Interpersonal Comparisons of the Good: Epistemic not Impossible

Utilitas 2015

MATHEW COAKLEY

London School of Economics & Political Science

Note: this is a pre-publication draft[1].

Please only quote from published Utilitas version.

Abstract

To evaluate the overall good/welfare of any action, policy or institutional choice we need some way of comparing the benefits and losses to those affected: we need to make interpersonal comparisons of the good/welfare. Yet sceptics have worried either: (1) that such comparisons are impossible as they involve an impossible introspection across individuals, getting 'into their minds'; (2) that they are indeterminate as individual level information is compatible with a range of welfare numbers; or (3) that they are metaphysically mysterious as they assume either the existence of a social mind or of absolute levels of welfare when no such things exist. This paper argues that such scepticism can potentially be addressed if we view the problem of interpersonal comparisons as fundamentally an epistemic problem – that is as a problem of forming justified beliefs about the overall good based on evidence of the individual good.

INTRODUCTION

A longstanding and prominent way of evaluating social, political and economic policies and institutions is by how they net affect people. If one policy benefits those affected more than another then, on an overall-good based moral theory, this potentially makes that policy

[1] I am very grateful for extremely helpful comments on this paper from participants of a 2011 NYU Political Theory group meeting, a 2011 LSE Choice Group session, a 2012 Harvard-MIT special political theory seminar, and a 2015 Paris seminar in Normative Political Philosophy at the *Ecole des Hautes Etudes en Sciences Sociales,* and also to detailed comments from Russell Hardin, Sean Ingham, Michael Kates, Dimitri Landa, Christian List, Bernard Manin, Michael Rosen, Kai Spiekermann and Lucas Stanczyk.

morally preferable. And as every actual conceivable collective policy or institutional change will have some who benefit and some who lose we need a way of comparing these loses and benefits: we need some way of making interpersonal comparisons of the good or of welfare. (As an aside on terminology: within moral philosophy the term 'the good' is frequently used, within economics 'welfare' is near ubiquitous, and historically 'utility' was dominant. 'The good' is perhaps a better term for any general discussion but – given the widespread contemporary use of 'welfare' – in particular examples this paper often reverts to that when it coheres with the existing literature, and nothing is meant to hang on which term is employed).

Interpersonal comparisons have, however, faced three distinct sceptical challenges. Firstly, there is 'the possibility-critique': that such comparisons cannot be made, specifically as doing so would require introspectively experiencing the internal mental phenomena of multiple individuals. It would, in a sense, require us to get into the heads of those affected by policies or institutions, something we cannot do. Robbins, for instance, justifies his scepticism by observing that 'Introspection does not enable A to measure what is going on in B's mind, nor B to measure what is going on in A's. There is no way of comparing the satisfactions of different people.'[2]

The possibility critique holds that while introspective comparisons of alternatives by an individual are possible, the mere existence of multiple minds renders such comparisons impossible across multiple individuals. In other words, if I ask you as an individual to compare two different scenarios in terms of your welfare then you might be able to do so as you have a unified perspective to do it: your own mind. Under the possibility critique, the fact that there isn't such a unified mind when thinking about different individuals' welfare in different scenarios entails that interpersonal comparisons are impossible.

The possibility critique arises because of one key feature: you can't see the welfare levels of others directly. If, for example, you were to add up the heights of a group of people you could ask them all to stand by a large tape measure. Here you can accurately measure that someone is 1m80cm and so on. Welfare, so the possibility critique goes, just isn't like this, and thus you can't make interpersonal comparisons (i.e. add up the values).

[2] Lionel Robbins, *An Essay on the Nature and Significance of Economic Science* (London, 1962), p. 140, and see paper VI for the general discussion.

Secondly, there is 'the determinacy-critique'. This typically assumes that only ordinal individual information is possible. Ordinal information is where all we know is rankings, that for instance we can only know for an individual that one arrangement is better than another, but not by how much. As such there are infinite possible numerical values that could be attached to these rankings that would be consistent with them when aggregating the good of multiple individuals, and therefore our overall ranking will be indeterminate except for when everyone has two options ranked non-conflictingly (in that there are not two people one of whom has A ranked higher than B and one of whom B higher than A).

According to Jevons, the key problem is thus that the 'susceptibility of one mind may, for what we know, be a thousand times greater than that of another. But, provided that the susceptibility was different in a like ratio in all directions, we should never be able to discover the profoundest difference. Every mind is thus inscrutable to every other mind, and no common denominator of feeling is possible'.[3] One notable consequence of this is that we cannot know that some individual is not what Nozick has dubbed a 'utility monster'[4] – satisfying their preferences actually contributes much more to overall welfare than that of others even though this fact is not observable. As a result, so the argument holds, we cannot have determinate knowledge of overall welfare.

Thirdly, and perhaps most fundamentally, there is 'the metaphysical-critique': that interpersonal comparisons are meaningless as they assume the existence of an entity – 'the overall good' or 'social welfare' – whose ontological or empirical status is mysterious. An individual's welfare is potentially based on facts about the mind / brain of the individual in question (her likes, dislikes, values, emotions etc.). But overall welfare would require, so the critique suggests, either (i) some social mind or (ii) the existence of *absolute* amounts of welfare for individuals in each situation even though this cannot be measured or empirically verified. The idea of the second is that while there is some truth to people being able to compare scenarios – such that one scenario is much better than another – there isn't any truth to a scenario representing or providing a particular 'amount' of welfare.

[3] Stanley Jevons, *Theory of Political Economy* (London, 1871), p. 21.
[4] Robert Nozick, *Anarchy, State and Utopia* (New York, 1974), p. 41.

These concerns seem ultimately to lie behind much scepticism. Arrow states that 'The viewpoint will be taken here that interpersonal comparisons of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility.'[5] This is justified for him both by the determinacy concern and fundamentally that 'it seems to make no sense to add the utility of one individual, a psychic magnitude in his mind, with the utility of another individual'.[6]

All three types of scepticism are important and interesting in their own right – not as mere instances of general moral scepticism – as they typically hold that while claims about the good of individuals are both possible and meaningful – via perhaps their preferences, what they value, their happiness, their well-being etc. – there is simply no way to correctly, determinately and meaningfully aggregate such claims, and thus to compare such claims. Sceptics need not deny that someone being robbed, made unemployed, being tortured, losing income, suffering disability or living in penury may be bad for them. What they deny is that when what is bad for some individuals clashes with what is good for others – when some people benefit and some lose from some change – either that we can be in a position of knowledge to compare these changes directly across individuals (the possibility critique), that we can use this information to gain determinate knowledge of overall welfare (the determinacy critique) or that the overall good or welfare meaningfully exists (the metaphysical critique).

The first part of this paper discusses how we might reject both the possibility and determinacy critiques and then, subsequently, why as a result the metaphysical critique may fail too. The means of doing so will be by treating the task of making interpersonal comparisons as an epistemic problem. To illustrate what is meant by this, consider the following.

