Comparing Evaluations

Richard Bradley
Department of Philosophy, Logic and Scientific Method
London School of Economics
Houghton Street
London WC2A 2AE
r.bradley@lse.ac.uk

January 7, 2008

Abstract

This paper explores the problem of comparing the strengths of different individual's attitudes, and especially their evaluative attitudes, by looking at how measures of these quantities are obtained. I argue that comparisons of both strengths of belief and relative strengths of preference and desire are justified by the causal role they play in the production of action.[1]

1 Introduction

It is commonplace that we compare people's beliefs and evaluations not only for their content but also for their strength, making such claims as that one person is less convinced than another that the picnic will be washed out or that one person's preference for tomato over cucumber sandwiches is much stronger than another's preference for cucumber over tomato. But while the consensus seems to be that comparisons of strength of belief raise no particular theoretical difficulties, however difficult they might be in practice, the same cannot be said of evaluations. On the contrary, there is a long tradition of scepticism about the empirical meaningfulness of comparisons of the strength of agents' desires, preferences or value judgements and correspondingly lively debate about the grounds for the comparisons that people are observed to make.

This debate has played an especially significant role in welfare economics and social choice theory. Arrow [1], for instance, took as a starting point for his famous impossibility theorem that interpersonal comparisons of utility were not meaningful, while many have argued that the 'solution' to this impossibility lies in the rejection of precisely this premise. The debate has not been aided by the presence of more than one interpretation of the concept of utility, as is clearly exhibited in Jevons' oft-quoted position on the matter:

> Every mind is ... inscrutable to every other mind, and no common denominator of feeling seems to be possible, ... the motive in

> one mind is weighed against other motives in the same mind, never against motives in other minds. - ([10, Introduction])

Jevons, and Robbins after him, argued that it was not possible to verify on the basis of empirical evidence whether or not one person experienced the same satisfaction or happiness as another from some state of affairs, a claim that has some resonance in contemporary debates about qualia. But whether or not this is true, it is largely irrelevant to the question of whether *motives* can be compared, for the satisfaction that someone gains from bringing about a state of affairs is not their motive for doing so, except insofar as their preferences for actions are formed in the light of true beliefs about the satisfaction these actions will secure. Nor do the qualitative aspects of our cognitive and evaluative attitudes play any direct role in motive-based explanations of behaviour of the sort typically provided by economists and other social scientists, as is easily seen by the fact that a single course of action may be utility maximising for two agents (which explains why they choose it), even if the feeling that accompanies their attitude to it is quite different. Degrees of belief and desire figure in these explanations as causal factors disposing an agent towards the performance of certain actions (and away from others) and it is the strength of these factors that are picked out by the standard choice-based procedures used to measure degrees of belief and desire.

It is the comparison of strengths of evaluations, qua causal factors in the determination of action, that will be our concern. A different sort of argument for incomparability is often made in this context. Arrow, for instance, argued that utility was not cardinally measurable and that if "... we cannot have measurable utility, in this sense, we cannot have interpersonal comparability of utilities a fortiori" [1, p.9]. The view that the empirical meaningfulness or otherwise of interpersonal comparisons depends on the measurability of the individual attitudes that are being compared is a common one, even amongst those who do not share Arrow's scepticism about the measurability of utility. The fundamental difference between comparisons of beliefs and evaluative attitude is one of its corollaries.

In this paper, I will examine and criticise this view, arguing that careful examination of how attitudes are measured shows that the problem of comparing beliefs and that of comparing evaluations are good deal more similar than usually thought. This in turn will provide a basis for explaining how meaningful comparisons of evaluations might be possible. In particular I shall argue that the construal of preferences as causal factors in action implies that interpersonal comparisons of relative strengths of preference are empirically meaningful.

2 The Problem

Let me begin by describing the framework within which our problem arises. The focus of our concern are the attitudes, or judgemental states, of agents of the sort we ascribe to them when offering intentional explanations of their

behaviour. I shall assume that there are at least two different classes of such attitudes: the cognitive and the evaluative. The former class includes attitudes such as believing, accepting or supposing, the latter such attitudes as desiring, preferring and valuing. To keep things simple we will work here with just one representative of each class; namely beliefs and preferences.

