This is a preprint of a paper published in Erkenntnis 58(2): 229-260 (2003).
http://dx.doi.org/10.1023/A:1022094826922

Are Interpersonal Comparisons of Utility Indeterminate?

Christian List

29 May 2002, forthcoming in *Erkenntnis*

Abstract. On the orthodox view in economics, interpersonal comparisons of utility are not empirically meaningful, and "hence" impossible. To reassess this view, this paper draws on the parallels between the problem of interpersonal comparisons of utility and the problem of translation of linguistic meaning, as explored by Quine. I discuss several cases of what the empirical evidence for interpersonal comparisons of utility might be and show that, even on the strongest of these, interpersonal comparisons are empirically underdetermined and, if we also deny any appropriate truth of the matter, indeterminate. However, the underdetermination can be broken non-arbitrarily (though not purely empirically) if (i) we assign normative significance to certain states of affairs or (ii) we posit a fixed connection between certain empirically observable proxies and utility. I conclude that, even if interpersonal comparisons are not *empirically* meaningful, they are not in principle impossible.

1. Introduction

This paper is concerned with our basis for making interpersonal comparisons of utility. Utility can be, and has been, interpreted in many different ways. The present argument applies to any conception of utility that has the following properties: (i) utility captures some form of *welfare*; (ii) utility is something that we *attribute* to a person; (iii) utility can be *experienced* (if at all) only from a first-person perspective; and (iv) utility may *surface observably* in the form of a person's choice behaviour and/or other observable proxies.[1]

Examples of interpersonal comparisons of utility are statements of the forms "Person i's utility in state x is at least as great as person j's utility in state y" (an *interpersonal comparison of utility levels*) or "If we switch from state x to state y, the ratio of person i's utility gain/loss to person j's utility gain/loss is λ", where λ is some real number (*an interpersonal comparison of utility units*). Below a third type of interpersonal comparison will be added (a *utility comparison with respect to an interpersonally significant zero-line*).

[1] Although this characterization of utility is deliberately left relatively open, so as to apply to a class of conceptions of utility, not all conceptions of utility fall into it. Even amongst those conceptions of utility satisfying (i), there is a great range of diversity. Elster and Roemer (1991, introduction) identify two dimensions that characterize different conceptions of utility (they actually address "well-being") that may lend themselves to interpersonal comparisons. On one dimension (call it the *subjective-objective* dimension), conceptions are divided into (a) subjective mental states (hedonic satisfaction), (b) degree of objective satisfaction of subjective desires, and (c) objective states. On a second dimension (call it the *relevance-irrelevance* dimension), conceptions are characterized by the criteria by which states of pleasure or desire-satisfaction are admitted or rejected as admissible components of utility. The present argument aims to be neutral with regard to the second dimension. With regard to the first, however, the present argument applies only to conceptions of type (a), as only those have properties (ii), (iii) and (iv).

In everyday life we often make (what look like) interpersonal comparisons of utility. We *attribute* certain utility levels or utility gains or losses to people and compare these across different people. Interpersonal comparisons play an important role in many choice situations, especially when several people are affected by a decision. We make choices to switch from *x* (e.g. "cooking Marmite paté") to *y* (e.g. "cooking Chocolate crépes") on the basis of whether we believe this switch incurs an immense utility gain for person *i* (e.g. "someone socialized in a Marmite-free part of the world and who finds Marmite revolting") that far exceeds a concurrent very moderate utility loss for person *j* (e.g. "a British Marmite connoisseur"). What exactly is captured by such attributions of utility is far from clear.

