WHE*, 7-8).

[1.1.3.] One approach is a test of *normative adequacy* (*WHE*, 8-10); but this may result in a theory which is not *really* about ‘welfare’ (as we commonly understand and experience it) at all. This is the case “[b]ecause the notion of welfare already has a vernacular currency [so] it is not available as a term of art, to be defined in whatever way will best suit some favoured theoretical needs. ... Whatever theoretical role we may have in mind for welfare, we will need an independent test of descriptive adequacy for theories about its nature” (*WHE*, 10).

[1.1.4.] But: “[i]f such a test is necessary, then is it also sufficient? ... There is certainly this difference between the two procedures for selecting the best theory: whereas normative criteria could not in principle be self-standing, descriptive criteria might be” (*WHE*, 10).

1.2. Descriptive Adequacy (pp 10-20)

[1.2.1.] **Fidelity:** “The basic test is easy enough to state: the best theory about the nature of welfare is the one which is

most faithful to our ordinary concept and our ordinary experience. That experience is given by what we think or feel or know about well-being, both our own and that of others. The data which a candidate theory must fit, therefore, consist of the prodigious variety of our preanalytic convictions. We manifest these convictions whenever we judge that our lives are going well or badly, that pursuing some objective will be profitable or advantageous for us, that a change in our circumstances has left us better or worse off, that some policy would enhance or erode our quality of life, that some measure is necessary in order to protect the interest of our family or community, that a practice which is beneficial for us may be harmful to others, that we are enjoying a higher standard of living than our forebears, and so on. A theory of welfare tells us what the world must be like in order for such judgements to be correct; it offers us truth conditions for them. Its degree of fit with our ordinary experience will therefore be a function of the extent to which the truth conditions it offers can support and systematise our intuitive assessments. A theory about the nature of welfare is a proposed interpretation of our preanalytic convictions, and the best interpretation is the one which makes the best sense of those convictions” (*WHE*, 10-11).

[1.2.2.] “In applying [a] test [of ‘descriptive adequacy’] one of our principal assets is the richness of our welfare vocabulary. The concept of welfare is analytically connected to such cognate notions as well-being, interest, good, benefit, profit, advantage – and their various antonyms. A theory of welfare must preserve this network of analytic connections; if it fails to do so then it is an interpretation of some other concept. ... [A] theory has given us truth conditions for welfare assessments only if it is capable of supporting all of these cognate judgements” (*WHE*, 11).

[1.2.3.] “Our other main asset is our familiarity with welfare, which is more internal and immediate to our lives than many other items of philosophical interest. ... [But d]etermining the fit of a particular theory cannot be a matter merely of checking its implications one by one against our intuitions, registering its various successes and failures as they occur, then aggregating these piecemeal results into an overall score. Any such mechanical procedure will overlook important features of those very intuitions. Some of them, for instance, will inevitably be more central to our unreflective practices than others, in which case their preservation should surely count for more than the preservation of more marginal elements. The test of fidelity therefore needs to be informed by a map of our network of preanalytic convictions which distinguishes its core from its periphery. Our core beliefs about welfare are those we hold with the highest degree of confidence, or which best survive the process of challenge or reflection. One sign, therefore, that we have moved from core to periphery is that our responses become tentative or divided” (*WHE*, 11-12).

[1.2.4.] “... [W]e need not confine the test [of ‘descriptive adequacy’] merely to our intuitive judgements; we are also free to consider the role which welfare plays in common-sense psychology. We often explain conduct by ascribing motives which in one way or another take well-being as their object, motives such as generosity, compassion, friendship, malice, greed, and envy. Just as we can require a theory of welfare to support our ordinary assessments, so we can also require it to support our psychological explanations ... [modes of which] might turn out to be misguided or mistaken. But it is inconceivable that all of them should turn out this way, and any theory which yielded such a result would thereby reveal that it was a theory about something other than welfare” (*WHE*, 12).

[1.2.5.] Additionally, “that my own well-being is just one possible end of action among others for me – is a core element in our ordinary concept [of welfare]. A theory of welfare ... [that] implies that all intentional action is, or must be, self-interested ... has mistaken welfare for some other concept” (*WHE*, 13).

[1.2.6.] “Fidelity to our ordinary experience has now come to embrace both our pretheoretical beliefs about well-being, whether in our case or that of others, and also the role of these beliefs in our practical deliberations and our common-sense explanations. The further criteria of descriptive adequacy are all latent in, or derivative from, this basic test” (*WHE*, 13).

[1.2.7.] **Generality:** “... [A] theory of welfare ... should give us truth conditions for all of the different sorts of welfare assessments we make. ... No set of truth conditions for welfare assessments is complete unless it covers all ... categories of [such] judgement[s] – positive and negative, of fixed levels, and changes in level. Furthermore, it must also be able to sustain the rough discriminations of extent or degree which we make in all of them. A theory of welfare must therefore tell us what it is for one person to be much better or worse off than another, or for a particular benefit or harm to be only slight or negligible. We will discard as incomplete any theory which supports some of these modes of assessment,

but can make no sense of others” (*WHE*, 13-14).

[1.2.8.] These assessments concern a wide variety of ‘subjects’ (of lives), so that from “the paradigm case of adult human persons, our welfare vocabulary applies just as readily to children and infants, and to many non-human beings”; so, “[l]ike the basic test of fidelity, generality ... presupposes a rough initial distinction between the core and the periphery of our concept of welfare ... [so that] we can demand [from a theory of welfare] a rationale for delimiting the class of welfare subjects in a particular way and we can fault the proffered justification if it seems arbitrary or *ad hoc*. In addition, whatever its resolution of peripheral cases, a theory should offer some illuminating explanation of what makes them peripheral” (*WHE*, 15).

[1.2.9.] This concern (for principled resolution of peripheral cases) extends beyond *individual* welfare subjects to *collectivities*, as “... our talk about collective welfare does not appear to be merely figurative or metaphorical. ... At the same time, we may legitimately wonder whether the interest of a collectivity is distinct from the aggregate interests of its constituent members, thus whether collectivities qualify as welfare subjects in their own right. ... Since our preinterpretative intuitions do not seem to settle this reductionist issue one way or the other, we ... cannot impose a particular settlement as a condition of descriptive adequacy” (*WHE*, 15-16).

[1.2.10.] **Formality:** “A theory [of welfare] must not confuse the nature of well-being with its (direct or intrinsic) sources; it must offer us, not (merely) a list of sources, but an account of what qualifies something (anything) to appear on that list. ... [T]he nature of welfare is one thing, its several sources quite another” (*WHE*, 16). Moreover, “a theory of welfare, because it must be both formal and general, must abstract from ... contingencies of history, culture, and even biology. For it must offer a common, and comprehensive, answer to questions” of wherein consists the well-being of a wide range of creatures who may be *very* different in their natures (“based on the hypothesis that, however plural welfare may be at the level of its sources, like causation it is unitary at the level of its nature”) (*WHE*, 17). “What reason could we have to accept this hypothesis? It is weakly supported ... by the linguistic facts. When we distinguish, for instance, among physical, psychological, and emotional well-being, this suggests that there is something – well-being unmodified – of which these are the several kinds or aspects. Likewise, we certainly seem to move freely around the welfare terrain, speaking of the well-being of women and men, children and adults, human beings and animals, without any consciousness of ambiguity or equivocation” (*WHE*, 17).

[1.2.11.] **Neutrality:** “The [previous] two criteria of [descriptive] adequacy combine to yield a third. If a theory is to be general then it must apply to a wide range of creatures diverse in their natures, tastes, and forms of life. If it is also to be formal, then it must not consist merely in a list of goods considered to be indispensable to well-being, whether the list generalises across all of these creatures or applies only to some particular variety of them. But this means that an account of the nature of welfare must not have built into it any bias in favour of some particular goods or some preferred way of life. Any such bias will inevitably violate generality ...” (*WHE*, 17-18; my emphasis). “To impose this condition of neutrality [though] is not to deny that creatures of a certain nature will standardly flourish better under some forms of life than others. ... [A] favoured way of life [for creatures with a shared nature] must fall out as a confirming implication of a formally neutral theory; it must not be built in as one of its presuppositions” (*WHE*, 18).

[1.2.12.] These, then, are the “four cardinal virtues for a theory of welfare: fidelity, generality, formality, and neutrality. A descriptively adequate theory will be faithful to our ordinary assessments of well-being, including the role they play in our common-sense psychology, will cover all core cases and provide a principled resolution of peripheral cases, will not confuse welfare with its sources or ingredients, and will be free of distorting bias” (*WHE*, 18). “... [T]here is room for a theory to satisfy some criteria better than others ... [so that] deciding which failures are forgivable and which fatal” in picking the best theory “means that there will inevitably be much room for judgement, and undoubtedly for stipulation and construction as well” (*WHE*, 18-19).

[1.2.13.] The upshot of this is that “it now seems likely that [a theory of welfare] cannot be selected on purely descriptive grounds”, so that normative considerations must also play a role in its selection or construction; hence, “the theories concerning the nature and value of welfare which I shall be defending are therefore more interdependent than their linear order of exposition might suggest” (*WHE*, 19). However, “[n]ormative considerations properly come into play only around the edges of our ordinary concept [of welfare], in its disputed or unsettled regions. When the

constraints imposed by that concept run out, or when the evidence provided by our ordinary experience is indeterminate or inconsistent, then there is a time for shaping a theory of welfare to fit some favoured normative niche” (WHE, 19). But: “[s]ince welfare is a quality of our lives of which we have direct lived experience, the first question for any theory of welfare remains its fidelity to that experience. In this respect, then, questions about the nature of welfare still logically precede questions about its [ethical] value” (WHE, 19-20).

1.3. Dimensions of Value (pp 20-25)

[1.3.1.] A further aspect of ‘descriptive adequacy’ for a theory of welfare is that it “will tell us what it is for our lives to have, or to lack, a certain kind of value. Lives, however, are complex things whose value can be assessed along a number of different dimensions or from a number of different standpoints. Welfare represents only one of these dimensions – one way in which a life can be going well – and it is important to distinguish it as clearly as possible from the others” (WHE, 20).

[1.3.2.] “Welfare assessments concern what we may call the prudential value of a life, namely how well it is going *for the individual whose life it is*. This relativisation of prudential evaluation to the proprietor of the life in question is one of the deepest features of the language of welfare: however valuable something may be in itself, it can promote my well-being only if it is also good or beneficial *for me*. Since an account of the nature of welfare is descriptively adequate only if it is faithful to our ordinary concept, any serious contender must at least preserve the subject-relativity which is definitive of prudential evaluation. If it cannot manage this much then, though it might be a plausible rendering of some other dimension of value, it is not a theory about welfare at all. But surely we should be even more demanding. A theory of welfare should not only fit our ordinary concept, it should provide an interpretation of the principal features of that concept. Subject-relativity is a key ingredient in our concept of welfare, the feature which differentiates prudential value from the other modes of value applicable to lives. It is not too much to expect a theory to explicate this feature, to tell us what it means for my life to be going well not just in itself or from some other standpoint but *for me*, to explain how it is that lives can have this peculiar perspectival kind of value” (WHE, 20).

[1.3.3.] “While subject-relativity is an essential aspect of welfare, it is also somewhat elusive. The best way to provide a preliminary sense of what it is about is to contrast prudential value with some of the other standpoints from which lives can be evaluated” (WHE, 21). While there are too many modes of value to discuss individually, three are worthy of special mention (i.e. aesthetic value, perfectionist value, and ethical value).

[1.3.4.] *Aesthetic value*. “It is therefore one thing for your life to contain features which augment its aesthetic value and quite another for it to be going well for you. Of course, if you have taken on the project of living an aesthetically valuable life then we should expect your success or failure in that endeavour to bear on your well-being. But the same will hold for all of your other projects: while success in one’s projects is arguably an important source of welfare, scoring high in aesthetic value, just considered by itself, makes no independent contribution” (WHE, 23).

[1.3.5.] *Perfectionist value*. “To say that something has this sort of value is to say that it is a good instance or specimen of its kind, or that it exemplifies the excellences characteristic of its particular nature. ... Although [aesthetic and perfectionist] modes of value are often near kin (we commonly find excellence appealing and praise it in aesthetic terms), the perfection of a thing is clearly a different matter from its beauty or nobility: a slug or a piece of kitsch may be a paragon of its kind despite being ugly or vulgar” (WHE, 23). Evaluations of lives for perfectionist value stress the presence (or absence) of certain excellences in them, according to a nature determined for specimens of a species according to some ‘species-identity’. However, “[w]hatever we are to count as excellences for creatures of our nature, they will raise the perfectionist value of our lives regardless of the extent of their payoff for us. There is therefore no logical guarantee that the best human specimens will also be the best off, or that their underdeveloped rivals will not be faring better. Like aesthetic value, the perfectionist value of a life is conceptually independent of how well it is going for its owner” (WHE, 24).

[1.3.6.] *Ethical value*. “In the modern era ... the ethical domain has come to be understood [as] isolat[ing] those practical considerations which have to do with the impact of our choices on the lives of others ... [so that] the ethical value of a life is sharply distinct from its prudential value, since the welfare of others will be directly relevant to

determining the former but not the latter" (*WHE*, 24). "... [O]ne possible answer to [the question of 'What is the good?'] is: welfare. If this is the right answer, as welfarists think it is, then prudential value is the sole ultimate bearer of ethical value ... [b]ut neither question nor answer would even be intelligible were ethical and prudential value not conceptually distinct" (*WHE*, 25).

[1.3.7.] "... [T]he fact that a life has a high degree of one mode of value entails nothing whatever about [its] score along any other dimension. ... [P]rudential value is unique in determining how worthwhile a life is, or how well it is going, from the perspective of the individual whose life it is. ... Even the best constitutive account of a different dimension of value cannot be a descriptively adequate theory of welfare" (*WHE*, 25).

2. Welfare and Subjectivity (pp 26-44)

[2.1.] There are lots of models of the good life for human beings (as such models are a 'staple' of Western moral philosophising since the time of the ancient Greeks). "Out of this array our aim is to locate the account which works best specifically as a theory of prudential value, the one which is most faithful to our ordinary concept of welfare and our ordinary experience" (*WHE*, 26).

[2.2.] Models of the good life are (obviously) candidates for the best 'descriptively adequate' theory of welfare; but there are too many to review individually. Categories for such accounts are needed (to be determined by their *salience*, that is, how well they identify features of different theories which bear positively or negatively on their 'descriptive adequacy').

[2.3.] Sumner's "preferred scheme begins by sorting all of the candidates into two mutually exclusive and jointly exhaustive categories: the *subjective* and the *objective*. I claim salience for this line of division because I believe there to be an interpretation of the subjective/objective distinction such that subjectivity turns out to be a necessary condition of success in a theory of welfare. If I am right then objective theories can be ruled out of consideration as a category, all of them inadequate precisely because they are objective" (*WHE*, 26-27).

[2.4.] But "... the concept of the subjective is one of the most treacherous in the philosopher's lexicon" (*WHE*, 27) because what constitutes 'the subjective' "has been delineated by means of a number of features ... [so that] while each of these features applies to some subjective phenomena, none of them is essential to the subjective as such. What we need is a conception of subjectivity which is more primitive than [traditionally identified features of 'the subjective'], one which is capable of explaining their range of application and [subjectivity's] limitations" (*WHE*, 27).

2.1. Subjective and Objective (pp 27-34)

[2.1.1.] "Abstractly, the subjective is that which pertains to, or is characteristic of, subjects. What then is a subject?" (*WHE*, 27). Sumner discusses the phenomenology of subjectivity (*WHE*, 27-33), concluding that "[t]he realm of the subjective, in its philosophically primary sense, thus begins with experience, moves on to appearance, and ends with mind-dependent reality. It is the third stage of this progression which is germane to our purposes: if welfare is subjective then it is by virtue of being mind-dependent" (*WHE*, 33).

[2.1.2.] To illustrate the epistemological relationship between subjectivity so defined and the external world, Sumner uses the analogy of our perception of the colour 'red' as a 'secondary quality' of that colour's otherwise primary existence in the world outside our apprehension of it: "... to say that colour and sound are secondary qualities is to say that they are powers or dispositions on the part of objects with certain primary qualities to arouse certain experiences in certain circumstances on the part of creatures with a certain perceptual apparatus. The placeholder 'certain' in this general formula will need to be filled out in different ways for each allegedly secondary quality. ... The case of [the] colour ['red'] serves to illustrate that a subjective analysis of any natural property will need to furnish values for four distinct variables: (1) some objective characteristic of things by virtue of which they evoke (2) some mode of experience on the part of (3) some reference group of subjects under (4) some set of normal conditions. Where perceptual properties are concerned the objective characteristic will be physical, the reference group will be human beings with the standard sensory apparatus [for members of our species], the experience will be some mode of perception, and the

normal conditions will be those which ensure the absence of [perceptual] distortion. But the same four slots will be filled rather differently in the cases of those non-perceptual properties which have also seemed plausible candidates for subjective analysis, properties such as boring, tempting, sublime, and funny” (WHE, 34).

2.2. Values and Subjectivity (pp 34-41)

[2.2.1.] “The claim that some particular category of values is subjective ... in the strict and proper [primitive] sense of subjectivity [just elucidated] ... only means one thing: that, like colours, values of this kind are mind-dependent” (WHE, 34-35). “Values have been prime candidates for subjective analysis partly because of their apparent motivational force. ... [T]heir psychological source has generally been sought ... in some mode of affect. ... [W]e need a neutral category which abstracts from [the] differences [between rival subjective theories of welfare]” (WHE, 35). To this end, “[w]e will do better to ... speak instead of *attitudes*. ... Generally speaking, we may say that I have an attitude toward something when the thing matters to me, or I care about it, or it is an object of concern to me, or I mind it, or (in the more formal psychological terminology) it is valenced for me. My attitude is positive (what philosophers used to call a pro-attitude) if I favour the thing or am favourably disposed toward it, a negative (a con-attitude) if I view it unfavourably” (WHE, 35-36).

[2.2.2.] Therefore, “[a]s in the case of natural properties, a subjective analysis of some particular mode of value [like *prudential value*] will need to fill four slots: (1) some objective characteristic of things [such as *sources* of prudential value] by virtue of which they evoke (2) some attitude [following *assessment* of these *sources* against prudential *standards*] on the part of (3) some reference group of subjects [‘moral agents’ perhaps, equipped with more than just ‘sentience’] under (4) some set of normal conditions [such as those constitutive of ‘authenticity’ for Sumner’s authentic ‘happiness as life satisfaction’ theory of welfare]. There is much room for discretion in specifying all four variables” (WHE, 37). However, in “the special case of prudential value ... there seems only one plausible choice of authoritative subject. Since the prudential value of my life is its value *for me*, it seems reasonable to expect that the attitudes or inclinations which will figure in a constitutive account of my well-being will be mine. In the case of other allegedly subjective properties, such as colours, our reference point is likely to be normal human beings, or the set of creatures whose perceptual apparatus is recognisably like ours and in good working order, or whatever. Unlike colours, however, prudential ‘properties’ such as beneficial and harmful contain relativising indexicals. In order to preserve this relativity, it looks as though a subjective theory will need to connect my welfare with some psychological processes in me” (WHE, 37-38).

[2.2.3.] To this end, then, “... a subjective theory will map the polarity of welfare onto the polarity of attitudes, so that being well off will depend (in some way or other) on having a favourable attitude toward one’s life (or some of its ingredients), while being badly off will require being unfavourably disposed toward it. Likewise, something can make me better off on this sort of account only if I have (or would have under the appropriate circumstances) a positive attitude (of the appropriate sort) toward it. These formulae are deliberately indeterminate, leaving room for different subjective theories to fill them out in different ways. But even at this preliminary stage two cautions are in order. The first is that a theory is subjective if it treats my having a favourable attitude toward something as a necessary condition of the thing being beneficial for me. It need not also treat it as a sufficient condition The second caution is that this formula applies only to direct or immediate benefits: those whose contribution to my well-being is not dependent on their further consequences. It is a commonplace that some remote event, such as the death of a distant and wealthy relation, can be instrumentally beneficial to me without my regarding it favourably, or even being aware of its occurrence. A subjective theory tells us only that this cannot be true of the intrinsic sources of my well-being” (WHE, 38).