Twenty students are asked to write down their birth city (or nearest city to birth location), the species of their first pet, and their favourite author's surname but are not to show anyone else. For example, perhaps five of them are (though nobody gets initially to see all of these):

---

[5] Kenneth Arrow, *Social Choice and Individual Values* (New York, 1963), p. 9

[6] Arrow, *Social Choice,* p. 11.

	City	Author	Pet
Mat	Cambridge	LeCarre	Dog
Barry	York	Orwell	Snake
Shahed	Pittsburgh	Brown	Cat
Kevin	Canterbury	Rowling	Rabbit
Sarah	Ramsgate	Rushdie	Alligator

We now ask the students to each individually indicate which has the most, medium and least letters, their city, pet or author. So with the previous example that would be as follows:

Mat:	C > A > P
Barry:	A > P > C
Shahed:	C > A > P
Kevin:	C > A > P
Sarah:	P > C > A

A key question is: once we know their individual rankings, can we use this to form justified – not infallible, but justified – beliefs about which of the combined words (the twenty cities, or the twenty pets, or the twenty authors) has more letters overall?

If the answer is yes, and we can set out the epistemic principles that permit this, then we can use those very same principles to form justified beliefs about the overall good of different alternatives given either partial or full ordinal individual information. We can, in other words, make justified interpersonal comparisons of the good given any conceivable set of individual information. (Ordinal information can be inferred from cardinal information, thus if we can show the possibility and meaningfulness of interpersonal comparisons given ordinal information we have shown the possibility and meaningfulness of them given cardinal information too.) If true this undercuts the possibility and determinacy critiques. The first part of this paper discusses how and why.

In doing so, it reframes the normal debate by making four key claims. Firstly, that the task of interpersonal comparisons is fundamentally epistemic: it is a problem of how to use evidence about

the good of individuals to form justified beliefs about the overall good. As such it seeks justified beliefs, not infallible ones. This is all we need.

Secondly, that to address scepticism about interpersonal comparisons of the good we do not need to specify what the correct account of the individual good is: whether it is well-being, happiness, preferences, valued-ends or whatever. All we need to determine is, given such an account, how to use evidence about the good of individuals to form justified beliefs about the overall good. By doing so our problem is greatly simplified and the key issues more easily identified.

Thirdly, that merely by virtue of seeking justified beliefs we must reject Arrow's Impossibility Theorem as correctly characterizing our task, for one of its key conditions – the Independence of Irrelevant Alternatives – logically entails the possibility of incoherent beliefs (a result proved in the Annex). No credible credence function or account of epistemic rationality would accept such a premise, and we should not: there is nothing impossible (or only possible if we accept a decisive dictator) about social welfare *viewed as an epistemic problem*. This is discussed and set out in the Annex.

Fourthly, if we do think of the overall good as epistemic, then we may remove a range of metaphysical concerns over the ontological status of this good. We may be able to think of the overall good as simply a product of combining all knowable information about the good of individuals in a way that is justified, that is in a way that is unbiased and coherent. As such we do not require the existence of a social mind nor that *absolute* individual welfare values be empirically 'real'. An epistemic solution to the possibility and determinacy critiques would potentially also therefore have the resources to challenge the metaphysical critique.

The paper proceeds as follows. Part I illustrates the logic of the overall approach using simple string and word-length analogies. Part II formalizes the results. Part III discusses three fallacies to avoid. Part IV considers the metaphysical implications of the overall argument. Part V revisits the case for scepticism.

I. ILLUSTRATING THE UNDERLYING LOGIC OF THE EPISTEMIC APPROACH

Ten students go into a room one by one and cut off a piece of red string and a piece of blue string from the respective balls of string. We want to subsequently form justified – not infallible, but justified – beliefs about which is longer overall, the red strings combined or the blue strings combined. We initially know nothing about the students and nothing about their strings. They each now individually compare which of their strings personally is longer, the red or blue. Seven have red strings longer than their blue ones; three have blue strings longer than their red ones. Which combined strings, red or blue, should we believe is overall longer? That is, how can we form a justified belief about overall string length from ordinal evidence – i.e. about which is greater, less or the same – about individual string lengths? One way we can do so is by adopting Unbiasedness and Coherence:

Unbiasedness: Beliefs that are justified should be unbiased in the inclusion and treatment of information from individuals unless there is evidence supporting, or a positive reason for, such bias.

Coherence: Justified beliefs about the overall amount of X positively supervene on beliefs about the particular amounts of X.

Given Unbiasedness and Coherence, if all – and this is a crucial clause – if all the evidence we have is that seven students have red strings longer than their blue, and three have blue strings longer than their red, then we would be justified in believing that the red strings ('R') collectively are longer than the blue ones ('B'), or giving a greater credence to the theory 'R>B' than the theory 'B>R'. (Hereafter a capital letter – such as R – refers to the overall amount of the quantity in the arrangement R, and r(i) that of individual i in the arrangement R.) Unbiasedness ensures that we give each student's information equal weight, and this combined with Coherence ensures that learning that there are more red strings longer than the blue justifies, *absent any other evidence*, believing the red is overall longer.

Now, it is true that in adopting Unbiasedness, and implicitly not assuming that anyone's overall combined string is greater in length than anyone else's, we have not guaranteed *infallible beliefs*: it might be that some student, Robert, took lots more string than all others combined, a veritable string monster. But with absolutely no evidence for this we would violate Unbiasedness in giving Robert's evidence

more weight than that of others (it might incidentally be that he took much less). In the string case our intuitions may be nebulous and not always support this. But in the welfare/good case Unbiasedness is a much more compelling general condition as it may be sufficiently justified in either of two ways: methodologically or morally. Which ultimately is adopted will potentially impact the type of result: if Unbiasedness stems from a methodological commitment then the result will not be a direct argument for overall-good based moral theories (merely permissive of them), but if it stems from a moral understanding of Unbiasedness then this seems to lend some credence to the idea that treating people equally involves giving equal weight to their interests.

The first broad justification of Unbiasedness is to see it as a basic methodological principle, roughly as a generalisation of something like Harsanyi's 'principle of unwarranted differentiation' whereby 'If two objects of human beings show similar behaviour in all their relevant aspects open to observation, the assumption of some unobservable hidden difference between them must be regarded as completely gratuitous hypothesis and one contrary to sound scientific method'.[7] This is a broadly evidentialist view of empirical method – that 'the epistemic justification of a belief is determined by the quality of the believer's evidence for the belief'[8] – though would also be straightforwardly entailed by Bayesianism with uniform priors.[9]

Secondly, however, Unbiasedness may be sufficiently justified not only by a methodological principle but also by a ethical one, by the principle of equal moral concern: if we show bias towards the good of some individuals without any good-based reason for so doing then this discrimination against others cannot be justified. All humans are worthy of equal moral concern, under this principle, and this means that the comparative weight we give them in evaluating what happens to them should be equal. A very, very broad range of approaches to morality seem capable of motivating this, such as that

---

[7] John Harsanyi, 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility', *The Journal of Political Economy* 63 (1955), pp. 309-321, at 317.