Cognitive and evaluative attitudes have both a content - the prospect towards which they are directed - and a force or strength. Issues of comparability could in principle arise with respect to both dimensions but since it is not really necessary to determine the content of an agent's attitudes directly we can treat these dimensions together. An analogy may make this claim clearer. Suppose we wanted to describe the movement of an object through space by giving its speed and direction of motion. One way of doing so would be to state the distance covered by the object in one time unit in each of the three directions: laterally, horizontally and vertically. This would determine exactly where the object was headed and at what speed relative to frame of reference of the observer. In a rather similar way an agent's belief state (or desire state) may be represented by a real-valued measure of her strength of attitude towards each possible prospect, with the number associated with any prospect specifying the strength of the agent's attitude towards that prospect.[2]

Why should utilities be regarded as non-comparable, but not probabilities? The 'standard' argument for the impossibility of interpersonal utility comparisons is most perspicuously presented, as it is in List [12], as an argument about empirical underdetermination. In a nutshell, the claim is that utility measures of agents' preference strengths are underdetermined by the available empirical evidence in a way that probability measures of their degrees of desire are not. The empirical basis for measurement is thought to consist primarily in what we are able to observe, or could in principle observe, by way of choices made by individuals between different prospects. Interpreted generously this basis suffices to determine, by means of a method that will be described in more detail later on, a unique measure of an individual's degrees of belief and a measure of her degrees of desire that is unique up to a choice of scale. The lack of uniqueness in the measurement of desire means that the number that we should assign as a measure of someone's degree of desire for some prospect depends on our choice of scale for the utility function being used to measure degrees of desire and in particular on which prospects we assign a measure of zero and one. No such a choice of scale is available when measuring degrees of belief, for by definition a probability assigns measure zero to the impossible or contradictory prospect and measure one to the necessary or tautologous prospect.

This has implications for comparisons. When comparing the probabilities of two individuals we are dealing with quantities that are measured on the same scale. In contrast, when we say that Anne has a utility of (say) 2 for some prospect X we are in part representing features of her attitudes and in part expressing an arbitrary choice of scale. So the fact that Bob has a utility of

[2] It might be possible for an agent to have no attitude towards some prospects, as opposed to merely having an attitude with no force. In this case we can let the function take some abitrary non-numerical value.

2 for prospect Y does not mean that he prefers that Y to the same degree that Anne prefers that X, unless these utilities are measured on the same scale. But there is no way of telling when this is the case. It follows that we cannot say when Anne and Bob's preference strengths are the same. To be sure we could choose two prospects to serve as the zero and units of both Anne's and Bob's utility measures (so long as the order of their preference for the two was the same). But this would not guarantee that the same quantities were being measured by the same numbers. The unit prospect might, for instance, be one that Anne dislikes and Bob likes.

There have been two main kinds of response to the problem as we have described it. The first consists in augmenting the empirical basis for attributing and comparing attitudes and has been pursued by, amongst others, Waldner [18], Harsanyi [5] and List [12]. The second consists in introducing non-empirical considerations into comparisons, either of a normative kind (as do, for instance, Schick [16] and Jeffrey [7]) or of a conceptual kind (as do, for instance, Harsanyi [5], Davidson [3] and Hausman [6]).

In general I do not see much hope for strategies of the first kind. Suppose that we have to hand some new kind of interpersonally comparable information about Anne and Bob to help in determining a co-scaling of their utilities e.g. concerning their brain chemistry or their childhoods. The question now is how to map this information to possible utility measures of their preference strengths in a manner which settles the question of how to co-scale them. It is hard to see how this could be done without having already solved the problem of interpersonal comparisons. If both Anne and Bob were for instance in the same relevant state (of having a certain brain chemistry, or a certain childhood) would this make it correct to assign the zero and one measures to the same prospects on their preference rankings? Only if we knew that in doing so we were in each case identifying the zero and unit with prospects preferred to the same degree by Anne and Bob. And we cannot know this unless we are able to compare their utilities.