[2.2.4.] “Subjective theories make our well-being logically dependent on our attitudes of favour and disfavour. Objective theories deny this dependency ... [they] do not merely deny the sufficiency of a reference to my attitudes in an analysis of my well-being ... [instead, the] crucial differentiating question is the necessity of such a reference. To this question subjective and objective theories give contradictory answers; the two categories are therefore both mutually exclusive and jointly exhaustive” (WHE, 38-39). This duality is useful because “... it will locate ... differences [between competing subjective theories] within the class of subjective theories, rather than between them and objective theories. Every candidate theory will either make welfare mind-dependent (to some degree and in some respect) or it will not. ...

[I]n the case of welfare the final defence of a sharp subjective/objective duality must be its salience [which requires demonstration]" (WHE, 39-40).

[2.2.5.] "Before setting out to construct a conception of subjectivity around the notion of a subject, I claimed that the result would be the philosophically primary sense of this multiply ambiguous expression. We are now in a position to appreciate wherein its primacy lies: it both underlies and explains all the further marks of the subjective. The secondary conceptions of subjectivity are all generated by the same process of over-generalisation: we begin by selecting some characteristic which belongs only to certain specific mental processes and then attribute it, first to the mental as such, and then to all mind-dependent features of the world" (WHE, 40). This means that the 'mind-dependency' of well-being in the 'philosophically primary' sense of 'subjectivity' need not give rise to a concern that 'subjective' assessments of the prudential value of lives are somehow 'unreal': "[i]f colour is a subjective property then its existence is dependent on our psychological processes. However, since we have no reason to think that these processes are themselves less real than physical ones, the subjectivity of colour gives us no reason to deny its reality. The same obviously holds for welfare. To claim that welfare is subjective is to claim that it is mind-dependent – that and nothing more. It is not to suggest that it is in any legitimate sense unreal" (WHE, 41).

[2.2.6.] "My explication of the subjective/objective distinction for theories of welfare differs from similar accounts elsewhere chiefly in its reliance on the very plastic notion of an attitude, which makes it possible to generalise over the many varieties of subjective theory. As it stands the line of division is still very rough and crude, but we will be able to add the necessary further refinements as we address particular instances of theories on both sides of it. First, however, we need some reason for thinking that it provides the right grid to be working with" (WHE, 41).

2.3. The Case Against Objectivism (pp 42-44)

[2.3.1.] "Defending the [subjective/objective] distinction as a working tool requires an argument showing that either subjectivity or objectivity is in itself a *sine qua non* in a theory of welfare. ... There are many ways in which a life can be a good life [from *outside* that life]; being high in prudential value is only one of them. What distinguishes welfare from all other modes of value is its reference to the proprietor of the life in question: although your life may be going well in many respects, it is prudentially valuable [to you as that life's proprietor] only if it is going well *for you*. This subject-relativity is an essential feature of our ordinary concept of welfare. ... Among the modes of value which can belong to individual lives, welfare stands out by virtue of incorporating an internal reference to its bearer" (WHE, 42).

[2.3.2.] "What is crucial [on a subjective account of welfare] is that you are the proprietor or manager of a set of attitudes, both positive and negative, toward the conditions of your life. It is these attitudes which constitute the standpoint from which these conditions can be assessed as good or bad *for you* [including under 'second-person' assessments of welfare]. It follows on this sort of account that a welfare subject in the merely grammatical sense – an individual with a distinct welfare – must also be a subject in a more robust sense – the locus of a reasonably unified and continuous mental life. Prudential value is therefore perspectival because it literally takes the point of view of the subject. Welfare is subject-relative *because* it is subjective" (WHE, 42-43; my emphasis on the word 'because').

[2.3.3.] This contention is however neither trivial nor analytic: "[a]ll we are given by the concept of prudential value is its characteristically positional or perspectival character; the claim that this is rooted in its subjectivity is a substantive thesis, analogous to the equally substantive thesis that causation is a matter of regular succession" (WHE, 43). This is because it is easily shown that systems of spatial ordering with subject-referential properties can also be constructed around objective characteristics (like the difference between subject-relative positional directions or coordinates, and directions or coordinates relying on positional data instantiated in terms of the Earth's poles).

[2.3.4.] "The thesis that welfare is subjective is therefore not merely a reaffirmation of the fact that it is subject-relative; instead, it is a putative interpretation or explanation of this fact. ... [I]t is clear from the outset what sort of account subjective theories have to offer. Since objective theories exclude all reference to the subject's attitudes or concerns, they will have to supply some alternative account. It is for this reason that the subject-relativity of welfare constitutes a deep problem for any objective theory. Where theories of welfare are concerned, objectivity appears to impede descriptive adequacy" (WHE, 43-44). That it is so easy to make a general case against objective theories of welfare

prima facie bears out the salience of the subjective/objective distinction (although, at this stage of the inquiry, we don't know yet whether one or more objective theories of welfare will be able to meet the 'challenge' posed by the substantive thesis that welfare is subjective in its essential *nature*).

3. Objective Theories (pp 45-80)

[3.1.] Rather than a theory of welfare's *nature*, all that most objective theories give us is a list of its various (intrinsic) *sources*, "[w]hich is why Derek Parfit refers to this [theoretical] option as the Objective List Theory (Parfit 1984, 4). But a list of virtues is not a theory of virtue, and a list of (human) goods is not a theory of welfare" (*WHE*, ch. 3, fn. 1). An example of this appears in the work of John Finnis, and David Brink: "... see Brink 1989, ch. 8, where David Brink, after rejecting the main forms of subjective theory, promises to sketch a plausible objective theory but instead merely goes on to enumerate what he regards as 'the primary components of valuable lives' (231). Since these components ('reflective pursuit and realisation of agents' reasonable projects and certain personal and social relationships') are likely to turn up on anyone's roster of the (intrinsic) sources of (human) well-being, including the subjectivist's, we still require a formal account of what it is, on an objective theory, for something to be such a source. At the level of lists, subjective and objective theories are indistinguishable" (*WHE*, ch. 3, fn. 2).

[3.2.] Sumner is aware of only two strategies employed by objectivists for meeting the 'challenge' posed by the essential subjectivity of welfare: the approach inherent in the *private ownership theory*, and the approach inherent in the *teleological theory*. (According to Sumner, the former is straightforward and easy to comprehend, but the latter requires some explication of 'needs', and 'functionings and capabilities', to understand).

3.1. The Private Ownership Theory (pp 46-53)

[3.1.1.] Its principal exponent was G.E. Moore. "... [I]n the course of attempting to refute ethical egoism Moore was led to provide an analysis of the notion of individual welfare. ... Moore's analysis reduces prudential value to two primitive elements: some dimension or other of intrinsic value and the relation of possession or ownership. ... On Moore's account my good consists of my ownership of something which is itself good [and therefore 'objectively' intrinsically valuable]" (*WHE*, 47).

[3.1.2.] In relation to 'objective intrinsic value': "[b]ut which mode of value did [Moore] think it was? To say that something is intrinsically valuable is to say that it is valuable in itself or for its own sake; the contrast is with instrumental value, which a thing has by virtue of leading to or bringing about something else which is valuable in its own right. This distinction between the ultimate bearers of value and their causal antecedents cuts across the four evaluative dimensions identified in the previous chapter; there seems every reason to think that prudential, aesthetic, perfectionist, and ethical value can all be possessed by things either intrinsically or instrumentally. Intrinsic value is therefore not itself a distinct dimension of value ..." (*WHE*, 48).

[3.1.3.] Instead, "[i]t is clear ... that the concept of intrinsic value which Moore had in mind is ethical: it is that category of final or ultimate value which we have a moral reason to produce, or whose production lends our actions moral value. [Moore's] analysis of prudential value thus reduces it to ethical value (plus private ownership)" (*WHE*, 48). "However, from the mere fact that some state of affairs is intrinsically good it plainly does not follow that it is good for me; there may be no connection at all, or no connection of the appropriate sort, between it and me. ... Ownership of [some] state by me – the fact that it is *mine* – is meant to supply the needed connection ... [to] the subject-relativity of welfare: in order for some intrinsically valuable state of affairs to be *for* a subject, it must be a state *of* that subject" (*WHE*, 49).

[3.1.4.] "The private ownership theory [postulates] that various states of the world have a special kind of intrinsic ethical value, and then identifies the sources of an individual's well-being with the subset of these valuable states which are in her vicinity or on her territory. ... According to the private ownership theory, the fact that [a] state [of the world] has intrinsic ethical value is both necessary and sufficient to make that individual better off. ... In that case, the theory entails that beautiful people are necessarily better off, at least in one respect, than plain or ugly ones. ... [It] will also entail that every beautiful object has its own welfare, since it will be true of each such object that an intrinsically valuable state is a state of it. This would be a spectacular failure of the test of generality; however vague the outer

boundary of the class of welfare subjects may be, mountains and sunsets and string quartets do not belong even to its periphery" (*WHE*, 49-50).

[3.1.5.] Moore tries to solve the problem of the proliferation of welfare subjects in his account by stipulating that consciousness is "an indispensable ingredient of any intrinsically [ethically] valuable state. Now if the list of intrinsic[ally ethically valuable] goods is restricted to states of consciousness, or to organic wholes which include such states, then it will follow that these goods can belong only to individuals with mental lives. Since it is just these individuals who qualify as paradigmatic welfare subjects, the private ownership theory will thereby be saved from grossly overpopulating the domain of such subjects" (*WHE*, 50).

[3.1.6.] But, in doing this, Moore reduces prudential value to ethical value (when the explanatory relation is more likely to run in the opposite direction, which is more "intuitively plausible ... [as] the fact that something makes our lives go better might be thought to give us an ethical reason for bringing it about" (*WHE*, 51). Vis-à-vis whether 'deep personal relationships' have prudential value, the private ownership theory "reaches an affirmative answer to this question by first establishing that these relationships have ethical value, and then concluding that they must therefore make our lives go better. ... It is believable that intimate connections with others make the world a better place by virtue of enriching our lives, but it is not believable that they enrich our lives by virtue of making the world a better place" (*WHE*, 51).

[3.1.7.] Moreover, we engage in a wide variety of everyday activities that have prudential value (for us), but little or no ethical value (for the world as a whole), so that "having [ethical] value seems not to be necessary in order for some state or activity of ours to contribute (directly) to our welfare" (*WHE*, 52). A way around this is to posit that it's not the state or activity that has ethical value, but the feelings or attitudes they elicit in us; which is more plausible, but still reverses the direction of explanation because "we have a moral reason for supporting or promoting these pursuits because they improve the quality of our lives, not the other way around" (*WHE*, 52). By reducing prudential to ethical value (when we can plainly see the two modes of value are conceptually and practically distinct), the private ownership theory fails to capture the subject-relativity of welfare because intrinsic ethical value is not subject-relative, as it exists in states of the world *independently* of how welfare subjects experience them. So, "[a]lthough subject-relativity is internal to the notion of possession, it remains external to the ethical standpoint. Private ownership is incapable of bridging the conceptual gap between ethical and prudential value" (*WHE*, 53).

3.2. Needs (pp 53-60)

[3.2.1.] "The private ownership theory is no more than an isolated episode, an idiosyncratic curiosity, in the history of objective accounts. The mainstream of the tradition has tended to rely on one or another conceptual resource which, while appropriately objective, seems intimately connected with well-being. A perennial favourite is the notion of a basic need" (*WHE*, 53). Needs satisfaction does not require intentionality (so the relevance of needs in an account of welfare implies no dependence on mental states on the part of welfare subjects). But it's also true that needs are determined by or contingent upon our projects, and other circumstances of our lives, so that "[w]hat the [welfare] objectivist (derivatively) needs is a set of needs which are not derivative" (*WHE*, 54), which is typically achieved by focusing on needs which are basic or fundamental. However, this shift from the range of needs dictated by someone's projects or other life circumstances to 'basic' needs involves a loss of generality, because "the [basic] needs which seem most hospitable to an objectivist treatment cannot be the whole story about well-being", although "they [also] cannot fail to be part of the [objectivist's] story [about what welfare is]" (*WHE*, 54). So: what are 'basic needs' anyway?

[3.2.2.] A *constructivist* (like David Braybrooke) identifies them via a consensus achieved amongst "some favoured group of agents" (*WHE*, 55), so that resultant lists of 'basic needs' are social constructs. But welfare objectivists can't rely on such a methodology because "[i]n order for basic needs to be objective their explication must include no reference to people's attitudes, including their preferences or choices. ... If an objective account of needs is possible, it clearly must eschew [constructivism] in favour of the *naturalist* route" (*WHE*, 56; my emphasis).

[3.2.3.] "A naturalist account will ground basic needs, somehow or other, in the nature of their subjects. ... While naturalism requires connecting our needs to some aspect of our nature, it is neutral as to which aspect this is to be"

(*WHE*, 56-57), so that ‘basic needs’ can instead be naturalistically defined as deeply subjective. “A particularly short route from needs to subjectivity can be traced in John Rawls’s notion of a primary good” (*WHE*, 57). Rawls’s theory concedes that not everyone will ‘need’ the same *amount* and *kinds* of ‘primary goods’ to achieve well-being, so is subjective in those terms.

[3.2.4.] Garrett Thomson’s naturalistic account of needs is no less subjective than the constructivist approach of Braybrooke, because Thomson argues that lack of access to what we need engenders harm (as the reason that certain ‘primary goods’ are basic in the first place is that they serve our *interests* as human beings, the range and type of which will of course often differ between *different* human beings *because* human beings are unique).

[3.2.5.] Sumner’s point in talking about ‘basic needs’ is to show that “[w]hile the concept of a need may be superficially objective, and while it may yet turn out to be the objectivist’s most valuable resource, its freedom from contamination by subjectivity cannot be taken for granted” (*WHE*, 59-60).

3.3. Functionings and Capabilities (pp 60-68)

[3.3.1.] “In several influential works Amartya Sen has sketched a model of well-being which, like accounts which invoke the concept of a need, has a decidedly objective look to it” (*WHE*, 60). Sen rejects the *utilitarian* account of welfare (as felt satisfaction, preference, or choice) because these can be the product of social conditioning (and not reflective of what a person might actually want in/for their lives); in any case, liking or wanting something is not the same as (prudentially) valuing it. Sen also rejects welfare accounts structured around possession of or access to *commodities*, as such accounts confuse *welfare* (as some state of being in the world) with its various *sources* (as the prudential value of commodities depends on what they do or enable *for* people, or rather, on what people can do *with* these commodities).

[3.3.2.] “While Sen faults the first family of theories for their subjectivity, he rejects the second for carrying objectivity too far. His ostensible aim is to construct an account which mediates between these extremes. The core notion in [Sen’s] account is that of a functioning, which Sen defines as anything which a person manages to do or to be. Functionings are therefore individual achievements or successes The ground floor of Sen’s account treats an individual’s well-being as a matter of his functionings. But Sen then adds a second level to the analysis. A capability is a freedom or opportunity to achieve a certain functioning Although capabilities may be prudentially valuable chiefly for the functionings which they make available, Sen argues that they also have a certain value in their own right: we are better off for having avenues open to us which we never actually choose to pursue. Our level of well-being is therefore determined both by our set of functionings and by our set of capabilities” (*WHE*, 60-61).

[3.3.3.] “When we ask what Sen has in mind as individual functionings, his list includes such familiar items as healthfulness, longevity, literacy and the like. For each of these items Sen recognises a distinction between the functioning itself and its various social indicators ... aiming to measure [typically] quite objective features of individuals’ lives: whether they are adequately nourished or clothed or sheltered, whether they can read or write or count, and so on. Since all of Sen’s functionings appear to be objective, it would seem to follow that his theory of welfare is objective as well” (*WHE*, 62).

[3.3.4.] However, to address differences in needs and priorities between different people (as differences in the subject-relative *prudential value* of various standard *sources* of well-being), Sen’s list of functionings is actually to be populated for *each* individual (elicited from them by means of self-evaluation techniques), and so is subjective in this sense. “... Sen’s introduction of personal valuations is meant to impose some order on the virtually infinite range of achievable human functionings ... [and also] converts a list of the (standard) sources of (human) well-being into a formal account of its nature. Instead of consisting in some stipulated set of functionings (an account which would vary for different welfare subjects), well-being now comes to consist in achieving whatever functionings an individual most values (an account which is common to all such subjects). If Sen’s aim was to develop a descriptively adequate theory about the nature of well-being then it is easy to see why he assigned a crucial role to individual valuation. But this way of both formalising and generalising his theory necessitated also subjectivising it” (*WHE*, 65-66).

[3.3.5.] This opens up Sen's theory of welfare to the problem of social conditioning affecting personal valuation ('the autonomy problem' that Sen himself poses): "Sen regards valuing as a more cognitive and reflective activity than enjoying or desiring. However this may be, personal values are also notoriously subject to influence by accustomed social conditions. If there is a problem here for theories which interpret welfare in terms of felt satisfaction or preference, there is equally a problem for a theory which assigns the same constitutive role to [personal] valuation" (*WHE*, 66).

[3.3.6.] Sen is also (deliberately it seems) unclear on the 'foundational question' of how the 'personal values' used to prioritise functionings originate (are they just rankings in a personal value scheme that might be a product more of social conditioning than individual choice? or do they acknowledge and reflect the urgency and importance of particular functionings in the lives of individuals?), so that his theory of welfare is 'incomplete' (because it 'falls foul' of the 'generality' criterion of 'descriptive adequacy'). There are "two importantly different ways in which Sen's theory of welfare could be completed. Should valuations be understood subjectively, then he will have no objective theory of welfare to offer. On the other hand, should the objective route be pursued, then personal rankings would play a merely evidentiary role in his account. In that case ... Sen would owe us an 'objective normative account of human functioning' as the foundation of his theory of welfare. The most likely source of such an account [would be the teleological theory of welfare]" (*WHE*, 67).

[3.3.7.] Sumner concludes his review of Sen's 'theory' of welfare by acknowledging that Sen may not *actually* be offering one. "This may explain why Sen is content with, even proud of, the incompleteness in his account concerning the character of the underlying valuations needed to pick out relevant or important functionings. If both subjective and objective interpretations of this valuational exercise seem likely to yield roughly the same list of functionings, then anyone whose interest lies primarily in social analysis need inquire no further. It is the peculiar burden of the philosopher to remain unsatisfied until the foundational issue has been resolved" (*WHE*, ch. 3, fn. 52). "As the materials for a normatively attractive account of the standard of living, the basic functionings and capabilities on which [Sen] focuses have excellent credentials. Objectivity is no liability in an enumeration of the principal sources of well-being, only in a theory about its nature" (*WHE*, 68).

3.4. The Teleological Theory (pp 69-80)

[3.4.1.] "It is clear now what an objective theory of welfare requires: some aspect of the nature of welfare subjects, other than their subjectivity, which can supply the standpoint or perspective characteristic of prudential value. And so we come finally to the second main strategy available to the objectivist, which attempts to supply this vital ingredient" (*WHE*, 69) by recourse to the 'telos' or 'end' or 'purpose' of human beings (and other welfare subjects besides). This 'telos' (for thinkers such as Aristotle, Attfeld and Taylor) derives from an organism's (biological) *function*: which may be, for example, to preserve its *existence* by protecting and promoting whatever is constitutive of its *well-being* (ignoring the circularity inherent in a descriptive statement masquerading as an explanatory statement).