[8] Earl Connee and Richard Feldman, *Evidentialism* (Oxford, 2004), p. 83. Evidentialism would, prima facie, treat informationally identical evidence as supporting the relevant hypotheses identically.

[9] The type of multi-parameter cases where uniform Bayesian priors are problematic, such as that of forming priors over a factory's cube sizes, do not apply in the case of beliefs about the good/welfare since this represents only one parameter. For the famous cube example see Bas C van Fraassen, *Laws and symmetry* (Oxford, 1989), p. 303.

biased treatment could be reasonably rejected[10], that it is not compatible with treating people as free and equal[11], or that it violates the principle that each counts for one and only one.[12] The sceptic, in arguing for biasedness, therefore minimally has to show the falsity of both the methodological and moral justifications of Unbiasedness.

Coherence is that beliefs about the overall amount of something should supervene upon beliefs about the individual amounts of that thing, that for instance your beliefs about whether there are more boys than girls in a particular class should supervene upon your beliefs about how many boys and girls there are in the class: if you think there are ten boys, six girls, and more girls than boys, then there is something wrong with your beliefs. In the string case Coherence entails that beliefs about the overall lengths of particular string colours should cohere with beliefs about the lengths of the particular bits of strings, such that if we find out that Jack has a red string longer than his blue we should be more confident the red is overall longer than if we had learned that Jack had identical length strings, or his blue longer than his red. (Absent other information. Obviously, if we have actual evidence that Bob likes to copy Jack, or that most of the class like to shun Jack's choices, then this needs including too.)

In the good or welfare case it means that beliefs about the overall amount of good or welfare are based upon beliefs about the amount of good or welfare of all the individuals. Coherence may, on some accounts, simply be a consequence of defining the overall amount of good / length of the red (or whatever) string as the sum of the particular amounts of good / bits of red (or whatever) string, and asserting that various constraints on justified beliefs necessary follow from this. It is included as a separate principle for clarity. It is, however, Unbiasedness that really does the major work in the results that follow, and that as such raises the most interesting meta-ethical issues, as discussed in section four. Unbiasedness and Coherence

[10] Timothy Scanlon, *What We Owe to Each Other* (Cambridge MA, 1998).
[11] John Rawls, *A Theory of Justice* (Cambridge MA, 1999). Note that unbiased beliefs about the overall good need not for a Rawlsian entail that maximin be rejected; the point of maximin is a rejection of maximizing the overall good, not of the possibility of the overall good.
[12] This idea is of course famously in Jeremy Bentham, *An Introduction to The Principles of Morals and Legislation* (Oxford, 1907), well-discussed in Henry Sidgwick, *The Methods of Ethics,* 7th edn. (London, 1907) and expanded to non–human animals and further defended as a basic commitment of morality in Peter Singer, *Practical Ethics* (Cambridge, 1993).

permit us to form justified beliefs about the overall lengths of the strings based on any information about the students' individual string lengths and, using the same logic, do so for the overall good.

One way the earlier example may mislead however is that when we are dealing with small numbers of individuals, or when we are dealing with large numbers but where the ratios of different options are fairly similar, we might as such just be not that confident in any conclusions. The thought would be: can we really be very sure?

This isn't a rejection of the overall epistemic approach however: it's an endorsement of it. It is only by having an account of how individual level information impacts overall beliefs that we can have a second order account of how confident we can be in the conclusions. Compare the following two scenarios where all you know is how each individual's author words relate to their birth-city words (and you know strictly nothing else, and only one scenario obtains – so it isn't indirect evidence for the other):

Scenario One: 30 students have authors longer than cities, 20 students the opposite.

Scenario Two: 45 students have authors longer than cities, 5 students the opposite.

In both cases, if this is all the information you have, under Unbiasedness and Coherence you will be justified in believing the author-words-combined have more letters than the city-words-combined. However, intuitively, in the second case you can also be more confident of this. Similarly, imagine you have just three students and they are asked to write down their favourite author, birth city, species of first pet, favourite singer's first name, last-sport-played, mother's maiden name and last holiday destination. They are as follows:

Student 1: Author > Pet > Singer > Sport > Mother > Holiday > City

Student 2: Author > Sport > Mother > Holiday > Singer > Pet > City

Student 3: Sport > Mother > Holiday > Singer > City > Author > Pet

Here it seems you can be more confident that the author words combined are longer than the city words than if you just knew the following:

Student 1: Author > City

Student 2: Author > City

Student 3: City > Author

Many will have intuitions that the more the categories are apart – both in terms of numbers of individuals and in terms of other categories – the more confident we can be that one is overall longer than the other. Unbiasedness provides a justification for these intuitions: each individual's information (and category) is in principle treated with equal epistemic weight (unless we have evidence to the contrary), thus the more individuals who have the ranking author>city, the more confident you can be that the author words combined are longer than the city words combined, and similarly the more words between someone's author and city, the more confident you can be the difference between these words is larger. Unbiasedness and Coherence endorse and justify all of these judgements. (One further thought here is that there are only so many letters, so we can guestimate some discreet differences. This is a side issue: if distracted by it then re-run the example with string lengths.)

More generally, Unbiasedness and Coherence are useful because often we have some evidence of underlying phenomena but not full evidence of them. Even if, for example, we cannot know the actual numbers of letters in all the words, we can still use evidence of the rankings of the students to form justified beliefs about which contains more words overall: the cities, authors or pets. Presumably there will be some underlying probabilities depending on where one does this and if we could obtain evidence of these then that could bear on our credences too. But if all the information we have is the individual rankings – in this case it might not be, but in the welfare/good case for the sceptic it is – then we can still form justified beliefs overall.

The string/word to welfare/good analogy is thus inexact. It is stipulated in these examples that direct individual level information is all the evidence we have, as that is true in the good/welfare cases. However, we probably have multiple rather diffuse intuitions about

pets, authors and cities. These examples, of strings and words, are thus meant to be illustrative of the underlying logic, even if some additional implicit background evidence – about dog versus kangaroo prevalence perhaps – requires ignoring. For someone who struggles to avoid bringing in such 'background evidence' then a better analogy might be if the example instead involved students in an unknown foreign language and place. All you know is the rankings (you don't even know the categories, nor how many letters in the alphabet or anything). This, after all, is the premise the sceptic asserts: you don't have evidence of the underlying values at all.

The overall epistemic approach simply says that we should use evidence about individual amounts (individual words/strings/good/welfare) to form coherent beliefs about overall amounts (overall words/strings/good/welfare), and do so in a manner that treats all unbiasedly (with equal moral concern). What you would do in the word and string cases you should simply do in the welfare case because you know the same sorts of things and are in the same epistemic situation (if you are an interpersonal comparison sceptic that is: if rough interpersonal comparisons are possible by empathy or such like then we don't need this, but then scepticism fails too).