Strategies of the second kind recognise that the bare facts alone cannot determine some particular way of equating utility scales of different people and aim to provide some extra-empirical ground for doing so. Some look to moral norms. Schick, for instance, argues that commitment to treating individuals on a par requires us 'assimilate' different individual's utilities by assigning zero to the least preferred prospect of each person and one to the most preferred prospect - a procedure known as the zero-one rule. As Jeffrey succinctly notes, it is far from clear that concerns of equal consideration require this rule when, for instance, the death of Bob is Anne's most preferred prospect and Bob's least. Nonetheless he too argues that the preferences of different individuals may be 'collated' by means of judgements that refer both to the observable facts and to norms of equal treatment. But while I have no doubt that we often make implicit use of moral norms in making interpersonal comparisons, I don't think this can be the whole story. For it seems to follow that a clean separation cannot be made between the question of how strongly or otherwise different agents prefer one or another prospect from the question of what moral weight

we should give to the fact that they do so. If, for instance, norms of equality or fairness must be appealed to in comparing your preference for going to an Indian restaurant to my preference for a French one, then having arrived at the conclusion that your preference is stronger, we cannot then ask further whether fairness requires us to choose Indian. But we sometimes do want to raise this further question.

No such a disadvantage would attach to the strategy of appealing to conceptual considerations, and it is just such a path that will be pursued here. Others have followed it before. Harsanyi and Davidson, for instance, posit the existence of a priori principles - respectively the principle of unwarranted differentiation and the principle of charity - that require in different ways that similarity between individuals in their evaluations be maximised to the extent allowed by evidence of difference. Unfortunately neither really explains how these principles solve the problem which precisely resides in the fact that we cannot tell whether an ascription of the same utility value to two individuals for some prospect counts as maximising sameness or not, unless we have some basis for equating their utility scales.

Hausman's argument for the zero-one rule is more helpful in this respect. He argues that (a) it belongs to the concept of degrees of preference satisfaction that total satisfaction and total dissatisfaction are interpersonally comparable states, (b) an agent's most preferred prospect is the one which totally satisfies her preferences and her least preferred one is the one which totally fails to satisfy them, and hence (c) the top and bottom of preferences rankings can be equated in accordance with the zero-one rule. What Hausman does not give is an argument for identifying degrees of preference satisfaction with preference strengths; nor, I think, can such an argument be given. For, as I shall argue later on, there is no reason to believe that the strength of the attitude of one individual to their most or least preferred prospect should be the same as another individual's attitude to theirs. But to make this argument, we must first look at how such strengths of evaluations are measured.

3 Measurement

In this section I will review the central elements of a procedure for measuring strengths of beliefs and desires that is explicitly based on the idea that these quantities are determined by the causal role they play in disposing the agent to act in certain ways. The procedure depends in a rather direct way on the kind of theory to which it is applied; namely that which takes an agent's choice of action to be determined by her expectation of benefit arising from its performance. The idea at its simplest is that the stronger an agent's preferences for an action's consequences and the stronger her belief in the pertaining of some precondition for the success of an action with desirable consequences, the stronger her disposition to choose it.

The version that I will draw on is more or less that sketched out by Ramsey (in [13]), but the essential parts are shared by all those methods that identify

strengths of attitudes with causal components of actions. Ramsey's own account is rather vague on the relationship between the various objects of agents' attitudes that he postulates (worlds, propositions and conditional prospects) and we shall instead postulate a single set of prospects to which agents can take both belief and desire attitudes. The set includes ordinary factual prospects such as that inflation will rise or that it will rain tomorrow as well as conditional prospects such as that agricultural prices will rise if crops are damaged and that we will be at the theatre on time if we take a taxi.

Let us represent prospects by capital letters, reserving F and T for the impossible and necessary prospects respectively. The prospect that both A and B will be written as AB, that A is not the case as $\neg A$, and the conditional prospect of A if P as $P \rightarrow A$. The logical relations between these prospects - or more exactly the sentences denoting them - place important constraints on the attitudes that agents can consistently take to them and when these matter I will identify the constraints being assumed.

The empirical basis for this account of the measurement for belief and preference strengths consists in what we are able to observe by way of an agent's choices amongst prospects and any verbal testimony as to their preference between them. It is typically supposed that empirical evidence of this kind suffices to determine a ranking that represents the agent's preferences over all prospects. The agent's (strict) preference that A rather B is written as $A > B$ and her indifference between the two as $A \approx B$. The expression $A \geq B$ means that $A > B$ or $A \approx B$.