[3.4.2.] "For Aristotle, therefore, a thing's excellence is conceptually connected to its function and thereby to its welfare. Simplifying somewhat, the argument to his formal theory of welfare contains the following steps: (1) something promotes my well-being just in case it enhances my distinctive function, (2) something enhances my distinctive function just in case it expresses my distinctive excellence, therefore (3) something promotes my well-being just in case it expresses my distinctive excellence. Ever since Aristotle first articulated it, however, this objectivist strategy has attracted as many critics as supporters. One classic line of resistance attacks the crucial claim that human beings are among the kinds of things which have a characteristic function" (*WHE*, 71).

[3.4.3.] Attfeld and Taylor address this criticism by pointing out that while we may not know the (ultimate) function (or purpose) of an organism, we know when it's functioning *well* (or *badly*) by standards appropriate to its kind. Following from this, "[t]he philosophical lesson here seems to be that we need not, and should not, expel teleology altogether from our biology. ... To say that teleological notions, like that of a function, are derivative or secondary in biological contexts is not to say that that they are inherently suspect" (*WHE*, 73).

[3.4.4.] Accounts like those of Attfield and Taylor react against the assumption that only human beings have a good to be promoted, by endorsing the subject-relativity of welfare, then arguing that having a distinctive evaluative standpoint does not require subjectivity (only a 'species-identity' for a species conceptualised as having a good to be regarded or promoted), then concluding that membership of a species according to some 'species-identity' is enough to determine whether things or experiences will be 'good for' *any* member of that species. While such accounts have intuitive appeal (because they make sense of much of our welfare vocabulary around the intrinsic value of a wide range of living things), they provide "us with no unequivocal direction when we are trying to decide whether to include or exclude peripheral subjects [of-a-life]. In this domain the best result is the one which follows from the best theory" (*WHE*, 75). So, while the teleological theory tells us that something has a welfare if it has a ('natural') good of its own (its own 'function' able to be promoted or impeded by various alterable states of the world), how far does welfare subjectivity on such accounts extend? Are artefacts welfare subjects? What about features of the natural wilderness or landscape? What of organisms whose evolutionary course has been diverted by human beings for their own ends, so that their 'natural' function has been 'contaminated' by our purposes? Do they still have a 'good' given that they have not evolved entirely 'naturally'? And what if human beings are creatures of some deity, created for *their* purposes; in such event do we even have our own independent 'good'? (as creation may be 'natural' for a deity, but then our function is derivative of the deity's function, and so not really 'ours').

[3.4.5.] In any case, such (objectivist, teleological) accounts of welfare work at least as well for bodily *organs* as for organisms: are organs to be regarded as 'welfare subjects'?

[3.4.6.] "These problems with the class of welfare subjects are ... mere embarrassments for the teleological theory. ... [W]e have not yet come to terms with the teleological theory's principal asset ... the notion of functioning [with which that theory claims it can explain the perspectival nature of prudential value]" (*WHE*, 77). "Thus far we have highlighted some of the pitfalls in the notion of a biological function. But the real problem with the teleological theory lies not in the coherence of this notion but in its relevance to the nature of welfare. ... A theory of welfare must be a theory about the nature of prudential value. ... Both Attfield's formula and Aristotle's argument equate a creature's welfare with its distinctive excellence [in terms of the achievement of its biological function]. But then both accounts conflate prudential and perfectionist value: they are really theories about the latter rather than the former" (*WHE*, 78).

[3.4.7.] "The[se] two evaluative dimensions come apart when we are dealing with subjects in the strict sense [rather than welfare subjects in the sense meant by the teleological theory], namely those with a subjective point of view. ... And they diverge most clearly for paradigm subjects such as us. Where human agents are concerned, it is a contingent matter whether the possession of some particular excellence makes us better off. There may, of course, be a strong empirical correlation between the excellences of mind and body and the well-being of their owners; it would be surprising if there were not. But as a conceptual matter the inference for any agent from perfectionist to prudential value is never guaranteed; there is always a logically open question. The gap between the two is opened by the agent's own hierarchy of projects and concerns, which is but one manifestation of her subjectivity" (*WHE*, 79). "In that case ... the teleological theory is fundamentally misconceived as a theory about the nature of welfare; it is really about something quite different. ... Our result concerns not the *value* of welfare but its *nature*. The teleological theory was the objectivist's last, best hope for responding to the challenge [of the essential subjectivity of welfare, because of the subject-relativity of prudential value, and therefore, prudential valuation]. With its demise goes all prospect of constructing a descriptively adequate objective theory of welfare" (*WHE*, 80; my emphases).

4. Hedonism (pp 81-112)

[4.1.] With the elimination of objective accounts, "we have merely succeeded in narrowing our original question a little: we now wish to know which subjective theory gives us the best account of the nature of welfare. Unluckily, however, not only are there many such theories available, they also differ among themselves in significant ways. We have made progress thus far by utilising a relatively simple subjective/objective grid; we now need some equally salient way of subdividing the category of subjective accounts" (*WHE*, 81).

[4.2.] "Philosophers confronted by this problem ... distinguish theories which make welfare a *state of mind* from those which make it a *state of the world* ... [a distinction which] does turn out to have a point. A theory is subjective if it

makes welfare depend *at least in part* on some mental state, but it may make it depend on something else as well. If we draw a boundary around (the subject's own) states of mind and call everything else a state of the world, then we can indeed generate two types of subjective theory: those on which my welfare is *solely* a matter of my states of mind and those on which it is *additionally* a matter of some states of the world. The ground for this distinction ... has been laid in our own account of the subjective, in which we worked outward from states of mind to states of the world. On this picture, then, 'state-of-mind' theories treat welfare just as some mental state, or some combination of such states, while 'state-of-the-world' theories bring into the picture some reference to mind-independent reality" (WHE, 82).

4.1. The Classical View (pp 83-92)

[4.1.1.] Hedonism "may be a theory about the nature of welfare. Such a theory will begin by taking pleasure and pain to characterise the primitive attitudes, positive and negative respectively, which subjects may have towards objects or states of affairs. Minding something, or caring whether it happens, will ultimately be a matter of finding the experience of it either pleasant or painful. The theory will then go on to assign these attitudes a basic constitutive role in a formal theory of welfare. Such a theory will map the polarity of welfare onto the polarity of pleasure and pain. In order for my life to be going well for me I must be experiencing it, or its principal ingredients, as pleasant or satisfying; conversely, if it is going badly for me then my experience of it must, on balance, be unpleasant or unsatisfying. Likewise, something can benefit me or make me better off only if I find it agreeable, and it can harm me or make me worse off only if I find it disagreeable" (WHE, 84).

[4.1.2.] Utilitarianism (as a moral theory) is underpinned by one such instance of a hedonistic theory of welfare. "The equation of well-being with happiness is implicit in the utilitarian tradition – too implicit [however] to count as a developed theory about the nature of welfare. Instead, it is an assumed conceptual identity: for the utilitarians the two notions were indistinguishable" (WHE, 85). "... [E]veryone knows that the classical utilitarians were hedonists both in their theory of the good and in their theory of welfare. ... [W]hat is important for our present purposes is that the utilitarians all shared the view that welfare consists in happiness and happiness consists in pleasure and the absence of pain" (WHE, 87).

[4.1.3.] Sumner associates utilitarianism with a mental state conception of welfare, for which utilitarians provided two models of pleasure and pain, "one which appeals to the internal qualities of these feelings while the other invokes their external relations. On both models pleasures and pains each constitute a class of distinctive feelings or experiences whose common properties can be identified by introspection" (WHE, 87). These two views on the nature of pleasure and pain differ, so that they can be characterised in terms of their differences as "the internalist *sensation* model with its emphasis on a homogeneous (positive or negative) feeling tone and the externalist *attitude* model with its reliance on a uniform (positive or negative) reaction. ... On both views pleasures and pains are experiences which can be identified as such on the basis of some introspectible feature, whether this is an internal quality (the way they feel) or an external relation (the fact of being liked or disliked). On both views, therefore, pleasures and pains are purely mental states" (WHE, 91).

[4.1.4.] Sumner concludes this section by stating that "[a]ny hedonistic analysis of welfare will need to endorse some version or other of [the theses that (1) welfare is the same as happiness and (2) happiness is pleasure in the absence of pain]" (WHE, 91-92), and that hedonistic theories of welfare locate our responses to pleasure and pain within mental states (thus making them 'mental state theories' of welfare). "Our next question is whether any such theory can be descriptively adequate" (WHE, 92).

4.2. Problems with the Classical View (pp 92-98)

[4.2.1.] "The two strongest objections to an account [like hedonism's] have been nicely summarised by James Griffin. The first is directed against the sensation model. Griffin argues that there is no positive quality of feeling such that having more of it invariably makes us better off, and no negative quality such that having more of that must make us worse off. ... Suppose, then, that all pleasures do share some uniform positive feeling tone, which can be present to a greater or lesser degree. A hedonistic theory will imply that the best choice for you is always the one which maximises your personal share of this feeling. But what if this does not match your own ranking of the available options? ... The

theory is then committed to disregarding your own priorities in determining what is best for you. ... A theory which stipulates that welfare consists in some distinctive feeling, regardless of the place which this feeling is actually assigned in the lives of reflective subjects, is only doubtfully subjective. ... What makes a theory subjective ... [is] the role it assigns to the subject's concerns in identifying those sources" (*WHE*, 92-93).

[4.2.2.] "Against [the attitude] form of hedonism ... Griffin brings a second objection" (*WHE*, 94), which is that welfare consists in not just 'illusory' experiences (and the mental states they may induce), but experiences which are 'veridical' (i.e. have *actually* happened, or reflect some *actual* state of the world). This is a version of Nozick's 'experience machine' objection to mental state theories of welfare (which is that, if all we care about are pleasurable mental states, then it doesn't matter if the stimuli inducing those pleasurable mental states are not 'real', or 'veridical' i.e. reflective of states of the world as the world *really is*; *WHE*, 94-95).

[4.2.3.] In practice, "[o]ur attitudes towards delusion, deception, and fantasy are more complex and ambiguous than the experience machine [objection] allows for. However, we will surely agree with Griffin in preferring reality [over] illusion [in] some key sectors of our lives. And we will agree that, in those sectors, being in touch with reality makes for a better life. ... If what you have accepted as an important constituent of your well-being – your achievements, say, or the feelings of others about you – turns out to have been an elaborate deception, you are likely to feel hurt and betrayed. How else to explain this, except to say that, in this area at least, what mattered to you was not merely how things seemed but how they actually were? Your reaction to the deception certainly looks, and feels, like a reassessment, in the light of your own priorities, of how well your life has been going. ... Since a subjective theory must be faithful to the full range of our concerns, and since these concerns typically extend beyond appearance to reality, such a theory cannot make welfare consist merely in having agreeable experiences, even of the eclectic variety that the utilitarians were prepared to recognise" (*WHE*, 96-98).

[4.2.4.] Therefore, *both* variants of classical hedonism fail because *each* overrides the authority of welfare subjects to determine for themselves wherein consists their good, either by "stipulating that subjects must always prefer more pleasurable feeling tone to less", or by "dictating that subjects must be indifferent between veridical and illusory experiences, as long as they are equally enjoyable. In the end, therefore, it does not matter which model the classical hedonists plug into their analysis of welfare. In either case the result will fail to preserve the individual [prudential] autonomy which is the most attractive feature of a subjective theory" (*WHE*, 98).

4.3. The Truth in Hedonism (pp 98-112)

[4.3.1.] Classical hedonism seems to tell (more or less) the right story about the place of pleasure and pain in our lives. But: not all pain or suffering is *physical* (i.e. responding to some threat to tissue integrity), as we can feel 'pain' at the loss of a loved one, or the failure of some important project. Sumner explores the phenomenology of pain (*WHE*, 99-103), noting the following along the way.

[4.3.2.] *Pain makes us feel bad.* "In the central cases of physical pain, then, it appears that at least part of what is bad about [the] condition is the way it makes us feel. ... If I am suffering physical pain then I can be quite wrong about the organic cause of my affliction, or even about whether it has one, without that error diminishing in the slightest either the reality of my pain or its impact on the quality of my life. Were the instance of physical pain in every respect typical, then there would be nothing wrong with a mental state theory of welfare" (*WHE*, 100).

[4.3.3.] *But: pain isn't always necessarily bad for us.* Sumner asks: when is pain not bad for us? It appears that there is an attitudinal or dispositional dimension to the experience of pain (so that, in some cases, lobotomy patients or those with battlefield injuries can endure levels of pain that would ordinarily be unbearable). This has "led some to deny that pain is a sensation at all, giving it instead a purely attitudinal analysis. ... [Yet the] similarity with our linguistic resources for characterising, say, sensations of taste or touch lends support to the idea that there is at least a sensory dimension to pain. At the same time, it also seems possible to distinguish between all of these properties, including intensity, and the extent to which we mind the pain or are bothered by it" (*WHE*, 101). But it's also true that in some situations (such as those associated with extreme, goal-oriented exertions, like running a marathon, or childbirth) pain is a welcome sensation (as it's reflective of *progress*, and, ultimately, *achievement*).

[4.3.4.] *So: it's the nature of our responses to pain that make pain bad for us. "The seeming (logical and psychological) separability of pain sensations themselves from our emotional or attitudinal responses to them forces a choice on us when we are trying to characterise the nature of pain" (WHE, 102). We can stick with the sensation model, or we can switch to the attitude model. But the attitude model can make no sense of 'phantom' pains, nor can it deal with physical feelings which are not in and of themselves 'painful' (i.e. nausea, hiccups, sneezing, dizziness, etc.) but to which we are normally averse (because, if these feelings persist and are severe enough, they can make our lives unpleasant, even miserable). "While none of these is quite the same as physical pain, we experience each as intrinsically disagreeable. The attitude model simply obliterates these categorical boundaries by treating all these states indifferently as pain. While there is something intuitively right about this approach (to which we will return), it is hopeless as a phenomenologically accurate picture of the nature of pain" (WHE, 103).*

[4.3.5.] *Which makes explaining the nature of our contingent or variable responses to pain difficult. There is a third, hybrid, option, which combines the sensation and attitude models into a complex whole. "However, if this account is to avoid the problem of over-breadth then it must somehow distinguish between pain and other subjective experiences which evoke aversive responses. And how is it to do this, except by saying that pain comprises those experiences which feel painful, or which hurt? And isn't that tantamount to conceding that the sensation model was right all along in identifying the pain proper by means of its purely phenomenal properties?" (WHE, 103).*

[4.3.6.] Eric Cassell advances a distinction between 'pain' and 'suffering', where "[s]uffering ... is a response of the whole person, which takes into account both the subjective experience itself (in the narrow sense) and its meaning or significance. It follows that episodes of pain which are intrinsically indistinguishable (being of the same kind, having the same intensity, and lasting for the same duration) may cause quite different degrees of suffering to different subjects, or to the same subject at different times. It also follows that conditions other than pain may cause suffering. ... In Cassell's conceptual framework, suffering results from 'injuries to the integrity of the person'" (WHE, 103-104).

[4.3.7.] Cassell's distinction has two important features: firstly, the range of the notion of suffering is much broader than for pain. So "Cassell's substantive point is that it is the elimination of suffering, construed as he does in terms of the individual's aversive reaction to her conditions, which deserves to be the central goal of medicine. We can express this as follows: it is suffering, in Cassell's sense, that is central to an account of happiness and well-being (or, rather, unhappiness and ill-being)" (WHE, 105). And secondly: "[n]o mere sensation is capable of playing the role assigned to pain in the classical hedonists' account of happiness, since sensations which are identical in their internal qualities may evoke very different emotional or attitudinal responses. However, whereas it is (logically and psychologically) possible to be indifferent to painful sensations, or even to enjoy them, this is not possible for suffering. ... It is plausible to say of suffering, as it is not of pain, that its presence necessarily compromises our happiness, and thereby also our well-being. Suffering seems just the sort of condition which, in itself and apart from any further accompaniment, makes our lives go worse" (WHE, 105).

[4.3.8.] On this analysis, then, Sumner says we must give up the view that pain *in and of itself* is an intrinsic evil for us (but: this has major implications for the viability of the classical hedonists' account of the nature of welfare, particularly for the sensation model). "... [H]ow physical pain feels to us, how much it hurts, is one thing; how much it matters to us is another" (WHE, 106). "The feeling tone of pleasure is one which we typically like for its own sake. But just as we need not mind pain, we need not welcome or enjoy physical pleasure. Too much of it can saturate or jade us; while this may result in a more muted sensation from the same stimulus, it can also lead us to care less about new episodes of undiminished intensity. We can also find the experience of embodiment occasioned by physical pleasure uncomfortable or unsettling; we can even regard the accompanying organic functions as vulgar or unclean. ... There is no linear dose-response relationship for pleasures, just as there is none for pains. It is our ability to distinguish physical pleasures from our emotional responses to them which forces a choice between the two models of the nature of pleasure: shall we identify it with the sensation itself or with our attitude of liking or enjoyment?" (WHE, 106-107).

[4.3.9.] The attitude model makes more sense for pleasure (as we can take pleasure in a wide range of activities or experiences that have no particular physical responses associated with them), but is ultimately not much more satisfactory than the sensation model (because it conceals important differences between physical pleasures

accompanied by sensations with a positive feeling tone, and all other sources of enjoyment or satisfaction, as the latter depend on our attitudes toward states of the world). “By labelling all enjoyable experiences indiscriminately as pleasures the attitude model erases this boundary between the intrinsically and extrinsically pleasant. In doing so, it loses sight of the features peculiar to strictly physical pleasures ... [and] can make no sense of being indifferent, or even averse, to a physical pleasure. But this phenomenon seems not only a logical possibility, but a common occurrence” (WHE, 108).

[4.3.10.] A more viable approach, following Cassell, is to adopt the sensation model, but restrict it to the core cases of physical pleasures, which distinguishes ‘pleasure’ from ‘enjoyment’. “Like suffering, enjoyment will consist in a response to a situation as a whole, to which a subject brings her entire hierarchy of values and concerns. ... The sources of enjoyment available to us will extend well beyond [the particular kinds of feelings associated with pleasure], so as to include the objective conditions of our lives” (WHE, 108). This distinction makes it clear that what matters to our welfare is ‘enjoyment’ rather than ‘pleasure’ *per se*. “While the sensation model tells the right story about the nature of both pleasure and pain, what it shows us is that these phenomena have no privileged place in an account of happiness or welfare. (They are sources of well-being, or ill-being, not ingredients in its nature.) On the other hand, while the attitude model tells the wrong story about pleasure and pain, the phenomena to which it properly applies – enjoyment and suffering – seem much more appropriate candidates for this role” (WHE, 108-109).

[4.3.11.] What this means for (classical) hedonism is that it can only be a mental state theory of welfare. But the utilitarians never intended their (sensation and attitude) models of pleasure and pain to apply *solely* to the core physical cases; they construed pleasure more broadly to encompass ‘enjoyment’, and pain more broadly to encompass ‘suffering’ (but failed to articulate this). “[On this basis] it looks as though it might be possible to rebuild hedonism ... around the notions of enjoyment and suffering. However, we still lack a solution to the most serious difficulty threatening any such enterprise – the problem of solipsism. On the attitude model, as the utilitarians interpreted it, enjoyment and suffering are all mental states. Or are they?” (WHE, 110).

[4.3.12.] J.S. Mill, in particular, always assumed (based on the technology of his times) that any mental state induced by some activity or experience would be accompanied by said activity or experience (that mental state’s ‘object’), and not artificially induced by some ‘experience machine’. “At this working level, therefore, Mill’s views about happiness (and welfare) are not vulnerable to the charge of solipsism. The problem is that neither he nor any of the other classical theorists ever seems to have developed a theory of the nature of pleasure/enjoyment which was adequate to their own understanding of it, and their view of its place in our lives” (WHE, 110).