II. FORMALIZING THE APPROACH

The substantive normative discussion continues in part III, this section simply sets out one (of the many) ways of formalizing the overall logic. It does not 'prove' the epistemic approach justified, the substantive arguments in favour of it are considerations to that effect. It merely hopefully makes it somewhat easier to understand and to criticize for some readers at least.

To formalize the above discussion we can adopt a credence framework or use the concept of expected value. (The two yield functionally identical results, but one may be easier to use than the other in particular contexts. Cognitively, an expected value approach is easier to follow, thus that is used here.) To do so, we need to set just three values. Firstly, we need to decide whether the identity of individuals should affect their relative weight in the calculation of expected values. Secondly, we need to decide in principle what relative expected value to give to B and R, before we have any evidence whatsoever. Thirdly, if we learn of an individual that for them R has more of the quantity than B – that $r(i)>b(i)$ – then we need to decide how this should affect our overall expected values compared

to had we learned different information. Unbiasedness and Coherence allow us to do all three.

Here are three basic axioms in an expected value framework, where E[R] represents the expected value of any arrangement R before the evidence, E[R / e] represent the expected value of R after learning e, and e*i = r(i)>b(i), r(i)=b(i), or r(i)<b(i) for individual i and all r, b such that r≠b.

1. Equal treatment 'ET': For all individuals i,j and arrangements R let E[R / e*(i)] = E[R / e*(j)].

2. Pre-evidential indifference 'PI': Before any evidence E[R] = E[B] for all R, B.

3. Equal positive supervenience 'EPS': For all R and all individuals i, j, where y>0 then E[R / r(i)>b(i)] = E[R] + y; E[R / r(i) = b(i)] = E[R]; and E[R / r(i) < b(i)] = E[R] - y

Equal treatment and pre-evidential indifference stem from Unbiasedness: to be biased towards one arrangement represents an implicit bias towards those favoured by that arrangement, to give the information of some individuals more weight than that of others is to be biased toward them. Equal positive supervenience is based the fact that the overall amount of something is comprised of the amount in its constituent parts. Rejecting equal positive supervenience would mean, for example, that the more students you learned had red strings longer than their blue ones, the more confident you would become that the blue strings were overall longer (violating Coherence) or that your beliefs about string lengths would be affected by the identities of individuals rather than information about their strings (violating Unbiasedness).

Given Equal Treatment, Pre-Evidential Indifference, and Equal Positive Supervenience it follows that, for body of evidence E, if n(r+) is the number of individual level pieces of evidence that r>q, if n(r=) is the number of individual level pieces of evidence that r=q, and if n(r-) is the number of individual level pieces of evidence that r<q, all where Q≠R, then for all R, Q, B:

$$E[R / E] > E[B / E] \text{ iff } n(r+) - n(r-) > n(b+) - n(b-).^{13}$$

To put this verbally:

> *The ordinal epistemic welfare/good principle*: If we have no evidence that an unbiased super-set of options would differ, then of two options we are justified in believing that A represents more welfare/good than B if and only if we are justified in believing the number of arrangements in which individuals have less good than A minus the number in which they have more is greater than the number of arrangements in which individuals have less good than B minus the number in which they have more.

It should be noted, however, that in the string or word cases this does not ensure we have infallible beliefs, merely that they are justified. It is possible, for instance, to think of possible distributions of string lengths that would permit us to form false beliefs, but we have no reason to think these distributions hold. This is the epistemic approach: we use evidence to form justified beliefs. Part IV discusses how in the welfare case true beliefs might in fact simply be justified beliefs given all possible evidence.

What is particularly striking about this overall result is how minimal are the conditions we require to produce it. If we seek beliefs that are unbiased and coherent then merely with ordinal information we can form justified beliefs about the overall good/welfare, and thus make justified interpersonal comparisons.

Since the aim of this paper is to show how scepticism about interpersonal comparisons of the good/welfare may be overcome by adopting the epistemic approach, the focus here is thus on explicating and defending comparisons given only ordinal information, as this is the case most favourable to the sceptic (from any cardinal set of information one can infer ordinal information, thus if comparisons based on ordinal information are possible, determinate and meaningful then so are ones given cardinal information). Discussion of comparisons using cardinal evidence is therefore put to one side,

[13] A proof: $E[R / E] = E[R] + yn(r+) - yn(r-)$; $E[B / E] = E[R] + yn(b+) - yn(b-)$; as $y>0$ and, from PI, $E[R] = E[B]$, therefore $E[R / E] > E[B / E]$ iff $n(r+) - n(r-) > n(b+) - n(b-)$. QED.

though it is implicitly covered for those that think that cardinal information is solely obtained via ordinal evidence (such as the lotteries over outcomes individuals would rationally choose[14]). The ordinal is however sufficient to illustrate the basic epistemic logic, and the ways in which the epistemic approach may be mis-interpreted. It is to these this paper now turns.

III. THREE FALLACIES TO AVOID

You are a peasant. Consider the following options:

1. The king takes all your property.
2. The king takes three-quarters of your property, but has to work for a week digging turnips.
3. The king takes half your property.

We could collect information on how these compare and then might find of them that the king has more welfare in 1 than 3, and more in 3 than 2 (he is very wealthy already, and really hates digging turnips). You have more welfare in 3 than 2, and more in 2 than 1. If this is all we know then it might seem we would be justified in believing that 1 represents more overall welfare than 2. But this seems absurd. Which it is: the belief that 1 represents more welfare than 2 is unjustified because it violates Unbiasedness, in that Unbiasedness covers both how we treat individual-level information, and how we select and structure it. More specifically, even if the above options were the three options we had to choose between, if we use them alone to structure the evidence in so doing we would be committing the first of three fallacies, namely:

The *'unjustified domain fallacy':* is that beliefs about the overall good (or welfare) are justified even if the choice-domain specification that underpins them is itself unjustified.

The issue here relates to how we select the different options about which we seek and incorporate evidence. If we (wrongly) think of

[14] See notably John Von Neumann & Oskar Morgenstern, *The Theory of Games and Economic Behavior* (Princeton, 1944).

welfare by analogy with voting, then the domain appears naturally to be the candidates or choices, and overall beliefs justified by the individual information about these. In contrast, however, for the epistemic approach the options are selected and distinguished based on our theory of the good and the evidence. The key point is this: if we have to choose between a set of options it is the alternatives about which we can gain distinct evidence that represent the potential domain which we use to form justified overall beliefs, not merely the options at hand. Even if we are only choosing between red and green strings, the blue strings can be informative as to that choice, and all such string information could be potentially included. Hence why for justified beliefs we need to seek out and include a set of evidence that can be justified by Unbiasedness and our theory of the good, not necessarily simply by the options under consideration.

The unjustified domain fallacy is very closely related to a second way we can go wrong in using the epistemic approach, namely:

The *'ignored evidence fallacy':* is that a method of forming beliefs about the overall good (or welfare) is unjustified due to the implausibility of the beliefs it produces where the justification of these beliefs is based on the exclusion of the very evidence that underpins the judgment of their implausibility.