To move from these preference orderings to numerical measures on prospects, Ramsey introduces the very useful notion of an ethically neutral proposition: a proposition whose truth or falsity is a matter of indifference to the agent irrespective of what else is the case. Formally:

Definition 1 *Prospect P is ethically neutral iff for all prospects A:*

$$PA \approx A$$

That a tossed coin lands or a rolled die lands on a five are typically taken to be instances of ethically neutral prospects, and though it is unlikely that they are really pure cases, they offer good enough approximations in most contexts to serve here.

The next step is the identification of a particular class of ethically neutral prospects, namely those that are as likely to be the case as not. Now if someone were indifferent as to whether a coin landed heads or tails and regarded each as equally likely then they should be indifferent between such prospects as winning £10 if the coin lands heads and nothing if it lands tails and winning £10 if the coin lands tails, but nothing if it lands heads. For if heads were more likely than tails, she should prefer the prospect of £10 in the event of heads to that of £10 in the event of tails. More generally we can say:

Definition 2 *Ethically neutral propositions P and Q are equi-probable iff for all prospects A and B:*

$$(P \rightarrow A)(\neg P \rightarrow B) \approx (Q \rightarrow A)(\neg Q \rightarrow B)$$

Whenever P and $\neg P$ are ethically neutral and equiprobable the prospect $(P \rightarrow A)(\neg P \rightarrow B)$ may be regarded as the midpoint in the preference ranking between prospects A and B. For in this case the realisation of the prospect makes it as likely that A as that B. Hence an arbitrary choice of utility values for the latter - say 1 and 0 respectively - will determine a utility value for the midpoint prospect $(P \rightarrow A)(\neg P \rightarrow B)$ and all other prospects C ranked with it - in this case 0.5. The midpoints between A and C and between C and B are in turn identified by $(P \rightarrow A)(\neg P \rightarrow C)$ and $(P \rightarrow C)(\neg P \rightarrow B)$ and are assigned utilities 0.25 and 0.75 respectively. Similarly utilities of 2 and -1 are assigned to prospects D and E such that the agent is indifferent between A and $(P \rightarrow D)(\neg P \rightarrow B)$ and between B and $(P \rightarrow A)(\neg P \rightarrow E)$. And so on.

By this method, and others of a similar ilk, measures of the agent's degrees of desire for all prospects are determined once an assignment of values has been made to the chosen reference prospects A and B. In this sense, utilities are measures of the agent's degrees of desire for prospects *relative* to the reference ones. More exactly, Ramsey's measurement process fixes only ratios of utility differences. In contrast, the related procedure for measuring degrees of belief fixes the scaling and hence determines a unique measure. Here is how.

Let P be any proposition (not necessarily ethically neutral) and let U be any measure of the agent's degrees of preference of the kind whose construction we have just described. Suppose that A, B and C are such that A and B respectively imply that P and $\neg P$, that $A > C > B$ and that the agent is indifferent between $(P \rightarrow A)(\neg P \rightarrow B)$ and C. Indifference between the two is rational only if difference between the gain (in utility terms) from C being the case rather than B and the gain from A being the case rather than C is exactly offset by the difference between the probability of P and $\neg P$. More exactly this indifference implies that the agent's odds on P are exactly equal to the ratio of the utility difference between C and B and the utility difference between A and C. Hence we may define a measure of the agent's degrees of belief, Pr, as follows:

Definition 3 *(Probability)*

$$\Pr(P) := \frac{U(C) - U(B)}{U(A) - U(B)}$$

It follows immediately from this definition that $\Pr(P) \geq 0$ and that $\Pr(\neg P) = 1 - \Pr(P)$. (It is also follows from the assumptions that Ramsey makes about preference that Pr is additive, though the proof of this is far from trivial). More to the point in this context, given that the function U is unique up to a choice of scale, the definition of Pr implies that the measure of an agent's degrees

of belief obtained in this way is completely unique. Stronger still, given that $(T \rightarrow A)(F \rightarrow B)$ is logically equivalent to A, the definition implies that:

$$\begin{aligned} \Pr(T) &= \frac{U(A) - U(B)}{U(A) - U(B)} = 1 \\ \Pr(F) &= \frac{U(A) - U(A)}{U(C) - U(B)} = 0 \end{aligned}$$

so that the scaling of the measure is determined independently of the choice of utility measure of the agent's degrees of desire and of its uniqueness.