[4.3.13.] Both enjoyment and suffering require intentional objects for their completion, which include (though they are not confined to) states of the world. The hedonist can therefore read enjoyment and suffering *extensionally* (so as to say that in order to actually enjoy, or suffer from, ‘something’, that ‘something’ must correspond to some *actual* state of the world), so that enjoyment and suffering “are complexes consisting of mental states plus their objects (which will usually, though not always, be states of the world)” (WHE, 111). Such an account would therefore be immune to the ‘experience machine’ objection, but might be too immune “having committed itself to the view that an experience has prudential value only when it is reality-based”, which seems “objectionably puritanical – surely a good fantasy can contribute something to a life. In any case, subjectivists should want to leave room for individuals to work out their own rates of substitution between illusion and reality” (WHE, 111).

[4.3.14.] Some other theory, working with the materials of ‘enjoyment’ and ‘suffering’, might do a better job of answering the traditional criticisms of classical hedonism (the main one being that hedonism treats enjoyment, and therefore also welfare, as a mental state, or a collection of such states, in *isolation from* actual states of the world). The core truth of hedonism (and one which must remain at the heart of any workable subjectivist account of the nature of welfare) “is that nothing can make our lives go better or worse unless it somehow affects the quality of our experience. ... [A]n adequate theory of welfare ... will have to be experiential: it will have to connect our well-being in some way or other with our experience of the conditions of our lives. The implications of this lesson are best appreciated by turning attention to a form of theory which has not fully absorbed it” (WHE, 112).

5. The Desire Theory (pp 113-137)

[5.1.] “Every subjective theory makes welfare depend, at least in part, on the subject’s attitudes toward the conditions of her life. What all attitudes share is that they are ways of being for or against things, ways of favouring or disfavouring them. Beyond this common defining characteristic, however, attitudes divide into significantly different categories [of pleasures versus pains, or wants, or preferences]. The generic formula of a desire theory [of welfare] is that a condition or state of affairs makes me better off by virtue of satisfying some desire [like wants, or preferences] on my part, that my life is going well to the extent that I am succeeding in getting what I want [or prefer]. The role played by desire in this formula ensures that the theory will be subjective, while the requirement that the desire be satisfied (thus that its object actually exist or come about) ensures that it will be a ‘state-of-the-world theory’” (WHE, 113).

[5.2.] “Desire theories have come to dominate the welfare landscape ... in the way that hedonism dominated in its own time” (WHE, 113).

5.1. Revealed Preference (pp 113-122)

[5.1.1.] This account is unpromising, but it has become popular because of its roots in classical utilitarianism (which held that welfare is the only thing worth promoting for its own sake as the sole foundational good of ethics). Utilitarian economists used welfare as an explanatory variable in their economic theories, the central assumption of whose theory of demand was that market choices would reflect maximisation of utility – or ‘welfare’ as the classical hedonists understood the term. But: ‘objectively’ measuring pleasure and pain turned out to be too unreliable to serve as the basis of an ‘economic science’, so economists detached the notion of ‘economic welfare’ from the (hedonistic) concept of ‘welfare’ employed by the classical utilitarians.

[5.1.2.] ‘Economic welfare’ was invented by A.C. Pigou who, as a *utilitarian* economist, still clung to the classical orthodoxy of a hedonistic account of wherein consists the welfare of people, while acknowledging the difficulties in accessing these states of consciousness for empirical purposes. So, he hived off from the broader notion of ‘welfare’ that part of it concerned with or affected by money (and transactions denominated in it), and called this ‘economic welfare’. “But if welfare is a matter of felt satisfaction how can any part of it be brought into relation with the monetary metric? Pigou’s answer to this question unfolded in two stages. ... [I]ntensity of desire was accepted as a surrogate for satisfaction ... [then] income expenditure was accepted as a surrogate for intensity of desire: how much I am willing to pay for something is an index of how much I want it. When these steps are combined we get the result that an individual’s market choices provide a measure of her economic welfare [which can thereafter be aggregated into national product or income]. ... [W]here individuals are concerned, [Pigou] explicitly acknowledged that economic welfare is but one small part of overall welfare, that an increase in the part may not entail an increase in the whole, that intensity of desire is only a crude indicator of degree of satisfaction, and that the budgeting of income is an index of intensity of desire at best in the special case of market goods” (WHE, 114-15).

[5.1.3.] ‘New welfare economics’ came to redefine utility, and therefore also ‘welfare’, as preferences revealed through choices in suitably idealised markets, in the process converting Pigou’s rough approximations into necessary truths. Over time, ‘economic welfare’ came to be seen (by economists) as the *entirety* of ‘welfare’. “The result was the view that the contribution a thing makes to my well-being is a function of the amount I am willing to pay to get (or keep) it. ... [That ‘economic welfare’ is only an *indicator* of broader ‘welfare’] was utterly unnecessary, at least for the purely explanatory theory of consumers’ behaviour. All of the central results of general equilibrium theory are derivable as long as the market choices of consumers are assumed to manifest certain basic forms of consistency; it is not necessary to suppose that in these choices consumers are seeking to maximise felt satisfaction, or any other experiential magnitude” (WHE, 116).

[5.1.4.] The reconceptualisation of utility in terms of revealed preferences also suited the more normative concerns of welfare economists (because behavioural notions of utility could be more easily made to figure in accounts of ‘rational choice’ by economic agents: ‘ordinal’ *instead of* ‘cardinal’ utility). But: “[i]f utility is given a technical definition as that which is maximised in consistent choice, then the classical utilitarian equation of utility and welfare has been effectively abandoned. In that case it is utility which is indispensable to demand theory, not welfare” (WHE, 117).

[5.1.5.] In fact, positive economics does not *need* to apply welfarist notions to the purely formal results of general equilibrium theory (although the equation of ‘utility’ in terms of money incomes with ‘welfare’ in the presentation and interpretation of such results is a seemingly irresistible temptation). ‘Welfare’ has come to play an important *rhetorical* role in positive economics, while in normative economics, the field is still called ‘welfare economics’ (despite not really being about broader ‘welfare’ as such). “It must be emphasised that none of the (alleged) welfarist results in either domain depends on any assumed empirical correlation between utility (choice) and welfare. Instead, these results, at least on their standard interpretation, presuppose a logical connection between the two. But in that case economists are (implicitly or explicitly) offering us an account of the nature of welfare, one which takes it to be constituted by ‘preference as revealed in choice’. It is this theory that we wish to assess” (*WHE*, 117-118).

[5.1.6.] Sumner begins this assessment by formalising the ‘revealed preference’ theory to make it determinate, as: “(1) *x* makes me better off than *y* just in case I want *x* more than *y*, and (2) I want *x* more than *y* just in case I choose *x* when I could have chosen *y* instead. The first step links welfare and desire, while the second links desire and choice. ... [T]he second link ... is the distinctive mark of revealed preference theories” (*WHE*, 118).

[5.1.7.] Sumner launches his critique of the ‘revealed preference’ theory of welfare by noting the obvious: that not *all* of our preferences for things or states of affairs will be reflected in our *market* choices (as some of these things or states of affairs, like loyalty and friendship, stand outside markets, while others are so fundamental to any decent standard of living, like education or sanitation or health care, that they cannot reasonably be available *only* via private sale: they are too important to us to be allocated by price *alone*).

[5.1.8.] Which means the restriction to *market* choices (required by the theory because preferences can only be ‘revealed’ when money changes hands) cannot be maintained (although dropping it will mean that the theory has no way of valuing non-market choices¹, and therefore no account of how these choices contribute to our well-being). “Dropping the restriction to market choices leaves us with the view that I want *x* more than *y* just in case I choose *x* when *y* is also available. This formula still restricts us to our actual (market and non-market) choices, a limitation which is imposed by the behaviourism of revealed preference theory: revealed preference, by definition, is preference as actually revealed in choice. But this restriction must also be dropped. If my preference structure is thought of as an ordered set of dispositions to make certain kinds of choices, then it seems clear that this will be revealed only very imperfectly by my actual choices [as there are choices I would choose, but the opportunity to choose them does not arise]. ... If my preferences are to connect in some important way with my welfare, then it seems reasonable to think that all of them should count, not just those I have had the opportunity to reveal in choice” (*WHE*, 119).

[5.1.9.] The inclusion of *hypothetical* choices renders it no longer a theory of welfare by ‘revealed preferences’ (as hypothetical choices cannot be so easily ascertained as those ‘revealed’ through market transactions). The resultant theory is a near relation to ‘revealed preference’, since it still attempts to explicate welfare in terms of (actual or hypothetical) choice, links choices to preferences, and preferences to welfare. But: to what extent do our choices *really* reflect our preferences? “[The] notion [of a preference] looks simultaneously in two directions: to my *attitudes* [what I like best or find most agreeable] on the one hand and to my *choices* [what I behaviourally prefer according to my choices] on the other. ... [The first meaning of ‘preference’ is] needed by a subjective theory of welfare, since in this sense it is arguable that the satisfaction of my preferences necessarily makes me better off. In the second sense [of the word ‘preference’], what I (behaviourally) prefer is just what I choose (for whatever reason). This is the interpretation of preference needed in order to connect it analytically to choice” (*WHE*, 120).

[5.1.10.] Acknowledging this ambiguity in the notion of ‘preference’ poses a dilemma for a ‘revealed preference’ theory of welfare: it is only plausibly connected to welfare in the *attitudinal* sense of ‘preference’, which then surrenders a tight connection to *choice*. My choice of *x* over *y* does not necessarily reflect a judgement on my part of a higher prudential payoff for me (as I may be acting out of altruism or a sense of obligation), and the fact that I can have behavioural motivations *other than* my likes and dislikes means that I will not *necessarily* choose *x* over *y* even if confronted by this choice. “Only by equivocating on preference, therefore, can a choice theory use it as a middle term

¹ Sumner was (presumably) unaware at the time of writing of the literature on ‘incentives compatible mechanisms’ and provision of public goods.

between choice and welfare” (WHE, 121), which results in an argumentative fallacy (of *equivocation*, which is the use of a key term or phrase in an argument in ambiguous ways).

[5.1.11.] “There is a parallel ambiguity in our vernacular notion of wanting. Wanting to do something, in the behavioural sense, is just having some reason or other for doing it, with no restriction whatever placed on the range of possible reasons. ... [When I act voluntarily it is always in accordance with what I ‘want’, but] ... [i]n the attitudinal sense ... wanting to do something requires finding the prospect of it pleasing or agreeable, or welcoming the opportunity to do it, or looking forward to it with gusto or enthusiasm. ... [Having this sort of reason to want to do something can easily conflict with other reasons, such as a feeling of obligation or the fear of disapproval.] ... In this sense, therefore, not doing what one most wants to do is depressingly familiar. (The behavioural notion is the one we use when we say, ‘Well, he must have wanted to go, since he went.’ We switch to the attitudinal notion when we ask ourselves, after the fact, ‘Is that what he really wanted to do?’)” (WHE, 121).

[5.1.12.] ‘Wanting’ (like preference) can also be linked with either choice or welfare; but it can’t (except by equivocation) be linked to both at once (and so *can’t* be the middle term between choice and welfare). Economists routinely make use of the ambiguity in the notion of preference to explain consumer behaviour, then shift to the attitudinal sense of the word in their assertions about the welfare implications of markets. This equivocation totally undermines the superficial plausibility of the revealed preference account, so that it can be safely discarded. “There is simply no interesting constitutive connection between welfare and choice. We are left then with the attitudinal notions of preference and desire. Do they provide the materials for a theory of welfare?” (WHE, 122).

5.2. Informed Desire (pp 122-137)

[5.2.1.] “The basic schema for this kind of account is that something benefits a person (directly or intrinsically) by satisfying some desire on that person’s part. Versions of the desire theory now define the orthodox view of the nature of welfare, at least in the Anglo-American philosophical world. In the theory of rational choice the equation of well-being and utility (preference-satisfaction) has achieved the status of an unquestioned axiom [and is defended by a number of prominent moral philosophers]” (WHE, 122).

[5.2.2.] The desire theory is attractive because it’s plausible that my life is going well for me when I am in the way of achieving my aims or goals, which have been determined according to a preference structure of my choosing. The desire theory both provides for a unified account of the *nature* of welfare, while respecting the range and variety of its *sources* (and so can make sense of the heterogeneity of tastes). “At the same time ... if there are states of affairs [for which everyone strives, or views as indispensable], then these will count as (either intrinsic or instrumental) common goods” (WHE, 123).

[5.2.3.] The desire theory also delimits the class of welfare subjects in an intuitively plausible way (because a structure of desires or preferences is required, which presupposes a capacity to form and order desires, and so clearly includes most human beings and many species of animals, but excludes plants and inanimate objects). The theory can also explain how a number of welfare notions have come to be applied to collectivities of one sort or another – these collectivities are analogical to ‘standard’ welfare subjects, and so can be regarded in like terms as requiring certain standard conditions to function properly according to their ultimate end, or purpose. “In this way the theory can draw a principled, though fuzzy, boundary between the classes of primary and secondary welfare subjects” (WHE, 123).

[5.2.4.] Additionally, the desire theory is in keeping with modern political liberalism (in that, unlike objectivist accounts of welfare, the desire theory puts us at the centre of our own lives as “shapers of our own destinies, determiners of our own good” – WHE, 123). As such, the theory’s ‘normatively adequate’ – but is it ‘descriptively adequate’? To consider this, we must make the theory more determinate, as incorporating two essential features of wants or desires: that they are *intentional*, and that they are *prospective*.

[5.2.5.] Desires are *intentional* in that they are focused on objectives the existence of which may not necessarily come to pass. I can desire things, or activities, and it is a simple step to homogenise a wide variety of materially different objects of desire into *states of affairs*: to want a book (for example) is to want to own or read the book, and to want to do

something is to want the state of affairs brought about by you doing that thing. It is then a further simple step to turn these states of affairs into *propositions*: to want the state of affairs which consists of owning the book is to want the proposition 'I own this book' to be true. "By this process of transformation, every desire comes to take some proposition as its intentional object [also referred to as a 'propositional object']. And of course the fact that I want a proposition to be true does not guarantee its truth" (WHE, 124).

[5.2.6.] Such transformations highlight the similarity between desires and beliefs. Like beliefs, desires can be understood as attitudes whose intentional objects are states of affairs or propositions. Like beliefs (which can be verified or falsified), desires can be satisfied or frustrated. A belief is verified by the occurrence of a state of affairs (or the truth of a proposition) which constitutes its intentional object, just like a desire is satisfied by the occurrence of the state of affairs (or the truth of the proposition) which constitutes *its* intentional object. Like beliefs, desires can take as their intentional objects states of the world which are spatially or temporally remote from their holders. Which means states of the world that I desire can come to pass (so that my desires for them have been satisfied) without my knowledge. By contrast, I can't (occurently) like or enjoy something without knowing this. My enjoyments must be experienced by me in a way that need not be true of the satisfaction of my desires.

[5.2.7.] The intentionality of desire or preference is awkward for a theory of welfare. Since what I desire can be spatially and temporally remote from me, and therefore be satisfied without my knowledge, it's hard to see how desires that have been satisfied without my knowledge can make my life go better (especially if they're satisfied after I'm dead). "... [W]hen we look closely[,] claims of posthumous benefits always seem to rest on conflating different modes of value. [One's life's work can have instrumental value for the world]; but this ... is not the question at issue. Somewhat closer to the mark, if achievement is a perfectionist value, as it might well be, then posthumous success could make a life a better specimen of its kind. Perfectionist value seems clearly capable of retroactive improvement [after one is deceased]. But from the standpoint of [one's] self-interest? That certainly does not follow, and seems much less plausible. [One] can think that a more successful life is a better life, without thinking that [one] is made better off by it [after one's death]" (WHE, 126-127). However, that a theory of welfare entails posthumous harms and benefits is not sufficient on its own to establish its descriptive inadequacy.

[5.2.8.] Posthumous cases are just the most dramatic instances in which the satisfaction of some desire provides no direct or indirect *experiential* benefit. The obvious remedy for such a defect is for the desire theory to impose an *experience requirement*. "Such a condition would stipulate that a state of affairs can make me better off only if, in one way or another, it enters or affects my experience. A version of the desire theory which incorporated such a requirement might look like this: x makes me better off (directly or intrinsically) just in case (1) I desire x, (2) x occurs, and (3) I am least aware of x's occurrence" (WHE, 127).

[5.2.9.] Some desire theorists reject an experience requirement, on the grounds that we desire things *other than* states of our own experience (presumably out of concern that, should a desire theory stipulate a requirement that we have mental *experiences* attendant upon satisfaction of our desires, in so doing it will *become* a 'state-of-mind' theory). However, such grounds confuse desire theories with 'state-of-mind' theories, when they're actually (and always) 'state-of-the-world' theories (since the *actual* occurrence of the desired state of affairs is *already* one necessary condition in the analysis). An experience requirement *by itself* cannot convert a 'state-of-the-world' theory into a 'state-of-mind' theory.

[5.2.10.] The 'descriptive adequacy' of a theory of welfare requires some form of experience requirement to insulate it against highly counter-intuitive results, like when the intentionality of desire settles on objects which are spatially or temporally distant from us. Desire theorists resist experience requirements because they fear they will eventually make the desire-satisfaction component redundant. This fear is well-grounded, as we shall see when we turn from the *intentionality* of desire to its *prospectivity*.

[5.2.11.] The analogy between desire and belief is instructive. A belief has two temporal indices, one for when I hold the belief (now) versus one for the intentional object of the belief (next Christmas, say, or last year). I can hold beliefs about intentional objects occurring at any time (past, present or future). Desires also are held at one time and fixed upon temporally indexed states of affairs; but unlike beliefs, states of affairs which are the intentional objects of desires must

be in the future. The future orientation of wanting contrasts with liking or enjoying. I can (occasionally) enjoy only what I already have, whereas I can want only what I haven't yet obtained.

[5.2.12.] This necessary future orientation of desire at best represents an *ex ante* expectation that some desired state of affairs will come to pass, which may be disappointed by my *ex post* experience of that state of affairs. I can realise some personal ambition, have done what I most wanted to do, but not have benefitted from doing it. The problem is not that a satisfied desire did not enter into my experience, but that my experience of the desired state of affairs turned out to be *negative*. I got what (I thought at some earlier time) I wanted, but I was mistaken in the belief that the satisfaction of this desire would make my life go better. Satisfaction of desires is therefore no guarantee that our lives will go well.

[5.2.13.] "Since it is the prospectivity of desire which creates this problem, ... it is common for desire accounts to stipulate that the only desires that count are those which are sufficiently rational or considered or informed, or otherwise 'corrected'. ... Whatever the idealising conditions adopted, their effect will be to screen out some of our actual desires. Only the satisfaction of the surviving subset of desires will count as enhancing our well-being" (*WHE*, 130-131). However, "... it appears that [the condition of appropriate information] is either redundant within a desire theory or inconsistent with it. Suppose that the satisfaction of one of my desires has left me worse off [and] satisfying the desire made me *to that extent* better off, but it also frustrated other, more important desires, so that on balance I ended up worse off. ... The desire theory is perfectly capable of explaining all this without the requirement that my preferences be informed. All it needs to do is bring into play the full structure of those preferences, including my priorities among them. In the light of all my actual desires I have suffered a net loss. Whenever this is the appropriate explanation of my plight ... an information requirement will be redundant" (*WHE*, 131).

[5.2.14.] But I can be worse off in cases involving no further desires. This can happen when my expectations are mistaken about how fulfilling I'll find the satisfaction of my desires *in the future*. "[The mistake can arise] from the gap between my *ex ante* expectation and my *ex post* experience. That gap exists by virtue of the prospectivity of desire: my preferences about the future always represent my view *now* of how things will go *then*. Because the gap results from the very nature of desire, it cannot be closed merely by requiring that desires be rational or considered or informed" (*WHE*, 131-132).