For example, consider someone arguing that slave holders prefer to keep a domestic slave, such a slave prefers to be free, and thus if this is all the information we have then we are justified in believing that either arrangement is of equal welfare, thus the epistemic approach must be wrong. This is fallacious as the correct judgment about the conclusion's implausibility stems from the vast, rich and detailed evidence and understanding we have about what it is like to be a slave, what slaveholding entailed, what the effects of it were as an institution, and how much of what is most valuable in human life was denied the slave. As social beings in a rich cultural environment we have an incredibly detailed and complex understanding of a wide range of patterns of human valuation, of suffering, hope, pain, dignity, freedom from domination, the enjoyment of security, the determinants of social respect, the ability to develop one's potential and shape one's life and so on. If we ignore vast swathes of evidence we will reach conclusions that, with this evidence in mind, seem unjustified. But it is

the ignoring of the evidence that is at fault, not the attempt to use evidence to form justified beliefs.

It is also for a similar reason that this paper explicitly separates out two questions: (1) What comprises evidence of the good of individuals? (2) How, in principle, given an account of the good of individuals, should we combine evidence thereof into beliefs about the overall good? If we fail to treat these as distinct then disagreement about 1 will risk fuelling disagreement about 2 as each theorist will, from the other theorist's perspective, have ignored evidence (the relevant evidence that arises from the different accounts). Finally there is:

The *'absolute comparison fallacy':* that scepticism about ordinal or cardinal epistemic interpersonal comparisons of the good/welfare can be justified by reference to claims about absolute levels of the individuals' good/welfare.

That is, it might be that in a range of situations we think that people's welfare is of a similar absolute level, that for instance we think the welfare level of everyone sleeping (without dreaming) or being dead is identical. However, if we use the epistemic approach based on only ordinal or cardinal information then these values might come to differ: someone who would knowingly and consistently take a higher risk of death for a large range of goods assuming they live could be assigned a lower level of welfare to death than someone who was much more relatively death averse. (It is assumed in this case that people's welfare is partially based on what they value, and that what they would choose is evidence of what they value.) But if we think being dead must represent the same welfare for all people then the epistemic approach seems to have given us the wrong answer. Such an argument does not support scepticism about the good/welfare however, as it presupposes the very falsity of such scepticism.

For example, consider the previous example where everyone writes down their birth city, favourite author and species of first pet. We learn how the number of letters in each of these compare for each individual, and form beliefs about the overall number of letters in the city, author and pet words combined. Now we learn that everyone was born in the same city and yet our previous comparisons had implicitly assigned a lower value to the cities of individuals who had more letters in their pet and author words than those who had the most

letters in their city words. This is mis-premised as a criticism as what we had previously were justified beliefs *given that we could not make absolute comparisons, i.e. directly compare people's city words*. The epistemic approach is needed when we only have ordinal or cardinal information: we know how options compare, but not their absolute values. If we assert such absolute values for some options however then our previous comparisons will not reflect that information. But this is not a problem, it is great: we can make direct interpersonal comparisons and scepticism fails.

As such we can see how the overall approach fits with other arguments in the literature about how we might come to know welfare information, such as via extended sympathy. These theories – if they succeed – are analogous to where we have evidence of absolute amounts, something with which the epistemic approach is perfectly compatible. Its real value however is if these arguments fail, for the sceptic still needs to provide an account of why we are not justified in using the epistemic approach.

The overall point of emphasizing all three fallacies is to try to pre-empt the conflation of localized disagreement over a particular conclusion based on the epistemic approach with generalized counter-arguments applying to the approach. If we choose biased domains, ignore evidence, or assert absolute comparisons then, having done so, we will risk conflict with epistemically justified comparisons implicitly based on unbiased domains, included evidence or absent absolute comparisons. This, however, is an argument for evaluative consistency and the updating of one's beliefs, not scepticism.

IV. METAPHYSICAL IMPLICATIONS

The role of the analogies with both string and word lengths was to highlight how we can use well established epistemic principles in order to combine individual level information about various quantities into beliefs about the overall amount of various quantities. There is one way this may be potentially misleading however: both the length of string and the overall number of letters in words are directly empirically verifiable properties in that they have values which we justifiably believe exist. There are eight letters in Melville, not nine or two. If we think that welfare is like that, in that there can be some empirical truth to Bob somehow having 57 (rather than any other number) units of welfare in situation A then we can still form justified beliefs and make, by implication, interpersonal comparisons. As such

the determinacy and possibility critiques fail. But welfare need not be like this, and as a result there potentially be nothing *necessarily* metaphysically mysterious about overall welfare or the overall good.

Consider the hypothetical case where we know all there is possible to know about the *relative* good or welfare of every possibility for each of the individuals affected: we have complete cardinal information about how much better or worse each possible scenario is for each individual. Now, if we assume that there is some 'overall good/welfare' that actually exists as the sum of individual-specific values that themselves empirically exist then it is possible that in using the individual level information to gain justified beliefs about overall good/welfare we would still have false beliefs, notably that some individuals may just be more important for the overall good in ways that cannot be known based on any information about them. In this case in combining their individual level information in a way that was unbiased we could form justified but false beliefs.

This we might label 'physical realism' about the good. This sort of position appears to be what the 'metaphysical critique' implicitly has in mind as both problematic and necessary. It seems to be a strong version of a cognitivist, truth-apt and non-error theory, one that conceives of the good very much in the same way as it conceives of empirical entities. Hopefully the label 'physical realism' manages to capture this.

The physical realist position is that the overall good exists in something like the same way that the number of letters in the previous sentence exists as the sum of its nine constituent words. If we make certain assumptions about how this good relates to the good of individuals we can still use evidence about the good of individuals to form justified beliefs about the overall good (and this is what we did, for instance, with string). Such beliefs are however justified but not necessarily true. For the physical realist the same applies to the overall good: there are true facts about the absolute level of an individual's welfare (for instance) – such as that in situation X Meghna has a welfare of 21 – even though we cannot ever conceivably gain direct knowledge of this number, nor indeed can Meghna. But it exists. Overall welfare is the sum of these numbers.

There is an alternative however: that the overall good is a construct that is created by combining information about the good of individuals in a justified manner. That is, once we have an account of justification then the true overall good is that which we would be

justified in believing were we to know all possible information about the relative good of individuals. There is nothing more: the status of Jack having a welfare of '57' in A is simply an epistemic construct, it is not that an omnicognisant being could look into Jack's head and count 57 welfare units or measure his welfare to be 57. The welfare/good as a construct position would perhaps be:

> 1. There are true empirical facts about how different arrangements relate to individuals.
>
> 2. A justified theory of the good/welfare would supervene on these facts to yield facts about how, for each individual, different arrangements relatively compare in terms of their good/welfare.
>
> 3. Combining all possible such facts in a justified manner would yield true facts about the overall good/welfare.
>
> 4. We can use individual level evidence to choose between theories about the overall good/welfare and, in principle, were we to believe all such potential evidence and reason correctly, we would have true beliefs about the overall good/welfare.