4 Comparisons

The 'Ramseyian' method allows for a unique measure of strength of belief, but measures of strength of preference that are unique only up to the choice of scale. It is tempting to move from this fact to the conclusion that strengths of belief are comparable across individuals, but strengths of preference are not. In fact, however, neither inference - to comparability from measurability and to non-comparability from non-unique measurability - is beyond criticism. That comparability does not presuppose unique measurability is in fact widely accepted; witness the attempts to provide non-empirical grounds for comparisons. But it is also the case that unique measurability does not provide grounds for comparability.

Consider the case of measuring degrees of belief. If we look closely at how unique measures for degrees of belief are determined by the 'Ramseyian' method, it is apparent that the procedure itself *assumes* comparability rather than establishes it, at least of states of full belief and disbelief. For the definition of the measure Pr (Definition 3) was chosen precisely so as imply that $\Pr(T) = 1$ and $\Pr(F) = 0$, a normalisation that implicitly assumes that the property of beliefs picked out by the zero and unit measures is the same for everyone. Hence, the uniqueness of the measure that depends on this normalisation cannot provide non-circular grounds for inferring that degrees of belief are comparable.

In general for comparability across persons it must be that the measurement made on one person picks out the same property of that person as does as an equivalent measurement on another person. This is so even if we accept that the basis for ascribing the property to someone is not purely empirical. Now it should be clear that the uniqueness of the measurements made on each person does not in itself guarantee that what is being measured in each person is the same, for it could be the case that the value of the measurement made on each individual is determined by individual-specific criteria or that the criteria pick out different properties when applied to different people. For interpersonal comparability we must have universal criteria - ones that apply to all - for determining the value of the relevant measurement. And these criteria must pick out the same property in each person when applied to them. Whether such criteria exist will in part depend on the type of property involved and in part on the nature of the evidence base.

What entitles us to assume interpersonal comparability of belief is not its unique measurability but that fact that within the theory of action which serves as a background for the measurement procedure, strengths of belief play the same causal role in one person as another. This is manifested by the existence of common criteria - ones that apply to all - for specifying a person's degrees of belief. A universal condition on preference for someone regarding two ethically neutral prospects as equally likely has already been given (in Definition 2). But common conditions can also be supplied for fully believing and full disbelieving that P, namely:

Condition 4 *Full Belief in P:* $(P \rightarrow A)(\neg P \rightarrow B) \approx A$

Condition 5 *Full Disbelief in P:* $(P \rightarrow A)(\neg P \rightarrow B) \approx B$

The first condition says in effect that strength of belief in P is maximal when the agent attaches no weight to the prospect that if $\neg P$ then B, no matter how desirable B is, while it is at a minimum when the agent attaches no weight to the prospect that if P then A, no matter how desirable it is that A. Since full belief and empty belief, as identified by these conditions, play the same causal role in each individual, these states may be equated across persons. So too with the state of regarding two prospects as equi-probable. (Conventionally, full belief and disbelief are identified by measures of one and zero respectively, a choice that then forces an assignment of one-half to any prospect that is regarded as likely or not to be true. But this choice is just a matter of mathematical convenience and one could equally well choose, for instance, an assignment of zero for the class of prospects that are believed as likely to be true as not and one and minus one for the most and least believed prospects.)

Sauce for the goose is sauce for the gander. Preferences too are causal factors of action and we may ask whether comparability of preference strength may not be justified in the same way that it is for strength of belief. In fact we already have a universal criterion for one state of preference; namely for when an agent is neutral towards a prospect, being disposed neither to make it or its contrary true. The criterion in question is that of the agent being indifferent between the prospect and its contrary (Definition 1). Crucially, the utility measure on an ethically neutral proposition is comparable across persons because it picks out the same causal force (or rather, the same absence of net force) in the production of action, everyone being equally disposed to promote the realisation of a neutral prospect as to hinder it. In this respect ethically neutral prospects are the analogues of the equi-probable ones: they represent points where the causal strengths of the attitudes in question are equally balanced.

Can considerations of causal role take us any further? In the case of belief, we were able to identify not only a common midpoint in credibility (the prospects considered to be as probable as their contraries), but also common upper and lower bounds marking states of full belief and disbelief. Similarly if the preference ranking has an upper and lower bound - the most and least preferred prospects - then one could define states of 'full' preference and its opposite as follows.