The orthodox view in economics is that interpersonal comparisons of utility are not empirically meaningful. Robbins (1932) famously argued that "[i]ntrospection 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 two different people" (p. 140). And Arrow's seminal contribution to social choice theory is premised on the view "that interpersonal comparison of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility" (Arrow, 1951/1963, p. 9). Although the discrepancy between this view and the ease with which we make (what look like) interpersonal comparisons in everyday life has been a continuing source of philosophical puzzlement, the orthodox view (or more refined versions of it) is strikingly persistent. Recently, Hausman (1995), for instance, argued that interpersonal comparisons of utility are impossible unless utility is interpreted as *preference satisfaction*.[2]

These arguments raise at least two different questions, which are sometimes confused with each other. First, are interpersonal comparisons of utility *empirically* meaningful, which we will take to mean: are they *determined*, in a relevant sense (spelled out below), *by empirical evidence*? Second, are interpersonal comparisons of utility *possible*? In particular, if they are not *empirically* meaningful in the sense of being determined by empirical evidence, are they meaningful in some other relevant sense (also spelled out below)?

It is often assumed that a negative answer to the first question (as given by the orthodox view on interpersonal comparisons) entails a negative answer to the second; in short, if interpersonal comparisons of utility are not determined by empirical evidence, *then* they are impossible. Such impossibility conclusions cannot be ignored. The question of whether or not

[2] If utility is interpreted as preference satisfaction, on Hausman's account, identifying the level of satisfaction for each person's top preference with 1 and for each person's bottom preference with 0 is warranted. Arguably, the conception of utility as preference satisfaction is a conception of type (b) on the *subjective-objective* dimension introduced in note 1 above. Hausman also holds that the interpretation of utility as preference satisfaction does not provide the kind of *morally relevant* notion of utility that is required by utilitarian welfare economics.

interpersonal comparisons of utility are possible has far-reaching implications for utilitarian theories of justice and for welfare economics.

Arrow's impossibility theorem (1951/1963) confirms that interpersonal comparisons of utility are relevant to whether or not certain collective decision problems can be solved: Arrow's theorem shows that, if the effects of outcomes on persons are specified in terms of (ordinal) utility (or some other evaluation standard) without interpersonal comparability, there exists no procedure for aggregating such individual utility information into collective preference orderings, where the procedure satisfies some minimal conditions (stated below). If the effects of outcomes on persons are specified in terms of interpersonally comparable utility (or some other interpersonally comparable evaluation standard), Arrow's theorem no longer applies, and there are aggregation procedures satisfying Arrow's minimal conditions (see, amongst many others, Sen, 1970/1979, d'Aspremont, 1985).[3] These results will be briefly reviewed in section 4.

This paper aims to reassess the status of interpersonal comparisons of utility. The paper draws on the parallels between the problem of interpersonal comparisons of utility and the problem of translation of linguistic meaning, as explored by Quine (1960, 1970).[4] According to Quine's indeterminacy of translation thesis (discussed in more detail below), there exist rival schemes of attributing meanings to speakers, where the different schemes are mutually incompatible, and yet equally empirically adequate. Crucially, however, indeterminacy of translation does not imply impossibility of translation. The underdetermination between different schemes of attributing meanings to speakers is broken non-arbitrarily by long-standing linguistic conventions.

I argue that we can reconcile the main insight underlying the orthodox view on interpersonal comparisons of utility – namely that such comparisons are empirically underdetermined – with an account of how interpersonal comparisons are nonetheless possible. Suppose we have a situation of empirical underdetermination between different schemes of attributing utilities to persons, where the different schemes yield mutually incompatible interpersonal comparisons, and yet each scheme is equally empirically adequate. I suggest that this underdetermination can be broken non-arbitrarily in a way that is similar to how the

[3] To pursue this escape-route from Arrow's impossibility theorem successfully we must *either* defend interpersonal comparisons of utility, *or* settle for a welfare-relevant evaluation standard different from utility that *is* interpersonally comparable, such as an index of Rawlsian primary goods or a suitable index of Sen's functionings and capabilities. The question of whether interpersonal comparisons are meaningful in a given sense depends on what evaluation standard we choose to compare: interpersonal comparisons of monetary income are unproblematic (leaving practical issues aside), but maybe not morally relevant, and interpersonal comparisons of the amount of health care or education a person has access to are also unproblematic (again leaving practical issues aside), and maybe more morally relevant in certain contexts. This paper, however, is not committed to any specific view on the question of which evaluation standards are morally relevant and why this is so.