[5.2.15.] That gap can be closed by stipulating that a desire is not truly satisfied if the desired state of affairs turns out to be disappointing or unrewarding on later experience of it. But then, this situates an account of welfare in *our* satisfaction, rather than the satisfaction of the intentional objects of our *desires*. "If an information requirement has any genuine work to do within a desire theory ... it will be inconsistent with the basic rationale of the theory" (*WHE*, 132).

[5.2.16.] "All of the difficulties for the desire theory which we have so far canvassed have a common logical form. Desires whose objects prove disappointing in the actual experience of them and desires whose objects never enter into our experience at all – these are both cases in which the satisfaction of our desires appears insufficient to make us better off. What they demonstrate is that when a desire is satisfied it is a logically open question whether the welfare of its holder has thereby been enhanced. However, ... [if] desire-satisfaction is not logically sufficient for well-being, it is not logically necessary either. ... Once again, the root of the problem lies in the very nature of desire. ... [J]ust as we can be disappointed when we get what we expect or have aimed for, so we can be pleasantly surprised when we get what we do not expect and have not aimed for" (*WHE*, 132).

[5.2.17.] In response to the expansiveness of desire (in space *and* time) which causes problems for the desire theory, desire theorists may try to narrow the theory by focusing on just those desires which are also aims or goals of some 'life plan'. By doing so, these theorists equate welfare with success in achieving our aims or goals. But: the fact that some benefits we enjoy are the result of good fortune rather than planning and achievement does not make them any less worthwhile (or prudentially valuable). Moreover, I can be introduced to new experiences not associated with an antecedent desire on my part, yet the state of affairs associated with the experience turns out to have enhanced my well-being. The antecedent desire was not logically (or practically) necessary for my enjoyment of the (hitherto unlooked for) experience.

[5.2.18.] The desire theory faces another sort of difficulty, from the fact that most of our desires have *grounds* (i.e. the reason one has for wanting a desire's object, which can be rooted in any one or more of our tastes, interests, ideals, commitments, and so on). Such concerns will vary from person-to-person (which is hardly surprising); but one we'll all have in common is our own welfare, so that self-interest is one possible ground of desire (which arises whenever a desire's object involves some anticipated payoff for us). Desires grounded in this way are *interested*. "It seems plausible to think that the satisfaction of our interested desires will standardly benefit us ... however, not all of our desires are interested in this way. ... [I can desire states of the world which entail no particular prudential payoff for me personally; such desires are therefore] *disinterested*. The desire theory tells us that the satisfaction of any (rational, considered, informed) desire makes one better off, regardless of its ground. But this is not plausible in the case of disinterested desires" (*WHE*, 134).

[5.2.19.] Disinterested desires are awkward for the desire theory (as that theory posits that doing what I want to do in the attitudinal sense is what's best for me, so I will *necessarily* do what's best for me). But: I need not always do what's best *for me*, because I can choose other than to do what I most want to do. Yet the theory says that what I'm doing *must be* what's best for me, *because* it's what I've chosen to do. Such a result renders the notion of a disinterested desire incoherent (even though such desires can obviously exist, because we are not psychological egoists). And, "[s]ince we are not psychological egoists, we are capable of finding enjoyment or satisfaction in doing good to others. Attitudinal desires therefore need not be self-interested; their satisfaction can be sought at considerable personal sacrifice. It is this possibility the desire theory threatens to erase. ... [I]f I am successful [in altruistic action] then I will have made no sacrifice at all, since I will [simply] have got what I most wanted. I can represent the desire to myself as disinterested, therefore, only if I do not subscribe to the desire theory of welfare" (*WHE*, 135).

[5.2.20.] Once again a gap opens up between desire-satisfaction and well-being, this time in the case of disinterested desires. The gap can only be closed if this class of desires is *excluded*. "The desire theory would then hold that a state of affairs benefits me whenever it satisfies some self-interested desire on my part. [But this is patently circular.] Our understanding of the nature of prudential value is not advanced by being told that it consists in the satisfaction of just those desires whose ground is prudential [i.e. 'what's good for you is what you do because it's good for you']" (*WHE*, 135).

[5.2.21.] Sumner concludes that the desire theory is both too broad (by admitting desires whose satisfaction is not beneficial) and too narrow (because it excludes benefits not produced by the satisfaction of antecedent desires): these are problems of *scope* (which do not prove that the desire theory is descriptively inadequate, only that it needs more work to better mesh desire-satisfaction with welfare). More seriously, Sumner also concludes that the desire theory operates at the wrong *level* to be descriptively adequate (because it confuses the *nature* of welfare with its *sources*, by stipulating that our welfare *ought* to consist in desire-satisfaction, regardless of whether I find satisfaction of some or all of my desires disappointing, or benefit from occurrences which I did not expect or plan for, and did not antecedently desire). "[The desire theory] purports to be a formal theory about the nature of welfare and not just a (very short) list of its sources. But if getting what we want is merely one way in which our good can be advanced then, despite this claim of formality, the ... theory is operating at the wrong logical level" (*WHE*, 136-137).

[5.2.22.] "The mistake made by [the teleological, classical hedonistic and desire theories of welfare] is the same: they all attempt to build an account of the nature of welfare around one of its standard generic sources. The symptom of the mistake is also the same in each case: whatever the favoured source may be, its possession is logically neither necessary nor sufficient for our well-being. The logical gaps which have opened up for all these theories can be closed only by a formal theory capable of explaining what it is about all these goods which makes them cardinal ingredients of our welfare" (*WHE*, 137).

6. Welfare and Happiness (pp 138-183)

[6.1.] Having concluded (over the course of Chapters 4 and 5 of *WHE*) that "neither of the historically dominant versions of subjectivism is adequate to [welfare's] nature", Sumner states that "the two rejected [subjective] theories are mirror images of one another, the strength of each corresponding to the weakness of the other. What [classical] hedonism lacks is a reference to the world outside the subject's mental states, while the desire theory, in correcting this problem,

loses all connection with the subject's experience of the conditions of his life" (WHE, 138). That prudential goods must actually be *experienced* to enhance our welfare makes (classical) hedonism (rather than the desire theory) the place to start in trying to construct a 'descriptively adequate' alternative theory of welfare.

[6.2.] Moreover, "[t]he idea that there is an interesting connection between well-being and happiness [as asserted by the classical hedonists] has at least some initial plausibility – much more plausibility than any account of happiness in terms of pleasurable feelings. As a new start, therefore, it seems worthwhile seeking a better (and less hedonistic) theory about the nature of happiness" (WHE, 138).

[6.3.] However, "... no simple theory about the nature of happiness enjoys much support among philosophers; there is not even agreement that such a theory is possible. About the only thing that everyone agrees on is that happiness is a complex and multi-faceted notion, one not easily reduced to a formula or slogan" (WHE, 139). Given this, getting to a novel theory of welfare based on happiness will be far from straightforward, so "[i]n working through the argument of this chapter, it may help to have some sense of its destination. The theory I shall defend does not simply identify well-being with happiness; additionally, it requires that a subject's *endorsement* of the conditions of her life, or her *experience* of them as satisfying or fulfilling, be *authentic*. The conditions for authenticity, in turn, are twofold: *information* and *autonomy*. Welfare therefore consists in *authentic happiness*. This theory is subjective, since it makes a subject's welfare depend on her attitudes, and since the function of the authenticity requirement is to ensure that these attitudes are *genuinely hers*. It satisfies the experience requirement, since a subject's happiness is a matter of her experience of the conditions of her life. However, it is not a mental state theory since [*circumstantially consensually sufficient*²] authenticity is a relation between the subject and the world. The happiness theory thus mediates between [classical] hedonism and the desire theory, exploiting the strengths of each while avoiding their weaknesses" (WHE, 139; my emphases).

6.1. Life Satisfaction (pp 140-156)

[6.1.1.] "Happiness, like welfare, is subjective ... [but while] the subjectivity of welfare is disputed and needs to be argued for ... no objective theory about our ordinary concept of happiness has the slightest plausibility Objectivism has at least superficial plausibility as a theory of welfare; an argument is required to show why it is mistaken. It is completely unintelligible as a view about the nature of happiness" (WHE, 140). [This is, presumably, part of the reason why Sumner has chosen a form of *sufficient* 'happiness'/'unhappiness' as *attitudes* toward one's life admitting of its (positive or negative) prudential *valuation*, wherein 'welfare' can take its object in his theory of it.]

[6.1.2.] Recapping his critique of classical hedonism, Sumner reminds the reader that 'happiness' does not reduce to simple states of feeling 'pleasure' or 'pain' "[b]ecause we can distinguish between the intrinsic phenomenal properties of pleasure and pain, on the one hand, and our attitudes toward them, on the other, [so that] it is possible for us to be indifferent to either, or even to dislike the former and like the latter" (WHE, 140). "In short, the proper place for physical pleasure within a life cannot be dictated by a theory, but needs instead to be fixed by the subject's own priorities. Which is to say that it is not the phenomenal properties of the sensation itself which determine just how much a particular pleasure will contribute to our happiness, but rather the meaning or significance which we attach to it" (WHE, 141).

[6.1.3.] Sumner argues that the same holds for pain, with some qualifications, such as in the case of "chronic pain, especially if it is intense but also at levels which would be easily bearable for shorter intervals. In these cases indifference to the pain, or assimilation of it into the normal course of our lives, becomes no more than a bare logical possibility. Intense chronic pain is absorbing, exhausting, debilitating, and depressing. It does not merely keep our lives from being happy, it makes them miserable" (WHE, 141-142).

² Or: *enough* information and/or autonomy in the circumstances of a particular case so that a reasonable and disinterested 'moral agent' sharing an 'ordinary concept' of welfare with a welfare subject would be *reluctant to defuse* that subject's (otherwise presumptively or substantively) authoritative self-assessment of their well-being. This can be referred to as the epistemic standpoint of an '*ordinary welfare assessor*' determining the 'phenomenal' well-being of *others* whilst undertaking ordinary assessments of welfare (in contrast to the epistemic standpoint of an '*authoritative welfare assessor*' determining 'essential' well-being whilst undertaking authoritative assessments of *their own* welfare).

[6.1.4.] So "... it is misleading to continue to use the labels 'pleasure' and 'pain' to identify these positive and negative attitudes toward the conditions of our lives, especially when our conceptual [repertoire] offers us ready alternatives ... [in the notions of] enjoyment and suffering [about which we can] then raise the question of the relationship between these phenomena and happiness' (WHE, 142).

[6.1.5.] Sumner notes that the relationship appears to be a close one at first glance: "[e]njoyment and suffering are plausible candidates for a constitutive role in a theory of happiness precisely because they are (positive and negative) attitudes or responses to the experiences which make up our lives. It seems roughly right to say that we are happy when we have a (preponderantly) affirmative attitude toward the conditions of our lives, and unhappy when our attitude tends toward the negative. And, indeed, this *is* roughly right. But only roughly. The appearance of a tight conceptual connection between enjoyment/suffering and happiness/unhappiness is fostered in part by a systematic ambiguity in these latter notions [which means that these ontological categories don't map 'one-to-one']. This is not the occasion to attempt an exhaustive inventory of the many ways in which we use the terms 'happy' and 'unhappy' (the results would in any case be of limited philosophical interest). However, four kinds or dimensions of happiness are worth distinguishing" (WHE, 143).

[6.1.6.] *Being happy with or about something* ('1st kind' happiness). Requires completion by an 'intentional object' (i.e. you have to *want* something in particular to *be* or to *happen*, or not to *be* or *happen*, to be this kind of happy or unhappy 'with' or 'about' it), but no affect or occurrent feeling on your part is required (so that *feeling* happy or unhappy *with* something is the same as *being* happy or unhappy *about* it). "The favourable assessment involved need not be a strong one. Being happy with something is roughly equivalent to being satisfied or content with it, i.e. finding that it measures up reasonably well to some standard. The degree of endorsement in question falls well short of enthusiasm or delight. Contrariwise, being unhappy with something means regarding it in an unfavourable light, though not as an absolute disaster" (WHE, 143-144).

[6.1.7.] *Feeling happy* ('2nd kind' happiness). Involves an occurrent feeling, but requires no intentional object. Defined by a mood of optimism or cheer (or pessimism or ill-humour), which situates you in a moment of your life (so that 'right here, right now' you are feeling happy, or unhappy). Feelings of (this kind of) happiness or unhappiness may be persistent, or short-lived. "In human beings, with their greater cognitive capacities, feelings of happiness or unhappiness may also have a judgemental dimension, an assessment of our lives as going well or badly. But even this assessment tends to be tied to the here and now: at this moment everything seems just terrific/awful. Feelings of [this kind of] happiness and unhappiness are frames of mind subject to fluctuation from day to day, rather than settled judgements about the quality of our lives" (WHE, 145).

[6.1.8.] *Having a happy disposition/personality* ('3rd kind' happiness). An 'orientation', or 'settled tendency', toward how you live your life (i.e. whether you *choose* to see the glass as 'half-full' or 'half-empty' most of the time). Where that 'orientation' is broadly positive, "we may say that you have a happy ... disposition, or that you are a basically happy person. This is the sense in which animals or infants can be happy, despite being incapable of sizing up their lives as a whole" (WHE, 145).

[6.1.9.] *Being happy/having a happy life* ('4th kind' happiness). This is the notion of happiness with which Sumner is principally concerned, "that in which you are (have been) happy or your life is (has been) a happy one. Being happy in this sense means having a certain kind of positive attitude toward your life [as a 'propositional object'], which in its fullest form has both a cognitive and an affective component. The cognitive aspect of happiness consists in a positive evaluation of the conditions of your life, a judgement that, at least on balance, it measures up favourably against your standards or expectations. This evaluation may be global, covering all of the important sectors of your life, or it may focus on one in particular (your work, say, or your family). In either case, it represents an affirmation or endorsement of (some or all of) the conditions or circumstances of your life, a judgement that, on balance and taking everything into account, your life is going well for you" (WHE, 145).

[6.1.10.] "Clearly this sort of prudential stocktaking is possible only for creatures capable of assessing their lives as wholes, either at a time or over some extended period of time. The cognitive component of [this kind of] happiness is therefore beyond the range of many subjects-of-a-life, such as small children and non-human animals. However, there

is more involved in being happy than being disposed to think that your life is going (or has gone) well [a point Sumner makes I think to avoid '4th kind' happiness collapsing into '3rd kind' happiness in the case of animals, or infants]. The affective side of happiness consists in what we commonly call a sense of well-being: finding your life enriching or rewarding, or feeling satisfied or fulfilled by it. Because it is less cognitively demanding than a judgement about how one's life is going as a whole, it is what we have in mind when we say that a child or an animal is happy, or is leading a happy life" (*WHE*, 145-146). {"Being unhappy [in the case of 'paradigm welfare subjects'] is therefore a matter both of evaluating your life (or some part of it) negatively, as failing to meet your standards for it, *and* of experiencing it as unsatisfying or unfulfilling" (*WHE*, ch. 6, fn. 4; my emphasis).}

[6.1.11.] '4th kind' happiness is best understood by comparing it with the other three kinds. For starters, you can be happy 'with' or 'about' a great many things, without the frequency and/or intensity of episodes of '1st kind' happiness correlating linearly with '4th kind' happiness. The only interesting overlap between '1st kind' and '4th kind' happiness is "when *what* you are happy with is your life as a whole (or some significant part of it). But even here more is involved in being ['4th kind'] happy than the bare positive evaluation; *you must also* experience your life as satisfying or fulfilling" (*WHE*, 146; my emphases).

[6.1.12.] There's also no straightforward relationship between '4th kind' happiness and episodes of '2nd kind' happiness. "The latter are related to the former as its contingent sources or ingredients: they will tend to make a life a ['4th kind'] happier one, but they are not necessary for ['4th kind'] happiness" (*WHE*, 146).

[6.1.13.] Similarly, '3rd kind' happiness is also not required for '4th kind' happiness because "[even] if one is predisposed [toward '3rd kind' unhappiness, like being] gloomy and sour ... [so that the] occurrent feelings [of '2nd kind' unhappiness] contribute toward making us ['4th kind'] unhappy ... they may be counterbalanced by other factors, such as a sense of hope or purpose" (*WHE*, 146).

[6.1.14.] "What being ['4th kind'] happy shares with feeling ['2nd kind'] happy is its affective component – the experience of something (in this case your life as a whole or some important sector of it) as worthwhile or rewarding. There is no sharp distinction between them. Feelings of ['2nd kind'] happiness (especially intense ones) are generally short-lived, but they are capable (at least in principle) of enduring for some time, at which point they become difficult to distinguish from a settled sense of satisfaction with the conditions of one's life" (*WHE*, 147).

[6.1.15.] Contemplating the various 'kinds' of happiness (or unhappiness) we can experience, it's clear that the relationship between enjoyment/suffering and happiness/unhappiness depends on the *kind* of happiness/unhappiness we have in mind.

[6.1.16.] There need be no necessary relationship between 'enjoying' something, and being 'happy' with or about it (as it follows that if you're 'enjoying' something you're 'happy' with it, but the converse need not hold, as the range of things about which we can be '1st kind' happy is much broader than the things we find experientially agreeable or rewarding for their own sake).

[6.1.17.] '2nd kind' happiness is much more closely related to enjoyment in that "a feeling of happiness is an occurrent (attitudinal and affective) response to (what you perceive as) the conditions of your life – one which is generally episodic, with a fairly definite duration" (*WHE*, 147) (although, unlike enjoyment and '1st kind' happiness, feelings of '2nd kind' happiness require no 'intentional object'; and while the notion of enjoyment "may be adequate for capturing many happy feelings, it seems much too tepid for the heights of rapture or bliss"; *WHE*, 147).

[6.1.18.] '3rd kind' happiness generally reduces to '2nd kind' happiness (so that one's disposition gives rise to certain affective *tendencies*, which then manifest as occurrent feelings in the domain of '2nd kind' happiness).

[6.1.19.] But what of the relationship between 'enjoyment' and '4th kind' happiness? (i.e. the leading of a happy life *overall*, which the classical hedonists identified with an existence typified by 'enjoyments' rather than 'suffering'). Were the classical hedonists right?

[6.1.20.] Sumner argues that “enjoyment and suffering are still too episodic, too tied to experiences of specific activities or conditions, to be identifiable with [‘4th kind’] happiness and unhappiness. ... [T]here is no algorithm for computing your level of [‘4th kind’] happiness from the intensity or duration of your particular enjoyments or sufferings. ... [E]njoyments and sufferings are typical sources of [‘4th kind’] happiness and unhappiness ... [but] success or failure in the pursuit of your aims count as well. The desire theory went wrong by treating desire-satisfaction not as one important source of well-being but as constituent of its nature. Hedonism, even the improved version which takes enjoyment and suffering as its central notions, likewise confuses an important source of [‘4th kind’] happiness with its nature” (*WHE*, 148).

[6.1.21.] Enjoyment and ‘4th kind’ happiness come closest to converging when it’s (the unity of) your *whole life* (as a ‘receptacle’ or ‘container’ for prudential value) that your *attitude* of ‘enjoyment’ is directed toward, either now or over time. But these notions *still* don’t quite converge (because enjoyment understates the judgemental component of ‘4th kind’ happiness for those welfare subjects capable of such prudential evaluations, and fails even to capture the full range of the affective dimension of ‘4th kind’ happiness, “which can extend from bare contentment to deep fulfilment. Enjoyment is simply too mild to cover the more intense regions of *this scale*. ... We will do better at locating the attitudes we are seeking here if we shift from the notion of enjoyment to something like *satisfaction* or *fulfilment* ... not desire-satisfaction [because of the logical gap between fulfilment of your *desires*, and *your* fulfilment] but *personal* or *life satisfaction*” (*WHE*, 148-149; my emphasis on the phrase ‘this scale’).