Condition 6 *Full Preference for X: for all prospects Y, $X \geq Y$*

Condition 7 *Full Dis-preference for X: for all prospects Y, $X \leq Y$*

Criteria such as these could be used to define a common zero-one scale for measuring and comparing degrees of preference strength. But stating such criteria is not the same as justifying them and in this regard the situation is less straightforward in the case of preference than that of belief. In the latter case we were able to claim that the unit measure on T and zero measure on F pick out the same property of individuals because it belongs to the concept of partial belief that it has both a maximum (full belief) and a minimum (no belief) and further that a rational individual will fully believe something that is necessarily the case and give no credence to something that is impossible. In contrast, it is no part of the concept of evaluative attitudes such as preference that either a common maximum or minimum exists, for there are no prospects that we are all rationally required to maximally value or disvalue. Nor does it belong to the concept of preference that each individual's evaluation of the prospect that they rank highest (or lowest) is the same. Of all the items on the menu, I might like the chicken most and you the fish, but it would not follow that my preference for chicken was as strong as yours for fish.

This objection might be defused to some extent if we read the conditions for full preference and dis-preference as licensing, not an equating of the maximal and minimal elements of the rankings of any arbitrary set of prospects, but of the maximal and minimal elements of the rankings of the set of *all conceivable* prospects.[3] The thought then would be that each person's attitude to their most (least) preferred prospect may be interpreted as one of full preference (dis-preference) in just the sense that there is nothing that they could prefer (dis-prefer) to it. Although the prospect towards which this attitude is directed may differ across individuals, the attitude itself is the same and may thus be equated.

This argument, were it convincing, would provide conceptual grounds for use of the zero-one rule in comparing strengths of preferences. Unfortunately however I do not think that it is. Firstly, it fails to establish that the maxima and minima of different individuals' preference rankings play the same casual role in the production of action in the way that the maxima and minima on the measures of different individuals' beliefs do. Whether my chicken-preferences dispose me to choose chicken on the menu to the same degree that your fish-preferences dispose you to make the fish choices depends, for instance, on how we regard these top items in relation to our second most preferred ones. If turkey is a close second for me, but whale a distant second for you, then my disposition to pursue chicken at all costs will be considerably weaker than your disposition to seek out fish. So while we might be able to state a common criterion for something called full preference, we have no basis for assuming that what we have defined picks out the same causal property of every individual.

[3] There are obvious practical difficulties here, not least that what different individuals are capable of conceiving may differ. But we set them aside here.

Secondly, the criterion for full preference is so demanding that it is doubtful that it is ever satisfied. For unless the set of prospects is finite, there need be no most preferred prospect. Indeed there is reason to think that there would not normally be. For suppose X was a candidate for the most preferred prospect. Then let X' be the prospect of X and my being given an extra hour of happy life or an extra £1 (over and above that associated with X). Unless X satiates my desire for wealth and longevity, I will prefer X' to X. Perhaps these kinds of desires could be satiated, but I see no reason to believe that all my desires should be. Similarly, note that the supremum of such a set of prospects would have to be unconditionally desirable i.e. if X were the supremum then for every prospect Y, both XY and $X\neg Y$ would be ranked as high as X (as long as both Y and $\neg Y$ are consistent with X). For if this were not the case, either XY or $X\neg Y$ would have to be ranked above X.[4] But I cannot myself find any candidates for such an unconditionally desirable state whose descriptions do not beg the question of their existence.

Even if we cannot rest interpersonal comparisons of preference strength on the existence of maximal and minimal elements of preference rankings, causal considerations do give us some purchase on interpersonal comparisons of preference strength. For, as we have seen, attitudes to ethically neutral prospects are comparable across persons in virtue of the common causal role that they play. Considerations of convenience thus suggests a common assignment of the zero to all such prospects and, in particular, to the necessary prospect T. (Note though that, in contrast to the belief case, a prospect of utility zero is not the least element of the preference ordering, there being prospects worse than the ethically neutral ones.)