The point of introducing the good-as-construct position here is not to demonstrate that it is the right meta-ethical stance, and indeed the outline above is at best a suggestive sketch, not a proper defence. The point instead is to stress that there is nothing *necessarily* metaphysically mysterious about the overall good, it being a justified construct being one (potentially among many) meta-ethical positions to be able to make sense of the overall good *without requiring the existence of a social mind or of empirically real absolute individual welfare levels.*

A different way of capturing the same conclusion is to consider what the underlying individual-level mental facts might be like. Assume – for argument's sake – that people's welfare is based somehow on their desires or what they value (the logic can be re-run using alternative conceptions). There seem to be two possibilities as to the underlying mental facts. Firstly, desires could have some objective mental strength or quantity, such that my desire for A will ultimately be derived from some quantity-like feature of part of my mental infrastructure (some feature of how strong the neurons link or how many link etc.). If so the string and word length analogies are directly

applicable, but as such there is the possibility of having justified but false beliefs (about overall welfare, overall string lengths, overall word lengths).

Alternatively, imagine that the only individual-level mental valuation facts are comparative, thus it is not that there are mental facts about the desire for A but rather only about the desire for A *compared to B*. Here, although the string and word analogies are less tight, if anything the epistemic approach can be more confidently used, for the more evidence we get the more confident we can be that epistemically justified beliefs are converging on true beliefs (if we treat people unbiasedly).

One notable implication is that if the mental evaluative facts are ultimately comparative then there cannot be a 'utility monster'. If they are absolute then there could be such a monster, but without evidence about who this is we would be unjustified in attributing this to a particular set of individuals, thus this does not entail the impossibility of justified interpersonal comparisons.

Which set of mental evaluative facts apply to humans (and non-human animals) is open to dispute. The point here is that whichever does we don't have to assert the existence of a social mind, and even if there are no absolute welfare-level facts for individuals in a situation, only comparative ones between situations, then the overall welfare/good is a meaningful concept about which we can gain evidence and form justified beliefs.

V. REVISITING SCEPTICISM ABOUT THE GOOD

For the epistemic approach, the possibility critique is addressed simply by virtue of the possibility of having evidence about the good of individuals: if we don't need to get 'into' the minds of individuals to gain evidence of their good or welfare, then when using epistemic principles to combine this evidence we don't need to do so either. The possibility of individual-level evidence about the good entails the possibility of justified beliefs about the overall good. Just as we don't need to see all the students' words to start to form justified beliefs about the overall length of the word-combinations, so we don't need to 'see' people's welfare in their heads to start to form justified beliefs about overall welfare as we gain evidence about their welfare.

The determinacy critique, that we cannot have determinate knowledge of overall welfare, is addressed in two ways. Firstly, the

epistemic approach denies that determinate knowledge is the necessary goal: what we seek are justified beliefs, for these allow us to evaluate actions, policies and institutions, the reason for seeking interpersonal comparisons in the first place. For the determinacy critique rests on the following mistaken reasoning:

Premise: The evidence does not deductively entail a unique hypothesis about overall welfare.

Conclusion: We cannot form determinate justified beliefs about overall welfare.

Now, if we assume a physical realist account of the good, then the premise will be almost always true. But even if so the conclusion is unwarranted. Indeed, if the general reasoning were correct then it would potentially entail the impossibility of much empirical method in general, for it is rare for evidence to prove a general theory and on some accounts it is normally impossible.[15]

The second way the determinacy critique may be addressed is that, if the welfare/good-as-a-construct approach is correct, then the overall good simply 'is' that represented by justified beliefs given all possible evidence and the determinacy critique cannot even arise as the gap between justified beliefs and true beliefs is, in principle at least, bridge-able. If the overall-good-as-construct approach is correct, as we gain more evidence justified beliefs will be expected to actually converge on true beliefs and, were we to gain complete evidence, represent true beliefs with no possible indeterminacy.

The ultimate problem for the possibility and determinacy critiques is that the conditions that supposedly justify them – our lack of access to the actual values of the various individual quantities, or inability to experience values across individuals – are exactly the conditions that hold in the string and word examples (and many, many, many more). A sceptic is potentially forced to either accept that determinate justified beliefs about string, word-lengths, welfare and the good are all possible; or that they are all not. That is, the difficulty for the sceptic comes because they must specify which of the following holds:

[15] This, for instance, is an enabling premise of falsificationism. Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (London, 2002).

1. The sort of epistemic information we have in the string/word case on the one hand, and in the welfare case on the other hand, is not of the same type.
2. The sort of epistemic information we have in the string/word/welfare cases are of the same type and we cannot form justified beliefs about the overall amounts of string/words in these sorts of situations.
3. The sort of epistemic information we have in the string/word/welfare cases are of the same type; we can form justified beliefs about the overall amounts of string/words in these sorts of situations but we cannot form justified beliefs about the overall amounts of welfare in these sorts of situations.
4. We can form justified beliefs about the overall amounts of welfare in these types of situations.

Now 1 is false: in both cases, by stipulation, we only have ordinal information, this after all is supposed to be the most sceptic-favourable case (if we have absolute information – such as via extended sympathy – then scepticism can be overcome straightforwardly).

To take 2 is to take a very odd general position in epistemology. Bayesians would certainly deny it. If this is necessary to sustain welfare-comparison scepticism then the position is a very strong one and has a large justificatory hurdle to clear.

Position 3 however seems arbitrary. It is conceded that we know the same sorts of things – that is we know ordinal rankings for the relevant variables and they alone. And it is conceded that in every other case (apart from welfare) we can use such rankings to form justified beliefs about overall amounts. But it is still maintained that we cannot in the welfare case. At a minimum the sceptic here has to address both the apparent arbitrariness and to set out exactly why welfare is relevantly different (it is obviously different in a range of ways, just as strings are different to words, but our epistemic situation in the various cases does not appear relevantly different).

Moreover, individual welfare is presumably based on *some* individual mental facts – see section IV – thus the sceptic needs to say what these are if they are not absolute or comparative in nature (and as

such welfare-comparison scepticism becomes a very contingent empirical hypothesis in neurobiology, and a slightly mysterious one at that). If 1, 2 and 3 fail however, then 4 follows as a matter of logical exhaustion. This is 'The Sceptic's Challenge', namely to do the following:

SC1: Provide a justified account of what principles should govern beliefs in empirical ordinal-information cases, such as with the word/string examples.

SC2: To demonstrate why those principles cannot be applied in welfare ordinal-information cases.

It's a challenge because (1) is a very tractable task. And applying the results of (1) to the welfare case is as such conceptually relatively trivial (just use the same principles). It is the sceptic here who owes us an explicit account: it is not enough to simply note that it is impossible to get into other minds nor to note that in the welfare case we cannot directly observe the values and thus infallibly know a determinate ranking.