[6.1.22.] “This identification of [‘4th kind’] happiness with life satisfaction is a theme running through most of the recent philosophical literature on happiness. It is also the accepted presupposition of most empirical studies carried out by social psychologists [O]ver the past three decades or so [to October 1995, when *WHE* was completed] a flourishing social-scientific industry has grown up which is dedicated to the measurement of [‘4th kind’] happiness, especially on a social scale, and which construes [‘4th kind’] happiness as life satisfaction” (*WHE*, 149).

[6.1.23.] This has come about because of increasing dissatisfaction amongst social scientists with objectivist notions and measures of ‘economic welfare’ masquerading as broader ‘welfare’, culminating in the ‘social indicators movement’ (a movement which seeks to identify those objective or subjective indicators of well-being most closely correlated with the observed *phenomena* of welfare; *WHE*, 149-151).

[6.1.24.] In response to weak correlations between readings on *objective* indicators of well-being or ‘quality of life’ and the reports of welfare subjects themselves, the ‘social indicators movement’ has turned increasingly to *subjective* indicators (with a consequent improvement in correlations between readings on such indicators and welfare assessments made by welfare subjects; *WHE*, 151-153).

[6.1.25.] “[This shift] reflects a number of assumptions, implicit or explicit: (1) that welfare is subjective, (2) that it is either identical with or at least closely related to [‘4th kind’] happiness, (3) that [‘4th kind’] happiness consists in life satisfaction, and (4) that people’s self-assessments provide the most reliable measure of how satisfied they are with their lives, or with particular sectors of their lives. We have by now [in the argument of *WHE* to this point] ample reason to accept (1), and the earlier part of this section [of *WHE*] has made a case for (3). What about (4)? Are people generally reliable informants about their own levels of [‘4th kind’] happiness or life satisfaction?” (*WHE*, 153). To which Sumner’s answer is: not always.

[6.1.26.] *A subject may get the questions wrong*. If that happens, their self-evaluations may be *irrelevant* (but, if the questions relate to their subjective well-being across various ‘life domains’, the importance of which they can prioritise for themselves, at least they’re reporting how they think and/or feel their lives are going *for them*).

[6.1.27.] *A subject may be insincere*. This can happen where a respondent reports how [‘4th kind’] happy he thinks (or thinks *we* think) he *ought* to be. “There is considerable experimental evidence which suggests that subjects systematically overstate their levels of life satisfaction, so as to represent themselves ... as happier than they actually are [for which compensatory adjustments in measurement can be made]. ... [S]elf-assessments will be reliable indicators of happiness only when they are frank and honest” (*WHE*, 155).

[6.1.28.] *A subject's assessments may be coloured by transitory factors (such as mood). "The solution to ... possible distortion by passing mood is to seek self-assessments which are considered or reflective – that is, consistent over time and representing your settled view of your level of life satisfaction" (WHE, 155).*

[6.1.29.] "... [P]eople's self-assessments tend to be reliable when they are relevant, sincere, and considered. ... [I]t is never possible to eliminate all sources of bias or distortion in self-reporting. But it is also worth keeping in mind that in determining how happy people are we are not solely dependent on what they say. We can also refer to behavioural signs (happy people tend to act happy) as well as second-person assessments by knowledgeable others (happy people tend to look happy to their friends and relations). In the case of very young children, as well as non-human animals, we have nothing but this non-verbal information on which to rely. ... [Nevertheless] ... we have a reasonably high level of confidence in our ability to determine how happy (or unhappy) subjects are. There is no reason of principle to think that this confidence is misplaced, any more than our similar confidence in our ability to determine what subjects want or believe on the basis of what they say and do. The real philosophical problems lie elsewhere, not in reliance on subjects as *authoritative sources concerning their own happiness* but in the bearing of this information on an assessment of their well-being. ... [Subjectivists] in the social indicators movement think that their data measure not merely the ['4th kind'] happiness or life satisfaction of their subjects but their welfare as well (see assumption (2), above). We must now ask whether they are right" (WHE, 155-156; my emphasis).

6.2. Authenticity and Autonomy (pp 156-171)

[6.2.1.] "So far we know that happiness, or life satisfaction [as '4th kind' happiness], is a positive cognitive/affective response on the part of a subject to (some or all of) the conditions or circumstances of her life. What remains to be decided is what all of this has to do with well-being. The simplest relationship between ['4th kind'] happiness and welfare would ... be identity – which is the operational assumption of the 'subjective indicators' school. However, there are two serious impediments in the way of accepting this assumption" (WHE, 156).

[6.2.2.] The first is the possibility of mistake (either through factual error or (self-)deception). So, "[i]f we identify well-being with happiness, and if we treat happiness solely as a function of a subject's *experience* of her life, then the result will be a mental state theory of welfare [and] it will ... assess a subject's well-being entirely 'from the inside' with no reference to the actual conditions of her life. ... [T]his simple identification of welfare with happiness – of well-being with 'perceived well-being' – must be rejected" (WHE, 156-157).

[6.2.3.] One possible response to this problem is to modify a 'happiness as life satisfaction' account "by adding the condition that a person's positive evaluation of her life will count as (real or true) happiness only if it is based on beliefs about the world which are true (or at least justified)" (WHE, 157). But what of a person's happiness that is the product of some deception: once the deception is exposed, does this mean that *now* that person wasn't *truly* happy at any time *before* exposure of the deception, even though the person felt they were *at the time*, and indicated this to others *at that time*? This is clearly implausible.

[6.2.4.] Philosophers have been tempted to add a truth or justification condition to their respective conceptions of happiness to ensure that a positive assessment of one's life is a well-grounded one (and avoid the problem of mental state theories of welfare being divorced from the real conditions or circumstances of welfare subjects' lives). But such accounts become a 'state-of-the-world' analysis of happiness (to preserve the identity with welfare), and objective conceptions of happiness have no plausibility.

[6.2.5.] A better way to proceed is to accept that happiness is a state of mind, give up a simple identity between ['4th kind'] happiness and welfare, and add whatever further epistemic conditions are appropriate to your conception of welfare. One might stipulate a (strong) *truth or reality requirement*, which would say that happiness only counts toward well-being just in case it is free from factual error (which appears to be 'unreasonably puritanical' to Sumner, given that we can be wrong about the grounds for our happiness at the time that we experience it, find out that we were wrong after the fact, but then still view the period during which our happiness was wrongly grounded as having prudential value). "When we reassess our lives in retrospect, and from a superior epistemic vantage point, *there is no right answer to the question of what our reaction should be* – that is surely up to us. Because a reality requirement stipulates a right

answer – any happiness based on illusion can make no intrinsic contribution to our well-being – it must be rejected as presumptuously dogmatic. It seems even more dogmatic from a third-person standpoint: who are *we* to dictate that the solace *someone else* finds in a comforting fantasy should count for nothing?” (WHE, 158-159; my emphases on the word ‘we’ and the phrase ‘someone else’).

[6.2.6.] A weaker epistemic condition would be a *justifiability requirement* “since it would not discount the meaning brought to our lives by assumptions which, though false, were at least reasonable under the circumstances” (WHE, 159), and would have the advantage of more closely aligning with a subject’s point of view about the conditions of her life. Yet, such a requirement still dictates how much departures from some ideal epistemic standpoint should matter to people in their self-assessments of prudential value (which is ‘unreasonably arrogant’ in Sumner’s view).

[6.2.7.] Another approach would be to say that our prudential evaluations should be (appropriately) *informed*. This is because happiness as personal or life satisfaction (‘4th kind’ happiness) requires the endorsement or affirmation of the conditions of one’s life. “In this context there is an evident rationale for requiring that the endorsement in question be informed. After all, what we are seeking is an adequate subjective theory of welfare, one on which the subject’s point of view on her life is authoritative for determining when that life is going well *for her*. By connecting welfare with happiness we have interpreted that point of view as an endorsement or affirmation of the conditions of her life. When that endorsement is based on a clear view of those conditions, we have no grounds for questioning or challenging its authority: in this respect, the individual is sovereign over her well-being. But when it is based, wholly or partly, on a misreading of those conditions then its authority is open to question, since it is unclear whether or not she is endorsing her life *as it really is*. When someone is deceived or deluded about her circumstances, in sectors of her life which clearly matter to her, the question is whether the affirmation she professes is *genuine* or *authentic*. In order for a subject’s endorsement of her life to accurately reflect her own priorities, her own point of view – in order for it to be truly *hers* – it must be authentic, which in turn requires that it be informed” (WHE, 159-160).

[6.2.8.] But: what *kind* of information requirement? A reality requirement is based on ‘ideal information’, whereas a justification requirement is based on ‘reasonable belief given the information available’, yet neither is “... compatible with the individual sovereignty which characterises a subjective theory, [so] we must find some other way of determining how well informed a subject must be in order for her happiness to count also as her well-being [that is, for prudential evaluations of her life to be *authoritative*]. The place to start is with a (slightly) different question: when is (more) information relevant? The obvious answer, on a subjective account, is: whenever it would make a difference to a subject’s affective response to her life, given her priorities. ... The problem with reality or justification requirements is that they impose uniform discount rates on everyone alike: happiness has no prudential payoff unless fully informed, or is discounted at a steady rate as it becomes less informed. The relevance of information for a person’s well-being is a personal matter to be decided by personal priorities; there is here *no authoritative public standard*. Still, the problem remains that the self-assessments which individuals report cannot merely be taken at face value; we need to know whether they are authentic. The best way to capture the conditions they must satisfy is to say that they are *defeasible* – that is, they are authoritative unless we have some reason to think that they do not reflect the individual’s own deepest priorities. Where someone’s endorsement of his life is factually uninformed, or misinformed, that gives us one reason for doubting its authority (whether it is a *sufficient* reason depends on whether the endorsement will, or would, survive the acquisition of the *missing* information)” (WHE, 160-161; my emphases on the phrase ‘no authoritative public standard’, and the words ‘sufficient’ and ‘missing’).

[6.2.9.] “Factual error is not, however, the only possible reason for questioning the prudential authority of self-assessments [of well-being], nor is it the most important” (WHE, 161). The malleability of personal preferences (or ‘values’, or ‘priorities’) as pointed out by Amartya Sen (in a famous passage) means that we can’t simply take for granted that what welfare subjects *say* matters to them isn’t *in fact* a product of some form of *domination* (to which they’ve accommodated themselves by ‘adjusting down’ their expectations of life), or *indoctrination* (which perverts the development of self in ways that implicitly limit or proscribe such expectations). “This problem cannot be met by merely stressing the cognitive/judgemental aspect of [‘4th kind’] happiness, since the extent to which people endorse the conditions of their lives will depend on their expectations for themselves, which are notoriously subject to external

manipulation through mechanisms of conditioning, indoctrination, or socialisation³. Clearly the requirement that endorsement be empirically informed will not suffice to exclude these social influences on the standards by which people judge how well their lives are going; the problem here is rooted not in the adequacy of people's factual information but in the malleability of their personal values. There seems to be nothing in [Sumner's] theory so far which would rule out finding fulfilment in forms of life which are trivial or exploitative or demeaning" (*WHE*, 162).

[6.2.10.] Some philosophers have responded to Sen's problem ("of the hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the over-exhausted coolie" leading lives in which they're forced to obtain what 'life satisfaction' they can from the 'small mercies' made available to them) by constructing 'hybrid' theories of welfare, which combine objective and subjective components (via the formula that something contributes to a subject's well-being directly or intrinsically if, and only if, (1) they find it satisfying or fulfilling, or endorse it as an ingredient in their life, and (2) it is *independently* valuable). On Sumner's analysis, such theories are still 'subjective' (since they preserve the necessity of the subject's positive attitude toward their life).

[6.2.11.] However, the second condition introduces a *value requirement* into the analysis of well-being (a feature typical of objective theories of welfare). "Here the subject is presumed to be mistaken not about some state of the world but about the value, from an independent standpoint, of some condition of his life ... [a] value requirement ... presupposes that there is an evaluative analogue to empirical truth or reality: a right answer to every question about value. How are we to determine which aims or activities or forms of life really *are* valuable? ... [I]f circularity is to be avoided, [the evaluative standpoint to be used] cannot be prudential value .. [is it to be instead aesthetic value? perfectionist value? ethical value?]. This option looks no more appealing in this context than it did as a self-standing account of welfare: it just does not seem true that my life automatically goes better *for me* if the goals I am pursuing rank higher rather than lower from [some] external standpoint" (*WHE*, 164-165).

[6.2.12.] Additionally, "a value requirement still seems objectionably dogmatic in imposing a standard discount rate on people's self-assessed happiness. ... When we undertake a retrospective reassessment of our lives we still judge the earlier part from the vantage point of our later standards. Thus from the first-person standpoint the question never arises of the bearing on our well-being of values which, however objectively 'correct', we never accept. To raise this question we need to take a second-person point of view on someone else's life" (*WHE*, 166).

[6.2.13.] Instead of a value requirement, we could instead try a justification requirement (on the basis that the prudential value of a subject's life would be discounted according to the extent to which the subject's belief in its value was unreasonable given the circumstances of that belief). "But it should be clear by now that all of these reality/value and justification requirements are unacceptably patronising and puritanical in their implications concerning the quality of people's lives. What we need instead is a counterpart to our earlier information requirement, with the *defeasibility it implies* for individual self-assessments. And once we begin to reflect on Sen's problem, its solution, within the framework of a subjective theory, seems pretty clear. ... The problem is not that [the welfare subjects in Sen's example] are objectively mistaken [in their values] but that they have never had the opportunity to form their own values at all. They do not lack enlightenment, or insight into the Platonic form of the good; they lack autonomy" (*WHE*, 166; my emphasis).

[6.2.14.] "Let us say, then, that (self-assessed) ['4th kind'] happiness or life satisfaction counts as well-being only when it is autonomous. But when is that? ... A person is autonomous when her beliefs, or values, or aims, or decisions, or actions are, in some important sense, *her own*. There is therefore an evident connection between autonomy and what we have been calling authenticity. We have said that a subject's affirmation or endorsement of her life is made from her own point of view, is truly *hers*, only when it is authentic. The demand that self-evaluations be authentic flows from the logic of a subjective theory, which grounds an individual's well-being on *her* (positive or negative) attitudes. One side of this demand is the requirement that subjects be informed about the conditions of their lives. The other is the requirement that they be autonomous" (*WHE*, 167).

³ In other words: a subject's personal values are *heteronomous* ("subject to a law or standard external to oneself").

[6.2.15.] “Authenticity is a core element in the concept of autonomy, just as subject-relativity is a core element in the concept of welfare. But it is not sufficient to yield a theory about the nature of autonomy, since it is the common property of all rival theories [of autonomy]” (WHE, 167). In the philosophical literature at the time WHE was completed (i.e. October 1995), two main types of theories of autonomy predominated.

[6.2.16.] The first type says that “a person’s values count as her own if she has identified with them, or acknowledged them as her own, or endorsed them as *her* standards for the conduct and assessment of her life. This process of identification requires the capacity for critical reflection on one’s aims or goals ... expressed in terms of a hierarchy of desires or preferences. First-order desires take as their objects actions of the agent or states of the world; second-order desires take as *their* objects first-order desires. On this hierarchical view to endorse or accept a first-order desire is to make it the object of a second-order desire” (WHE, 168). But: not taking a *particular* first-order desire as the ‘object’ of a second-order desire is really just a conflict of desires (rather than the necessary result of some process of critical reflection and acceptance/rejection). Moreover: why assume that first-order desires are *less* reflective than their second-order counterparts? And, in any case: how do we know that in the construction of this hierarchy of desires or in the conduct of the second-order reflective process that a subject’s underpinning values or standards have been accepted autonomously anyway? So, “... [although an] analysis of autonomy which highlights the psychological process of identification, and which therefore requires the capacity to step back from one’s goals or values, seems to be on the right track ... it cannot tell the whole story about autonomy. For the rest of that story we need another approach” (WHE, 169).

[6.2.17.] The second type (exemplified by John Christman) focuses on the manner in which desires have been formed (paying particular attention to processes of preference formation which are ‘manipulative’ in ways which are crucially different from ‘normal’ processes): an *historical* approach to theorising autonomy. With reference to Sen’s problem, “the reason we are reluctant to take at face value the level of life satisfaction reported by Sen’s beggars, labourers, and housewives is that we are suspicious of the socialisation processes by means of which they internalised the standards they use in assessing their lives. Why not just stipulate what these processes must be?” (WHE, 169). Because: how *exactly* is the distinction between ‘manipulative’ and ‘normal’ socialisation processes to be determined? as all of our aims, values and beliefs come from somewhere *outside of us*, and have been influenced *to some extent* by our particular personal histories, and the socio-cultural factors which have shaped our identities. “Roughly speaking, an autonomy-preserving [‘normal’] socialisation process will be one which does not erode the individual’s capacity for critical assessment of his values, including the very values promoted by the process itself” (WHE, 170). Which is all well and good: but how is the historical approach to be applied in cases where socialisation processes *were* ‘manipulative’ and a person’s resultant values, attitudes and beliefs therefore ‘tainted’ by the circumstances of their historical formation, but the person, following a period of genuinely critical reflection, comes to accept those values, attitudes and beliefs as *their own*? In such cases, and “[d]espite the fact that his values have a tainted history, we would not wish to exclude the possibility that his present endorsement of them counts as autonomous. In that case, the history of the initial formation of a value or goal cannot tell the whole story about its autonomy” (WHE, 170).

[6.2.18.] Sumner therefore concludes that “*neither* of the currently dominant theories about the nature of autonomy is self-sufficient. ... *However the details of a fully adequate view* [of wherein consists ‘autonomy’] *are worked out in the end*, the implications for our theory of welfare are clear. Self-assessments of [‘4th kind’] happiness or life satisfaction are suspect (as measures of well-being) when there is *good reason* to suspect that they have been influenced by autonomy-subverting mechanisms of social conditioning, such as indoctrination, programming, brainwashing, role scripting, and the like. Since these are all socialisation processes, and since we are all historically embedded selves, the practical question becomes *how much* emancipation from her background and social conditions a subject must exhibit in order for her self-assessment to be taken *at face value*. *As in the case of the information requirement*, the best strategy here is [simply] to treat subjects’ reports of their level of life satisfaction as *defeasible* – that is, authoritative *unless* there is evidence that they are non-autonomous” (WHE, 170-171; my emphases).

[6.2.19.] “On a subjective theory, individuals are the *ultimate authorities* concerning their own welfare. Their self-assessments are therefore *determinative* of their well-being [i.e. such assessments *instantiate* it as an *ontological* ‘fact’] *unless* they can be shown to be inauthentic, i.e. not truly theirs. The requirements that these assessments be [*circumstantially consensually sufficiently*] informed and autonomous spell out the conditions of authenticity. A

person's own view of her life satisfaction carries an *initial presumption* of authenticity, and thus of authority [*unless and until* evidence to the contrary becomes available]. It can be mistaken, or even deeply distorted. *But it must be shown to be so* before we have any ground for discounting it" (*WHE*, 171; my emphases again).

6.3. A Theory of Welfare (pp 171-183)

[6.3.1.] Sumner starts this section with a recap of the chapter so far: "... [T]he linkages between welfare and [4 kinds of] happiness [have been explored, so that 'welfare'] ... can be equated with life satisfaction [or '4th kind' happiness], which has both an affective component (experiencing the conditions of your life as satisfying or fulfilling) and a cognitive component (judging that your life is going well for you, by your standards for it). The best way of determining people's happiness levels is to ask them, and a sophisticated methodology for eliciting self-assessments of life satisfaction [as '4th kind' happiness] has been developed by social scientists. ... [A]n individual's report will accurately reflect his perceived ['4th kind'] happiness only if it is relevant ..., sincere ..., and considered The question then is whether ['4th kind'] happiness, as so measured, is identical to well-being. We have found two reasons for thinking that it is not: a person's self-evaluation may not be informed and it may not be autonomous. In either case it is inauthentic, in that it does not accurately reflect the subject's own point of view. Welfare therefore consists in authentic ['4th kind'] happiness, the ['4th kind'] happiness of an informed and autonomous subject" (*WHE*, 172).