Once the zero for the utility scale has been fixed, ratios of utilities are uniquely determined by the measurement procedure. Hence it may be said to determine a unique (scale-independent) measure of *relative* strengths of preference. Given our background theory of action, comparisons of relative strengths of preference are meaningful. To say that my strength of preference for X relative to Y is the same as your strength of preference for W over Z is to say that these quantities measure equal causal forces in the production of our actions; it explains why ceteris paribus I am as inclined to favour X-promoting actions over Y-promoting ones to the same degree as you favour W-promoting actions over Z-promoting ones. I see thus no conceptual or technical obstacle to interpersonal comparisons of utility ratios. No technical obstacle because these ratios are uniquely determined by our measurement procedure once the zero point of the utility scale has been identified with the desirability of T; no conceptual obstacle because the measurement of the relative strengths of a causal factors determining agents' actions should not depend on whose actions are being caused.

To argue that relative strengths of preference are comparable, like strengths of belief, on grounds of common causal role, is not to rule out the possibility

[4] This follows from the axiom of averaging: if $XY = F$, then $X > Y$ iff $X > X \vee Y > Y$. See Jeffrey [9].

of providing other grounds for equating absolute strengths of preference. But I would suggest that such further grounds will be of a normative character. Furthermore, for many purposes, comparisons of relative strengths are sufficient: for example in reaching aggregate judgements about the relative desirability of two courses of action. Both these claims deserve further exploration, but to do so would take us beyond our initial problem. In this latter regard we can now conclude that they standard view is only half-correct. Comparisons of absolute preference strengths are indeed empirically underdetermined (by even the most generous of sets of evidence), though comparisons of relative preference strengths are not. This marks a difference with comparisons of strengths of belief. But this difference is not to be explained in terms of differences in measurability. For both differences reflect the same fact, namely that while strengths of belief are causally relevant in the production of action, only relative strengths of preference are.

References

[1] Arrow, K. E (1951, 1963) *Social Choice and Individual Values*, New Haven and London: Yale University Press

[2] Bradley, R. W. (2001), "Ramsey and the Measurement of Belief" in D. Corfield and J. Williamson, eds., *Foundations of Bayesianism*, Dordrecht: Kluwer, pp. 263-290

[3] Davidson, D. (1986) "The Interpersonal Comparison of Values" in J. Elster and A. Hylland, eds., *Foundations of Social Choice Theory*, Cambridge: Cambridge University Press, 195–211.

[4] Davidson, D. (2004) "Expressing Evaluations" in Davidson, D. *Problems of Rationality*, Oxford: Oxford University Press: 19-38

[5] Harsanyi, J.C. (1955) "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility", *Journal of Political Economy* 63: 309-21

[6] Hausman, D. (1995) "The Impossibility of Interpersonal Utility Comparisons", *Mind* 104: 472-490

[7] Jeffrey, R.C. (1971) "On Interpersonal Utility Theory", *Journal of Philosophy* 68: 647-56

[8] Jeffrey, R.C. (1992) *Probability and the Art of Judgement*, 2nd edn, Chicago: University of Chicago Press

[9] Jeffrey, R.C. (1983) *The Logic of Decision*, Cambridge: Cambridge University Press

[10] Jevons, W.S. (1871, 1970) *The Theory of Political Economy,* London: Macmillan

[11] Krantz, D. H., R. Duncan Luce, P. Suppes and A. Tversky (1971) *Foundations of Measurement*, vol.1, London: Academic Press

[12] List, C. (2003) "Are Interpersonal Comparisons of Utility Indeterminate?", *Erkenntnis* 58: 229-260

[13] Ramsey, F. P. (1926) "Truth and Probability" in D. H. Mellor, ed., *Philosophical Papers*, Cambridge: Cambridge University Press, 1990.

[14] Robbins, L. (1938) "Interpersonal Comparisons of Utility: A Comment", *The Economic Journal* 48: 635-641

[15] Rossi, M. (2006) "Interpersonal Comparisons of Utility", *Choice Group working papers* 2 (2).

[16] Schick, F. (1971) "Beyond Utilitarianism", *Journal of Philosophy* 68: 657-66

[17] Sen, A. (1982) *Collective Choice and Social Welfare*, Edinburgh: Oliver and Boyd

[18] Waldner, I. (1972) "The Empirical Meaningfulness of Interpersonal Utility Comparisons", *Journal of Philosophy* 69: 87-103

[1] I am indebted to Mauro Rossi and Christian List for their helpful comments on an earlier draft.