Finally, as noted previously, if the possibility and determinacy critiques can be addressed, the metaphysical critique loses much of its force and, if the good-as-construct is correct, it is removed. In a sense, the metaphysical critique rests on a mis-analogy. The individual good is potentially based on the brain / mind of the individual involved. But we need no social mind or group brain to form justified beliefs about the overall good, and we need no such group brain or group mind to account for the potential meaningfulness of this good.

Viewed through an epistemic paradigm, scepticism therefore looks problematic. This potentially has very important implications indeed. For it is virtually impossible to justifiably say anything about overall welfare or the overall good without interpersonal comparisons, a fact that is brutally unavoidable as there quite plausibly has never been a pareto-improving economic or political policy or institutional change undertaken by any state, at any time, at any point, in all of recorded history, and likely never will be. Indeed, I have a long-standing bet to this effect. No suggested pareto-improving policy has yet withstood scrutiny. Moreover, as a colleague Hakon Tretvoll pointed out, even if such an unlikely policy did exist, one final impact would be to cause me to lose the bet, which would leave me worse off,

and entail it couldn't be a pareto improvement. (This delightfully ingenious result is not terribly important, but it does however highlight the sheer improbability of saying anything useful about social welfare without some sort of interpersonal comparisons).

There are always winners and losers in terms of their personal welfare. A welfarism based on Pareto rankings must either idealize human agents or causal chains to the point where they are barely recognizable, or opt for productive efficiency as a duplicitous welfare proxy and by doing so abandon seeking an unbiased normative justification.[16]

Fortunately such a negative conclusion is not necessarily warranted, for the problem of interpersonal comparisons is ultimately a problem of how we justifiably combine evidence of individual amounts into beliefs about overall amounts, something that is both a tractable epistemic problem (assuming a basic account of justification), and one with intriguing implications for how we might understand the ontological status of the overall good.

VI. CONCLUSIONS

The main argument of this paper has been that, if we treat it as an epistemic problem, then beliefs about the overall good, and by implication interpersonal comparisons of the good, may be possible, determinate and meaningful. With this positive conclusion in mind, here, however, are two semi-qualifications.

Firstly, the overall argument assumes the possibility of an account of the individual good. This may – considering only goods, services, leisure and income – be relatively tractable, and indeed we can fairly straightforwardly use the approach to produce estimates of the impact on economic welfare of any policy or institutional change. But if we then include life expectancy, other more rich qualitative determinants of quality of life, the effects on potential beings who may and may not exist, and endogenous changes to the future good of individuals, then a range of potential problems arise.[17] We can form

[16] Hammond rightly, but perhaps over-diplomatically, characterizes using monetary value as an implicit welfare proxy as 'almost certainly unethical'. Peter Hammond, 'Interpersonal Comparisons of Utility: Why and How They Are and Should Be Made', *Interpersonal Comparisons of Well-Being*, eds. Jon Elster and John Roemer (Cambridge, 1991), pp. 200-254 at 201.

[17] See for instance John Broome, *Weighing Lives* (Oxford, 2006) or, for a famous outline of the general temporal difficulties, Derek Parfit, *Reasons and Persons* (Oxford 1984), esp. chapters 8, 16, 17, 18 and 19.

justified beliefs and produce estimates without address these, but a complete account would ultimately need to resolve these issues, something compatible with, but not determined by, the epistemic approach.

Secondly, nothing in the epistemic approach *per se* entails that morally right actions are those that best contribute to the overall good. The approach instead addresses one historically influential critique of theories that require interpersonal comparisons (such as utilitarianism) by showing how and why we can in principle potentially evaluate policies, actions and institutions by their impact on overall welfare or the overall good, and why scepticism to that effect is quite possibly the harder position to justify.

m.p.coakley@lse.ac.uk

Annex: why any general epistemic approach has to reject Arrow's Independence of Irrelevant Alternatives or risk incoherence.

Before setting out the proof, here is an illustration of the logic so it is hopefully easier to follow. Imagine we have a large group of people and three scenarios about which we are going to gain some evidence about their welfare. Initially we know nothing about the people nor the scenarios. Assume we have initial beliefs about the relevant possibilities (about the possibility that scenario A will contain more overall welfare than scenario B and so on). We then learn some things about the individual level welfare and, perhaps, change some of our beliefs. But we also adopt the Independence of Irrelevant Alternatives, that beliefs about whether X overall represents more or less of some quantity than Y should solely be derived from evidence about the relative amount of the quantity in, or comparing, specific instances of X and Y alone.

That is we don't change our beliefs if the evidence is – according to this condition – irrelevant: so if we learn that Bob has more welfare in A rather than B we shouldn't change our beliefs about the overall amount of welfare in B compared to C. Is there any set of initial beliefs we can have so we don't risk ending up with incoherent overall beliefs? The proof below demonstrates that the answer is no.

A quick illustration would be: imagine you start with a greater belief that A>B than your belief that A>C. You now learn that all the individuals have equal welfare in B and C. Under IIA you should still have a greater belief that A>B than that A>C. But you should also

now believe that B and C have the same welfare. You have incoherent beliefs. In fact there is no initial set of beliefs immune to this possibility, not because coherent beliefs are impossible, only because they are impossible to guarantee if you adopt the Independence of Irrelevant Alternatives. Fortunately we can just reject this.

Now the logic behind these sorts of results will be very familiar (from voting cycling amongst other things). The key point is this: adopting the IIA in an epistemic context where you are using ranking information to produce beliefs about overall amounts leaves you vulnerable to incoherent beliefs. The rejection of IIA does not need us to weigh its intuitive plausibility and so on: we simply need to reject a set of principles that result in us both believing P and believing not-P.

If we want to have justified beliefs about string-lengths, word-lengths or welfare, we should thus reject IIA. Hence why, if we seek justified beliefs about welfare or the overall good, Arrow's Impossibility Theorem should not be accepted as validly characterizing our task. In fact, any credible credence framework or account of epistemic justification would violate IIA, so long as it is well-defined (credences in mutually exclusive theories are additive), coherent (we are not more confident in a proposition than in any proposition that is logically entailed by it) and non-dogmatic (possible theories are not believed impossible). Formally:

> (*Axiom 1: Additivity*) If h1 and h2 are mutually exclusive (such that $h1 \rightarrow \neg h2$) then $Cr(h1 \vee h2) = Cr(h1) + Cr(h2)$
>
> (*Axiom 2: Coherence*) If $X \rightarrow Y$ then $Cr(Y) \geq Cr(X)$. This axiom of Coherence is not strictly equivalent to the previous principle of the main article, and in this Annex all uses of Coherence refer to this specific axiom.
>
> (*Axiom 3: Non-dogmatism*) Before evidence, we assign some positive credence to all possible theories: prior to evidence $Cr(h)>0$ so long as h is not a logical contradiction ($h \neq A \wedge \neg A$)

From Coherence we can derive a further condition of upper and lower boundedness (our credence in any theory is at least as great as our credence in logical falsehoods, and not greater than our credence in logical truths):

(*Boundedness*) $Cr(h \wedge \neg h) \leq Cr(h) \leq Cr(h \vee \neg h)$ for all h.[18]

For convenience we can map all credences to the unit scale, such that $Cr(h \vee \neg h)=1$. [We already know that $Cr(h \wedge \neg h)=0$ from A1 as $h \wedge \neg h \rightarrow h$ therefore $Cr((h \wedge \neg h) \vee h) = Cr(h \wedge \neg h) + Cr(h)$.] Mapping to a specific interval is convenient, but not strictly required, as we could simply express all credences as fractions of our credence in all logical truths.