[6.3.2.] This theory is subjective because it makes a subject's well-being dependent on their positive or negative attitudes. "The result [of the way Sumner's theory instantiates the four variables of a subjective analysis of welfare in section 2.2 of *WHE*] is an analysis on which some condition of a subject's life is (directly or intrinsically) beneficial for him just in case he authentically endorses it, or experiences it as satisfying, for its own sake. The intrinsic sources of welfare will be whatever conditions of subjects' lives elicit this response" (*WHE*, 172-173).

[6.3.3.] This theory of welfare as 'authentic happiness' preserves the analogy of welfare with perceptual properties (i.e. a cognitive/affective response by a welfare subject to their experience of 'states of the world' inducing in them various 'states of mind' out of which the inherently subject-relative prudential value of their lives can be determined). Some might think this analogy lays bare (seemingly) subjective theories of welfare to a charge of 'covert' objectivity (as, for example, "[s]pecifying the reference group of perceivers and the conditions of perception for a subjective analysis of redness thus appears to presuppose the logical priority of *being red* to *looking red*, so that the dispositional formula [of subjectivity] really analyses the latter in terms of the former" (*WHE*, 174)⁴).

[6.3.4.] A subjective analysis of welfare as 'authentic happiness' can be defended from this charge because "[t]he aim of such an analysis is to ground an individual's well-being in her attitudinal point of view on the world. This has been accomplished only when a subject's responses to the conditions of her life are authentic, i.e. truly hers. The requirements that her responses be informed and autonomous flow from this basic demand of authenticity. If a subject's endorsement of some particular (perceived) condition depends on a factual mistake, or results from illusion or deception, then it is not an accurate reflection of her own underlying values. And if those values have been engineered or manipulated by others then they are not truly *hers*" (*WHE*, 174). Prudential standards autonomously *chosen* or *accepted* by a welfare subject (following their experience of various 'states of the world'), and which are therefore 'authentically' *hers*, are *necessarily* subjective (thus excluding 'covert' objectivity from Sumner's theory of welfare).

[6.3.5.] Sumner's 'happiness theory' of welfare (as he refers to it) avoids the solipsism of (classical) hedonism and the disengagement from our lived experience implicit in the desire theory, constituting 'something in between' these two types of subjective welfare theories. But is it *superior* to them? To decide this we must review the features of the 'happiness theory' against the various criteria of 'descriptive adequacy' for *any* theory of welfare.

[6.3.6.] *Fidelity*. "[A theory of welfare] should provide an interpretation of [the] subject-relativity [of welfare as prudential value *for the individual whose life it is*], by explaining what it means for a life to be going well not just in itself or from some other standpoint but *for its subject*. This is, of course, the demand that objective theories of welfare are

⁴ A fully subjective account of our perception of the ontological quality of *redness* reverses the normal order of explication, i.e. "it's 'red' just because we all think 'redness' is what we're perceiving" (which is circular), rather than "we know it's 'red' according to objectively ascertainable and measurable properties of 'redness', which is why 'redness' is what we're perceiving".

unable to meet. Subjective theories satisfy it by referring to the subject's attitudinal point of view; in the case of the happiness theory, that point of view is made up of the values or standards which the subject uses in determining whether her life is satisfying or fulfilling. The authenticity requirement, which is an essential part of the happiness theory, guarantees that the operative point of view in a subject's self-assessments of her happiness is genuinely hers" (WHE, 175).

[6.3.7.] "Determining just how well the happiness theory fits our many preanalytic judgements about welfare would be a complex, and probably endless, task. However, it is possible to point to the ways in which its fit is superior to that of its subjective rivals. ... The happiness theory avoids [hedonism's preoccupation with 'pleasure' and 'pain', or 'enjoyment' and 'suffering', and its mental statism] by shifting attention to the more global attitude of ['4th kind'] happiness or life satisfaction and by introducing an information requirement *which is capable of overriding* subjective reports of happiness. On the other hand, the desire theory failed [because it lacks an experience requirement, opening a logical gap between desire-satisfaction and *personal* satisfaction]. By incorporating an experience requirement, and shifting to the notion of *personal* satisfaction [a phrase Sumner uses interchangeably with 'life satisfaction'], the happiness theory is able to distinguish between those desires which, when satisfied, make a subject's life go better and those which do not ... thus [enabling] the happiness theory to make better sense of *at least many* of our judgements about well-being" (WHE, 176; my emphases).

[6.3.8.] "A further component of fidelity was adequate fit with the role which welfare plays in our common-sense psychology, and especially rejection of the thesis of psychological egoism. ... [The happiness theory avoids reduction of behavioural motivations to psychological egoism via] the relevance condition for subjects' reports of their own happiness. Subjects are being asked to reveal how well they feel their lives are going *for them*, not for self-assessments in any other value dimensions (except in so far as their self-rating in these dimensions affects their level of life satisfaction). The personal values which are pertinent to an individual's ['4th kind'] happiness are prudential ones. An autonomous agent is therefore free also to embrace aesthetic, perfectionist, or ethical ideals whose pursuit may conflict with her well-being" (WHE, 176-177).

[6.3.9.] *Generality*. "... [A] theory of welfare must be general or complete, by virtue of covering all of the different sorts of welfare assessments we make ... [like] those which situate a subject at some particular welfare level and those which chart a subject's gains or losses (movement from level to level). ... [The happiness theory supports such assessments because] people seem capable of determining both how satisfied they are with their lives and whether some particular development has made them more or less so. ... [The happiness theory's] analogous treatment [of unhappiness] is straightforward, resting as it does on the negative counterparts of *personal* satisfaction and endorsement. A life is therefore going badly for someone when she (authentically) experiences its conditions as unsatisfying or unfulfilling, or disclaims or disowns them" (WHE, 177; my emphasis).

[6.3.10.] "Our everyday welfare assessments also include interpersonal comparisons, both of levels of well-being and of gains and losses. ... *How these various kinds of comparisons are to be interpreted and supported is too large a topic to be dealt with adequately here*. ... [Subjective accounts have been thought to make interpersonal comparisons of welfare impossible, but] [s]uch accounts are now commonplace in the research into ['4th kind'] happiness or life satisfaction carried out by the 'subjective indicators' school. Ordinary people always seem to have managed them easily on a daily basis. What we rely on in the straightforward cases is a rich vein of information, both verbal and non-verbal, about the way other people's lives are going for them. The hard cases are ones ... [where] we find the other relatively opaque, as when we attempt to compare ourselves with members of very different cultures or with people who have serious mental or physical disabilities. Where we are unable to represent to ourselves how the life of the other is for him, from the inside, then we may be at a loss to decide whether that life is better or worse, for him, than ours is for us" (WHE, 178; my emphases).

[6.3.11.] "... [O]pacity is exacerbated when we move beyond the realm of human subjects. ... [Generality requires a] plausible rationale for delimiting the class of core welfare subjects, plus some illuminating explanation of what makes other subjects [of-a-life] peripheral. For the happiness theory, the minimum wherewithal for having a welfare is being a subject who is capable of being satisfied or unsatisfied by the conditions of one's life. In the case of paradigm human subjects with complex cognitive capacities, more is necessary as well: their judgements about the quality of their lives

must be authentic. Where these more sophisticated skills are absent, the *sine qua non* is the baseline ability to experience one's life, in the living of it, as agreeable or disagreeable. The most primitive form of this ability is the capacity for enjoyment and suffering, or for pleasure and pain. If we call this capacity *sentience* then we may say that on the happiness theory the class of core welfare subjects is populated by all sentient creatures" (WHE, 178). {"If a creature is incapable of acting inauthentically, then the authenticity requirement is trivially satisfied. Can non-human animals respond inauthentically to the conditions of their lives? Can this be true, for instance, in the case of domesticated animals whose affective responses have been deliberately engineered, for our convenience, so as to lead them to be satisfied by ways of life which are [otherwise] unnatural to them? If so, then the authenticity requirement will apply (non-trivially) in those cases as well, and it may not be safe to conclude that an animal is faring well just because it is [seemingly '4th kind'] happy" (WHE, ch. 6, fn. 47).}

[6.3.12.] Creatures which the happiness theory consigns to the periphery of welfare subjectivity (because they lack sentience) include: lower-order (non-sentient) animals and plants, human embryos and fetuses up to a certain stage of prenatal development, and persons who have permanently lost the capacity for conscious awareness (because they are in a persistent vegetative state). The happiness theory also excludes the dead and collectivities from welfare subjectivity (as they do not have 'lives' which they can find satisfying or unsatisfying; they are not 'subjects-of-a-life').

[6.3.13.] *Formality*. "[T]he happiness theory offers ... an account [of what makes something a source of well-being]: a condition of someone's life counts as an intrinsic source of well-being for her just in case she authentically endorses it, or finds it satisfying, for its own sake. The subject-relativity of prudential value requires the reference to the subject's own endorsement. However, it does not prevent us from generalising over the conditions which will standardly be experienced as intrinsically rewarding or fulfilling by creatures who share a common nature (such as us). The happiness theory is therefore capable of generating a set of *standard human goods*: such items as health, mental and physical functioning, enjoyment, personal achievement, knowledge or understanding, close personal relationships, personal liberty or autonomy, a sense of self-worth, meaningful work, and leisure or play. These items should look familiar, since they commonly turn up on the lists of human goods invoked by objectivists. Each of these goods has enormous instrumental value for us, since each is the condition of realising many other goods in our lives (for one thing, the items on the list are mutually interdependent). And each is something whose relative value in our lives may vary considerably from person to person: you may place much greater store on achievement or success and less on maintaining close personal relationships than I do. But for each item on the list it is plausible to say that everyone cares about it *to some extent* for its own sake, thus that its presence in a life makes that life *to some extent* more satisfying or fulfilling, quite apart from its instrumental value. If asked why it should be that just these goods achieve the status of intrinsic sources or constituents of our well-being, the only intelligible answer is: because that is the kind of creatures we are" (WHE, 180-181). {"The analogous list of intrinsic prudential goods for members of other species will likewise be determined by their nature" (WHE, ch. 6, fn. 53).}

[6.3.14.] "[The formality of subjective theories means that, unlike objective theories, they are] able to explain why the list of standard prudential goods contains just these items and not others. They are also able to explain why the relative importance of these goods can vary so much across individuals as well as cultures. ... The root problem with [objective] theories is the same: they all attempt to build a formal theory about the nature of welfare around one of its standard intrinsic sources. Only the happiness theory provides a general rubric capable of explaining the prudential significance of all of these goods, as well as many others" (WHE, 181).

[6.3.15.] *Neutrality*. "... [R]equires that a theory not exhibit any bias in favour of some particular list of goods or some favoured way of life. Objective theories have difficulty with this requirement, since they typically stipulate a pattern of the good life for all members of a particular natural kind (such as us). ... [This] is generally not an issue [for subjective theories]" (WHE, 181), unless they incorporate some objective elements (like ranking forms of life on some independent non-prudential scale, thus becoming 'hybrid' theories of welfare). "The happiness theory avoids [the] trap [of violating neutrality by 'rigging' its normalising conditions so as to ensure selection of some predetermined set of goods] by deriving its information and autonomy conditions from an authenticity requirement, which is in turn entailed by subjectivity itself. It will therefore endorse as prudentially valuable whatever ways of life are found to be fulfilling by informed and autonomous subjects. No independent restrictions of content are imposed" (WHE, 181-182).

[6.3.16.] “[Under] ... a neutral theory of welfare [like the happiness theory] ... ways of life which we regard as trivial or demeaning or depraved can ... be prudentially valuable for their subjects ... as there is no way of excluding the possibility, for any such way of life, that it can be endorsed by an informed and autonomous subject. ... [While] a life of servility or subservience is rarely embraced under conditions of full information and autonomy⁵ ..., [i]n other cases, such as a life devoted to watching daytime talk shows or pursuing a career in the sex trade, we may have to discipline ourselves not to make *a priori* assumptions about the degree of autonomy manifested by those who make choices which we find personally distasteful. ... [A]n open mind and a willingness to attend to the specifics of people’s particular circumstances seems less patronising than simply assuming that [their] lifestyle ..., besides failing to measure up to some other value dimension, is also necessarily bad for them” (*WHE*, 182).

[6.3.17.] Sumner concludes that his ‘happiness theory of welfare’ “scores no worse than any rival [objectivist, hedonistic or desire] theory [of welfare] on any criterion [of ‘descriptive adequacy’⁶] and better than all of its rivals on most. It therefore provides us with the best picture of the nature of welfare” (*WHE*, 183).

7. Welfarism (pp 184-223)

[7.1.] Having the best theory of welfare in hand (i.e. Sumner’s ‘happiness theory’) allows us to turn from an examination of the *nature* of welfare, to its ethical *value*.

[7.2.] ‘Welfarism’ (a normative view of the foundations of morality, which posits that nothing but welfare matters, basically or ultimately, for ethics) has fallen out of favour, in large part because the various other theories reviewed in *WHE* make it hard for moral philosophers to accept that any of *those* conceptions of welfare could possibly be the whole point of the ethical enterprise.

[7.3.] Sumner asserts that welfarism deserves to be tested on the basis of the best available theory of the *nature* of welfare (i.e. his ‘happiness theory’ of it). He states that opposition to welfarism is best understood in terms of opponents’ reactions to various of its (facilitative) methodological assumptions.

[7.4.] These methodological assumptions are as follows.

Ethical realism. An epistemological standpoint which says that there are ‘right answers’ to ethical questions.

Ethical theory. The belief that theory-building in ethics is neither futile nor misguided (as it helps you ask and answer the ‘right questions’).

Foundationalism. All other categories or dimensions of value are either derived from, or justified by, prudential value (so only prudential value is *truly* foundational).

The priority of the good. The purpose of ethics is to bring about ‘good’ states of the world (by furnishing answers to the ‘right questions’).

Agent-neutrality. All ‘moral agents’ have the same reason to care about prudential value (because prudential value is the *currency* of well-being).

[7.5.] “Someone who shares [all of these methodological assumptions] thinks that the foundation of ethics consists in a *theory of the good*: a list of foundational agent-neutral values whose pursuit and promotion [via ‘realist’ ethical theory-building] is the point of the whole ethical enterprise. Welfarism is a theory of the good whose list of foundational values contains a single item: well-being. ... [W]e will concern ourselves with only one class of anti-welfarists: those who share all of the welfarist’s assumptions about ethics but support a rival theory of the good. ... A theory of the good can take [one of] two possible forms: it is *monistic*, if it includes only one item on its list of foundational values, or *pluralistic*, if it makes room for more. Welfarism is monistic; its rivals are therefore both other monisms and any form of pluralism” (*WHE*, 185-186).

⁵ I believe the qualifier ‘full’ applies to the noun phrase ‘information and autonomy’, not the noun ‘information’, and is to be read as synonymous with the adjectives ‘adequate’ or ‘sufficient’, rather than ‘comprehensive’ or ‘total’.

⁶ So that “the ideal theory of welfare ... will be faithful to our ordinary assessments of well-being, including the role they play in our common-sense psychology, it will cover all core cases and provide a principled resolution of peripheral cases, it will not confuse welfare with its sources or ingredients, and it will be free of distorting bias” (*WHE*, 182).

7.1. The Case for Welfarism (pp 187-200)

[7.1.1.] Sumner begins this section by asking: how is one to make a case for welfarism? J.S. Mill's 'psychological hedonism' won't work, as it is "a commonplace that people are psychologically capable of valuing and pursuing ends other than happiness (either their own or that of others) [so appeals to psychological egoism cannot suffice]" (*WHE*, 187).

[7.1.2.] Sumner also rejects 'practical reason' (which is about what constitutes a *rational basis* for ethics) as a justification for welfarism. Some philosophers confuse the *self-interest theory* of practical rationality (which says that individuals are behaving 'rationally' when they *maximise* their own welfare) with its *instrumental theory* (which says that individuals will *therefore* maximise desire-satisfaction). But (as Sumner has demonstrated earlier in *WHE*): well-being and desire-satisfaction are *logically distinct, not equivalent* (and these two theories of practical reason are in any case incompatible, in that the former assumes a single unique 'end' of 'their own welfare' for each individual, while the latter is concerned with 'means' *not* 'ends'). This is "easy to see ... [in the case of] someone who knowingly makes himself worse off in pursuit of some rival [non-prudential] category of value: the ethical, perhaps, or aesthetic, or perfectionist [categories]. We do not normally regard such persons as irrational, or even unreasonable, unless their degree of self-sacrifice is carried to the extreme that they become a danger to themselves" (*WHE*, 188).

[7.1.3.] Practical rationality is about reasons for action, "reasons [which] can emanate from distinct, independent, and possibly incommensurable points of view, of which ethics is but one. Besides ethical value we can be called on to protect or promote other dimensions of value ... [and, in the case of a conflict between values] practical reason appears to furnish no higher court capable of adjudicating the conflict, and no higher mode of value capable of subsuming these particular standpoints. Pluralism therefore seems to be the right story about practical rationality. So why [then is pluralism] not also the right story for ethics?" (*WHE*, 188).

[7.1.4.] Because: a pluralist account of what matters for practical reason cannot be symmetrical with what matters for ethics due to the essential *subjectivity* of welfare (which is ultimately about the *prudential* value of a life for the person living it; to the extent that someone pursues some other mode of value, such as aesthetic value, to the exclusion even of their own 'happiness', the concomitant increase in the aesthetic value of that person's life from some external standpoint doesn't compensate *them* for *their* loss of well-being).

[7.1.5.] "Furthermore, we should not suppose that the rules of a welfarist morality will forbid individuals to make non-welfarist choices. Any reasonable set of arrangements will protect autonomy, among other values, as an intrinsic prudential good ... ensur[ing] that individuals have sufficient space in which to set their own priorities for their lives, imposing limits only at the margins (when others will be adversely affected, when you become a danger to yourself, when you are not competent to make such choices, etc.). ... A welfarist ethics ... should be flexible enough to make room for idiosyncrasy, eccentricity, even a little craziness" (*WHE*, 189-190).

[7.1.6.] So, instead of arguing for welfarism from psychology or rationality, Sumner sets out some relevance conditions for choosing between (different, sometimes competing) values. If *prudential* value really is the *only* value ultimately worth pursuing or promoting via ethics, as welfarism says it is, then this (or any other supposedly foundational) mode of value should be able to satisfy these conditions.

[7.1.7.] The relevance conditions are as follows.

Intrinsic. "[W]orth having or pursuing for its own sake, not merely by virtue of some further good with which it is connected or associated ... [thus] eliminat[ing] anything whose value is merely instrumental or contributory" (*WHE*, 190).

Abstract or generic. Our ethical concerns range over so many ethical 'goods' in so many contexts, "[i]f there is a single foundational value then it must be broad enough to encompass all of [our ethical 'goods', such as liberty, autonomy, equality, sociality, loyalty, intimacy, security, health, achievement, enjoyment and so on] and to explain their appeal. Nothing short of an abstract category or mode of value will have the requisite degree of generality" (*WHE*, 190).

Important. “Moral considerations are advanced as constraints on our wishes or desires, as demands we must heed even when we would prefer not to. If these constraints and demands are all in the service of some category of the good, then the case for promoting this good had better be pretty compelling” (WHE, 191).

Ethically salient. A relation between a value and the ethical theory for which it is foundational. As the content of ethical theories includes such ‘weighty’ matters as rights and duties, blame and punishment, guilt and shame, and so on, “[n]ot all values ... can bear this particular [foundational] weight. ... If there is a foundational value for ethics, then [that value] must be not only worth pursuing for its own sake but also capable of supporting ... [the] structure of judgements and practices [characteristic of ethics]” (WHE, 191).

[7.1.8.] Welfare meets all four of these relevance conditions (as benefits and harms matter in their own right, it embraces a wide variety of goods as sources of well-being, it tracks whether people’s lives are going well or badly, and the fact that something makes someone’s life go better or worse for them counts for or against that thing).

[7.1.9.] “Welfarism also gives a credible account ... of the conditions under which [ethics] is both possible and necessary. These conditions are ... (1) a set of *moral agents* who possess the requisite rationality and autonomy to recognise and act on moral reasons, thus to regulate their lives by means of ethical standards, and (2) a set of *moral patients* who must be taken into account in the deliberations of moral agents. Welfarism ... give[s] us a *criterion of moral standing* for determining who qualifies as a moral patient. To have moral standing is to count or matter morally in one’s own right or for one’s own sake. Welfarism entails that moral standing is shared by all creatures with a welfare, thus (on the happiness theory) all sentient creatures” (WHE, 192).

[7.1.10.] “[W]elfarism [based on the happiness theory] also tells us that the existence of rational agents is not sufficient by itself to make moral thinking possible. It is conceivable that such agents might be utterly lacking in affect (they might be superintelligent computers, for instance), thus quite incapable of finding their lives (if that is the appropriate term for [their existences]) either satisfying or unsatisfying. If that were the case, then nothing could go either well or badly for them. If the world were lacking in any other sentient creatures then, whatever other intellectual abilities these rational beings might have, ethics would lie beyond their range: there would be nothing for moral thinking to be *about*. *The preconditions of ethics include not merely agency but also sentience*” (WHE, 192-193; my emphasis on the sentence ‘The preconditions of ethics include not merely agency but also sentience’).

[7.1.11.] “As a monistic theory of the good, welfarism claims that (1) welfare matters (ultimately and for its own sake), and (2) nothing else does. It is the second claim that is the sticking point for most critics of welfarism” (WHE, 193), many of whom prefer pluralist theories of the good. “Against this pluralistic alternative, welfarists are forced onto the defensive, arguing for each such additional good that it does not deserve a foundational role in ethics” (WHE, 193).

[7.1.12.] Applying his four relevance conditions to various modes of value, Sumner concludes that welfare’s only serious contender as a monistic axiological foundation for ethics is *perfection*. “To say that something has perfectionist value is to say that it is a good instance or specimen of its kind, or that it exemplifies the excellences characteristic of its particular nature. ... [P]erfection is objective; the criteria which determine the perfectionist value of a life are derived entirely from the natural kind to which the subject of a life belongs, and not at all from her own attitudes or values. ... [P]erfection seems ... to be intrinsically valuable, it is a broad, abstract category of value which embraces and supports many more specific goods, it is seemingly important, and it is plausible to think of it as ethically salient. Could a case then be made for perfectionism as a monistic theory of the good?” (WHE, 194).

[7.1.13.] Sumner concludes that it can’t, as “[a] monistic perfectionism attributes no value to happiness or satisfaction or enjoyment (however authentic), except in so far as these conditions are indicators or accompaniments of proper (mental or physical) functioning. ... Worse ... [monistic] perfectionism cannot recognise pain or suffering as intrinsic evils; indeed, it has no room for the very concept of an intrinsic evil. ... No theory which fails to find a place for well-being (and ill-being) could possibly tell the whole story about the good” (WHE, 195).

[7.1.14.] Lastly, Sumner responds to Thomas Nagel’s argument that welfare does not have ‘agent-neutral’ value. Nagel’s argument is that there are two generic categories of prudential goods: physical (dis)pleasure or (dis)comfort (the value of which is ‘agent-neutral’), and the achievement (or frustration) of aims or goals (the value of which is necessarily

'agent-relative', as achievement of *personal* goals is something welfare subjects want *for themselves*). If Nagel is right, and any aspect of welfare is agent-relative, then welfarism (which assumes agent-neutrality) is defeated.

[7.1.15.] But: Nagel's argument relies on a form of desire-satisfactionism (i.e. that desire-satisfaction in relation to our aims or goals cannot be agent-neutral because of the *heterogeneity* of tastes and preferences, and so cannot have agent-neutral *value* either; but desire-satisfactionism lacks agent-neutral value *because getting what we want does not reliably correlate with welfare*). When desire-satisfaction *does* (correlate with welfare), then the reason that I have to help you achieve your aims or goals (and that I have for wanting your help in the achievement of mine) is *because* it will make our lives go better (so that the promotion of prudential value for welfare subjects is agent-neutral).

[7.1.16.] Sumner argues that what motivates Nagel's concern to differentiate 'categories' of prudential goods are the "foreseeable implications of an ethical theory ... according agent-neutral value to the achievement of aims (where this has prudential value) ... result[ing] in unreasonable burdens being imposed on others. ... Welfarism works at a prior stage [to the development of an ethical 'theory of the right']: the formulation of a theory of the good. It seems premature to contour [a 'theory of the good'] to forestall unwanted deontological implications, especially if doing so seems to rely on ... an arbitrary distinction between equally important generic sources of welfare" (*WHE*, 198).

[7.1.17.] Rejecting that distinction, "there can still be further reasons for questioning whether all prudential goods have positive ethical value. What about happiness or satisfaction derived from a lifestyle which is trivial or demeaning or degrading? ... On the happiness theory a lifestyle genuinely enhances a person's well-being only when it has been endorsed by her under conditions of *adequate information* and by means of values or standards which she has autonomously adopted. These authenticity conditions are themselves sufficient to exclude many cases of repugnant lifestyles, since they will not count as prudentially valuable for their subjects. ... But not all ... [yet] [h]owever distasteful we may find her choice, for all we know it is the lifestyle which will be most fulfilling for that person in her full particularity. ... A step further takes us to forms of satisfaction which are morally repugnant ... [although] [t]he authenticity conditions [of the happiness theory] may ... suffice to screen out many of these unpalatable desires as sources of well-being: in particular, we may doubt the extent to which they can be autonomous. ... These kinds of satisfaction [such as rape or torture] stretch welfarism's toleration to its limit: if our aim is to build a welfarist moral theory, how can we admit such tainted sources of welfare onto its ground floor? The solution to this problem, I think, lies once more in reminding ourselves how little welfarism determines of the structure of a moral theory: while it tells us wherein the good consists, it does not dictate how the right is to be derived from it. ... [A variety of different moral theories can address] the moral problem raised by evil appetites elsewhere, in [an] account of the right. In general, this seems the appropriate way to deal with [the problem of evil appetites], rather than by launching a pre-emptive strike at the deeper level of the good" (*WHE*, 198-200; my emphasis).

7.2. Pluralism (pp 200-217)

[7.2.1.] Pluralism maintains that there are many irreducible foundational goods. Sumner asserts that it is "tailored to fit the modest ambitions of recent moral and political philosophy which ... has turned its back on grand, comprehensive, reductive, universalistic theories. ... Pluralism ... [can] accommodate or adjust to any shifts in the prevailing social or philosophical winds; it is the low church of ethics and politics" (*WHE*, 200-201). Pluralism is (seemingly) a 'hard target' for a monistic theory of the good (like welfarism) to defeat.

[7.2.2.] Rival goods can be divided into two categories: personal goods (i.e. valuable states or activities realised within the lives of welfare subjects, such as health, pleasure or enjoyment, achievement or success, rationality, knowledge or understanding, close personal relationships, safety or security, risk or adventure, liberty, autonomy, self-worth, meaningful work, and leisure or play), and impersonal goods (belonging to entities other than welfare subjects, and manifesting in their lives as functioning or flourishing, values such as collective achievement or self-determination realised for groups, and environmental values including the survival of species and the integrity and stability of biotic communities).

[7.2.3.] When considering these rival *personal* goods or values, the welfarist can either co-opt each as a welfarist good or deny that its pursuit as an 'end' in itself has any place in ethics. Co-option is the most promising strategy (as any

personal value which could plausibly be advanced as basic or foundational for ethics will *also* be a standard and intrinsic source of well-being).

[7.2.4.] By contrast, the pluralist believes that these personal goods are basic ethical values, alongside welfare (and that they are therefore worth promoting *for their own sake and independently of the extent to which they make our lives go well*). Is this a plausible view? Sumner thinks not: “I can find no ethical value in promoting achievement or knowledge or liberty, or any other personal good, if *no one at all* will be better off for it: neither the person whose good it is nor anyone else. To my mind, the value of these states of affairs is adequately captured by the role they play in enriching our lives; there is no remainder which requires independent acknowledgement beyond this prudential payoff” (*WHE*, 202). Pluralists need to be able to respond to Sumner’s point.

[7.2.5.] The force of Sumner’s point is easier to see when contemplating a test case. A suitable candidate is Sen’s version of pluralism which proposes two co-equal foundational values: well-being, and agency. Sen’s case for pluralism rests in part on the possibility of conflict between the two values, as in the hypothetical example of your unconscious friend who will be administered either ‘treatment A’ or ‘treatment B’ on your say so, both of which are equivalently effective, although ‘treatment A’ will engender fewer side-effects for your friend. However, ‘treatment A’ is associated with animal testing, to which you know your friend is morally opposed. In fact, you know that your friend would agree that ‘treatment A’ is better for his well-being, but as a free agent he would choose ‘treatment B’, if he were able to exercise this choice. If you authorise ‘treatment B’, you are attaching an independent value to your friend’s agency, which Sen thinks is a reasonable thing to do (given what you know about your friend’s preferences).

[7.2.6.] But: “[i]n order for agency to function here as a strictly non-welfarist good [and so be co-equal with welfare], we need to suppose that your friend would not regard his life as going worse for him *in any way* were he to discover subsequently that his agency had been overridden or compromised. ... [If this is psychologically] coherent then it seriously weakens your reason for choosing treatment B for him. Now you must think that this choice will not just be worse for him on balance (as it would be if the additional suffering were great enough to outweigh the benefits of preserving his agency) but a dead prudential loss. It is one thing for your friend to choose to sacrifice his own interest in this way for an ethical cause, and quite another for you to make this choice on his behalf. Your friend has two commitments: to his own well-being and to the cause. You have only your commitment to him; the ethical cause matters to you only to the extent that it matters to him. If protecting what matters to him does not make his life go better, then you have no further reason to care about it. Being dedicated to doing what is best for your friend, all things considered, will normally require you to take his commitments seriously, just because they are his. But that is because doing so will normally make him better off. Where this condition fails, where honouring his commitments will not be a good for him, you have only the remaining aspects of his well-being to take into account” (*WHE*, 204-205). In which case, your friend’s ‘agency’ is not really a co-equal good with his ‘welfare’, and Sen’s pluralism collapses.

[7.2.7.] Agency is a concomitant of autonomy, and autonomy is “something whose presence in our lives makes them go better in itself. *But it is also one of the conditions of anything having prudential value, since it is embedded in the nature of welfare*. A person’s endorsement of the conditions of her life is determinative of her well-being only when that endorsement is authentic, and it is authentic only when it is autonomous. In the absence of autonomy, a person’s welfare is *indeterminate or unknowable*. Fostering autonomy is therefore necessary, not only as one particular way of enhancing people’s welfare, but also as a condition of *knowing wherein their welfare consists*” (*WHE*, 205; my emphases).

[7.2.8.] Sen’s version of pluralism reminds us of the inevitability of conflict among goods in value pluralism (such conflicts are a familiar feature in the landscape of practical reason). Conflicts in the domain of practical reason can be ignored when we are doing ethics, except to the extent that they affect well-being. But what of conflicts among goods or values within the ethical domain? “Welfare is not some overriding or higher-order value to which we can appeal in order to resolve conflicts among more local goods; rather, it is the outcome when we have settled our priorities among such goods. Welfarism feeds information about individuals’ interests into an ethical decision-making procedure (whatever it may be), but since welfare is subjective the determination of those interests is left (within the limits defined by the requirements of authenticity) to those individuals. ... As welfarists all we can, and should, say is that the

choices a subject makes should be as *informed and autonomous as possible* – that is, that they should genuinely reflect the subject’s own priorities” (WHE, 206-207; my emphasis).

[7.2.9.] Moreover, “[i]f non-welfarist goods are considered item by item, then it is not difficult to make their ethical value seem to depend on the extent to which they make our lives go better for us. The pluralist would have a more convincing case if her list of basic values, to be added to welfare, were less eclectic and piecemeal. This would be true if there were some other evaluative dimension, just as abstract and generic as prudential value, to which appeal was needed in order to capture some of the value of these personal goods. ... We are therefore led once more to perfectionist value as the only viable candidate. ... [H]ere we encounter a more modest and circumspect version of the view. The perfectionist can claim, with some plausibility, to have a rival story to tell about the value of personal goods. Whatever direct contribution they may make to our well-being, they also have intrinsic perfectionist value because their presence in a life *makes it a better life of its kind*. ... [As a foundational value in a *pluralist* theory of the good] this is ... as plausible as perfectionism gets” (WHE, 208-209).

[7.2.10.] For the welfarist to reject a pluralist theory of the good which fuses welfare with perfectionism as its foundational values requires the welfarist to show that there is some salient difference between perfectionist and prudential value. And this difference is: *whose* good is being promoted by perfectionism? Answer: anything with a *nature*. “An account of the nature of a thing, we are to suppose, will identify certain properties which are essential to it – properties in the absence of which it would not be a thing of *that* kind. From this account it will then be possible to extract criteria for determining how well a particular thing measures up to the standards of its kind. However this very abstract schema comes to be filled out in detail, it will clearly apply across the full range of natural kinds. Inevitably, therefore, many things will be candidates for perfectionist value which could not possibly have a welfare” (WHE, 209).

[7.2.11.] “Part of the problem with perfection as a fundamental value for ethics is that as a category of the good it settles on just about everything. ... Excellence or perfection is ... relative to kinds. Thus it is possible, in principle at least, to compare things belonging to the same kind in terms of their perfectionist value. However, we are given no resources for carrying out comparisons across kinds ... [and] [w]hen a category of value is distributed so widely, ... conflicts will be legion. How are any of them to be adjudicated?” (WHE, 212-213).

[7.2.12.] In any case, “[t]he superabundance of perfectionist value entails that there will often be no ethical point to promoting or protecting it. ... We already know that individuals can face a choice between their perfection and their welfare: what makes them better persons may fail to make them better off, if it gives them no satisfaction or fulfilment. When you are choosing between such options for yourself, then the sacrifice of either to the other seems intelligible; we can understand the point of the choice in either case, and we may take the view that it is simply your decision to make. (On the other hand, if you seem to be too willing to sacrifice your well-being in the pursuit of some perfectionist ideal, we may stop being so supportive.) In cases of personal choice, where we are determining the array of goods within our own lives, we all assume that individuals enjoy a broad prerogative within whose limits the balance to be struck between perfection and welfare (and other goods as well) is simply up to them” (WHE, 213-214).

[7.2.13.] “But we must also make choices whose impact will fall on the lives of others. If we adopt a welfarist/perfectionist pluralism then we will sometimes sacrifice other people’s welfare for the sake of their perfection. ... Their perfection is in no way determined by their interests or concerns. ... [Such] [o]bjective values are quite literally alien to us because they emanate from a standpoint which is external to us as individuals, and because their status as values requires no affirmation or endorsement of them on our part ... [which] infringe[s] autonomy or individuality. ... The deep problem for any objective theory is that personal concerns play no role in determining why something (anything) counts as a good for an individual in the first place, or why one thing counts as a greater good than another. Any such theory will therefore be committed in principle to overriding the autonomous choices of individuals concerning their own lives, imposing on them what they themselves value less” (WHE, 214-215).

[7.2.14.] Moving on to *impersonal* goods, denial is welfarism’s only option (because ‘welfare’ is restricted to sentient creatures). The only serious question raised by this denial is: does welfarism’s ‘individualism’ (i.e. that only individuals can be a welfare subject, and only welfare subjects are capable of having interests) draw the circle of moral standing too narrowly, by excluding too many things from consideration?

[7.2.15.] Sumner thinks not because “the only good which we can promote for non-sentient organisms is their perfection, rather than their welfare ... [however] to deny intrinsic value to the lives of non-sentient organisms is not to deny them all value. In most cases, we will have (derivative) welfarist reasons for protecting and preserving these organisms, especially when we take into account [the] interests ... of all sentient creatures” (*WHE*, 215-216).

[7.2.16.] “Welfarism implies that collective goods have no intrinsic value; instead, they are worth fostering only to the extent that they enrich the lives of individuals. ... The flourishing or survival of a natural species or of a biotic community, or the integrity and stability of the ecosystem, ... are values worthy of our support when, as is usually the case, they are means of ensuring the welfare or survival of sentient beings. Where they are not, they have no ethical value, whether intrinsic or instrumental” (*WHE*, 216-217).

[7.2.17.] Sumner regards the individualism of welfarism as its greatest strength: “that nothing which requires ethical notice has occurred unless someone’s (or something’s) life has been made to go well or badly ... [has] enormous intuitive appeal The localisation of welfare to individuals provides a ... tolerably clear ... criterion for the scope of its ethical domain, which is capable of differentiating it from other realms of value (such as the perfectionist or aesthetic). ... Moral considerations matter because it matters how well lives are going, and how well the lives of other [sentient] creatures are going must matter to us *because* it matters (from the inside) to those [sentient] creatures” (*WHE*, 217).

7.3. Welfarist Ethics and Politics (pp 217-223)

[7.3.1.] “In affirming a welfarist foundation for ethics we have remained studiously agnostic about the further shape of an ethical theory. ... I do not intend to abandon this agnosticism at this late stage in the argument. *The development of a full welfarist ethical theory is a task for another occasion.* Instead, I want to consider the possible impact of welfarism on some prominent issues in political theory” (*WHE*, 217-218; my emphasis).

[7.3.2.] The traditional view (that ethics is logically prior to politics) has come to be challenged (by those who claim that ethical values *themselves* presuppose antecedent social and political arrangements, and are therefore *ideological* rather than purely axiological). Sumnerian welfarism, based as it is on a subjective (‘happiness’) theory of welfare, can be neutral among different ways of life and cultural traditions (except for the operational importance of autonomy in that theory, which Sumner’s welfarism privileges over other prudential goods).

[7.3.3.] Autonomy is the exception “... because where the social conditions for the autonomous adoption of personal values are lacking, *we cannot know wherein people’s well-being consists.* There is therefore a deep affinity between welfarism (when coupled with a subjective theory of welfare) and the liberal conception of the person as self-determining and self-making. Once autonomy has been recognised as a foreground value then it seems inevitable that defence of the traditional liberal freedoms (especially freedom of expression) will become a high priority for a [Sumnerian welfarist] political theory” (*WHE*, 218-219).

[7.3.4.] In part, this is because Sumnerian welfarism “... capture[s] ... one very powerful idea concerning the relationship of ethics to politics, namely that the state needs to justify its policies by showing how they will make the lives of its citizens go well, and that these citizens are the final authorities on when this condition has been satisfied” (*WHE*, 220).

[7.3.5.] Sumnerian welfarism also accords with liberalism’s acceptance of, and respect for, widely divergent lifestyles (as it reaffirms the agent-neutrality of ‘welfare’ as a good, while making sense of why the contribution various intrinsic *sources* of welfare will make to people’s lives is necessarily agent-relative).

[7.3.6.] However, Sumnerian welfarism does not lend itself to serving as G.A. Cohen’s ‘currency of egalitarian justice’, because (outcome) equality of ‘welfare’ ignores the heterogeneity of people’s conceptions of the ‘best life’ *for them*, and ‘welfare’ is simply ‘the good’ to be promoted or advanced by ethics and/or politics (so that welfarism is completely silent on how ‘welfare’ should be *distributed*, lest it violate agent-neutrality). “From the fact that making people’s lives go better is the ultimate rationale and justification for liberal policies, it does not follow that well-being must figure directly in [their] formulation In principle at least, a welfarist liberal state ... can fall back on the value of fostering

what we have called standard goods: those goods which correlate reliably with well-being, though their role and importance may vary widely from person to person. *It is some mix of these goods* which will then constitute the currency of egalitarian justice" (*WHE*, 223; my emphasis).