Finally, under the 'Independence of Irrelevant Alternatives', if one learns about an individual that 'r>b', 'r=b', or 'r<b', then one's credence in 'B>G' should not change. That is to express this formally:

Independence of Irrelevant Alternatives (IIA): $Cr(B>G \,/\, b(i)>r(i)) = Cr(B>G \,/\, b(i)=r(i)) = Cr(B>G \,/\, b(i)<r(i)) = Cr(B>G)$ for all overall amounts B, G, and individuals i where it is not that G=R.

If A1, A2 and A3 then IIA must be false. That is, for any non-dogmatic well-defined credence function, adopting Arrow's Independence of Irrelevant Alternatives entails the possibility of incoherent beliefs. Here is one proof. Let us define the overall quantity of something as the sum of its individual amounts (*Def 1:* Let R=r(i) + r(j)..... + r(n) for all individual instances i, j....n) and before any evidence, consider three possible overall quantities, A, B and C.

Case one: There exists any two credences such that $Cr(A>B) \neq Cr(A>C)$ or $Cr(A<B) \neq Cr(A<C)$.

That is either (i) $Cr(A>B) > Cr(A>C)$; (ii) $Cr(A>B) < Cr(A>C)$; (iii) $Cr(A<B) > Cr(A<C)$; or (iv) $Cr(A<B) < Cr(A<C)$. If so, imagine we

[18] Proof: As $(h \wedge \neg h) \rightarrow h$ therefore $Cr(h \wedge \neg h) \leq Cr(h)$. As $h \rightarrow (h \vee \neg h)$ therefore $Cr(h) \leq Cr(h \vee \neg h)$. (In fact, given Additivity, not only does Coherence entail Boundedness, but the converse is true too. Proof: As $Y \rightarrow \neg(\neg Y)$ therefore $Cr(Y) + Cr(\neg Y) = Cr(Y \vee \neg Y)$ [From A1]. If $X \rightarrow Y$ then $Cr(X \vee \neg Y) = Cr(X) + Cr(\neg Y)$ [From A1]. Thus $Cr(X \vee \neg Y) + Cr(\neg X \wedge Y) = Cr((X \vee \neg Y) \vee (\neg X \wedge Y)) = Cr(Y \vee \neg Y)$. Therefore $Cr(X) + Cr(\neg Y) + Cr(\neg X \wedge Y) = Cr(Y) + Cr(\neg Y)$. As $Cr(\neg X \wedge Y) \geq 0$ [From Boundedness] therefore if $X \rightarrow Y$ then $Cr(Y) \geq Cr(X)$ QED.)

learn body of evidence E1, that each individual instance of B is equal to each accompanying instance of C. That is, let Cr(E1)=1 where E1=[b(i)=c(i); b(j)=c(j); ... b(n)=c(n)] for all individuals i, j...n. Therefore Cr(B=C) = 1[From Def 1, A1, A2]. Therefore Cr(B>C) = 0 and Cr(B<C) = 0 and thus: (i) Cr('A>B' ∧ 'B=C') = Cr(A>B); (ii) Cr('A>C' ∧ 'B=C') = Cr(A>C); (iii) Cr('A<B' ∧ 'B=C') = Cr(A<B); (iv) Cr('A<C' ∧ 'B=C') = Cr(A<C) [A1, A2]

Under IIA, E1 is irrelevant to the choice between A and B and to the choice between A and C. Therefore: (Cr('A>B' / E1) = Cr(A>B); Cr('A>C' / E1) = Cr(A>C); Cr('A<B' / E1) = Cr(A<B); Cr('A<C' / E1) = Cr(A<C). That is, under IIA, in each of the respective cases (i)-(iv): (i) Cr('A>B' ∧ 'B=C') > Cr('A>C' ∧ 'B=C'); (ii) Cr('A>B' ∧ 'B=C') < Cr('A>C' ∧ 'B=C'); (iii) Cr('A<B' ∧ 'B=C') > Cr('A<C' ∧ 'B=C'); (iv) Cr('A<B' ∧ 'B=C') < Cr('A<C' ∧ 'B=C').

However, ('A>B' ∧ 'B=C') and ('A>C' ∧ 'B=C') are logically equivalent theories, as are ('A<B' ∧ 'B=C') and ('A<C' ∧ 'B=C'). We have incoherent beliefs. ['Coherence' is violated: each of the logically equivalent theories entails the other, thus their credences must be the same].

Case two: There do not exist any two credences such that Cr(A>B) ≠ Cr(A>C) or that Cr(A<B) ≠ Cr(A<C) for all A, B, C.

It therefore follows that Cr(A=B) = Cr(A=C) where Cr(A=B) > 0 [From Case two, A1, A2, A3]. As such Cr(A=C) + Cr(A>C) > Cr(A>B). Lemma 1.

Imagine we now learn body of evidence E2, that each individual instance of C is greater than each accompanying instance of B. That is, let Cr(E2) = 1 where E2 = [c(i)>b(i); c(j)>b(j); ….. c(n)>b(n)] for all individuals i, j .... n. As a result Cr(C>B) = 1 and Cr(B≤C) = 0 [From Def. 1, A1, A2]. As such Cr('A=C' / E) + Cr('A>C' / E) ≤ Cr('A>B' / E). Lemma 2.

However, under the IIA, "E" is irrelevant to ranking beliefs concerning A and C, and also to those concerning A and B. Hence Cr('A=C' / E) + Cr('A>C' / E) = Cr(A=C) + Cr(A>C) and Cr(A>B /E) = Cr(A>B). If so, we have incoherent beliefs (in combination with Lemmas 1 and 2, 'Coherence' is violated). QED.

That is, no matter what one's initial credences, so long as they are non-dogmatic and well-defined, then adopting IIA entails the possibility of incoherent beliefs. (And while this result requires three or more arrangements to hold, if there are less than three arrangements then the IIA itself is irrelevant: there are no independent alternatives).

Arrow's Independence of Irrelevant Alternatives should not be a condition upon beliefs about how the overall amount of something relates to its constituent components if such beliefs are to be justified, and as such should not be a condition on how we form beliefs about overall welfare or the overall good.