[4] Davidson (1974, 1986) hinted at these parallels without developing them in detail.

underdetermination between different rival attributions of meanings to speakers is broken non-arbitrarily.

The realization that underdetermination, or even indeterminacy, does *not* imply impossibility then tames the problem of interpersonal comparisons of utility. The main insight underlying the orthodox view on interpersonal utility comparisons – namely that such comparisons are empirically underdetermined – remains correct, but its implications are far less negative than commonly assumed. A negative answer to the first of the two questions raised above (are interpersonal comparisons of utility empirically meaningful?) does not force us into a negative answer to the second one (are interpersonal comparisons of utility possible?).

2. Underdetermination and Indeterminacy

To define underdetermination and indeterminacy, I will follow the traditional syntactic approach to theories. Both a theory and a set of empirical observations will be represented as a set of sentences of a formal language. Given a set of (empirical) observation sentences Φ, a theory T is *adequate* with respect to Φ if T implies all the sentences in Φ. In other words, a theory is adequate with respect to a given set of observation sentences if these observation sentences are amongst the ones the theory would have led us to expect, i.e. if they are amongst the implications of the theory. Thus the basic logical relation between theory and empirical observations is a relation of one-way implication. A theory, if it is adequate, implies the observation sentences, but the observation sentences do not in general imply the theory. A theory T (or a specific theoretical statement τ) is *determined* by a set of observation sentences Φ if Φ implies T (or τ). A theory T (or a specific theoretical statement τ) is *underdetermined* by Φ if T (or τ) is consistent with, but not determined by, Φ. If T (or τ) is underdetermined by Φ, then there exists an alternative theory T' (or an alternative theoretical statement τ') such that T' (or τ') is also consistent with Φ, but T and T' (τ and τ') are mutually inconsistent (see also List, 1999).

Underdetermination, thus, is a purely logical concept. Indeterminacy, by contrast, is a metaphysical concept stronger than underdetermination. Given a set of alternative theories and a set of observation sentences Φ, we have a situation of *indeterminacy* if each of the given alternative theories is underdetermined by Φ and there exists no independent fact of the matter as to which of the alternative theories is the 'true' or 'correct' one (for a more detailed account of the relation between underdetermination, indeterminacy and facts of the matter, see Gibson, 1986).[5]

[5] On Quine's account, physical theories are underdetermined by the totality of observable evidence without being indeterminate, while translation is indeterminate, in so far as translation schemes are underdetermined by the totality

3. Profiles of Utility Functions and Interpersonal Comparisons

Let $N = \{1, 2, ..., n\}$ be a set of persons, and $X = \{x, y, x_1, x_2, y_1, y_2, ...\}$ a set of options or states of affairs (for simplicity, we assume that X is finite or denumerable). A *profile of utility functions* $\{u_i\}_{i\in N}$ is an assignment of one utility function $u_i : X \rightarrow \mathbf{R}$ to each person $i\in N$. For each $x\in X$, $u_i(x)$ is interpreted as the utility experienced by person i in response to option x or in state x.

An *interpersonal comparison of utility levels* is a statement of the form "Person i's utility in state x is at least as great as person j's utility in state y", formally

(LC) $u_i(x) \geq u_j(y)$, where $i, j\in N$, $x, y\in X$, $i\neq j$.

An *interpersonal comparison of utility units* is a statement of the form "The ratio of [person i's utility gain/loss if we switch from y_1 to x_1] to [person j's utility gain/loss if we switch from y_2 to x_2] is λ", where λ is some real number, formally

(UC) $\dfrac{u_i(x_1) - u_i(y_1)}{u_j(x_2) - u_j(y_2)} = \lambda$, where $i, j\in N$, $x_1, y_1, x_2, y_2\in X$, $i\neq j$ and $\lambda\in \mathbf{R}$.[6]

We will add to these two familiar types of interpersonal comparisons a third, less familiar one (List, 2001). A *utility comparison with respect to an interpersonally significant zero-line* is a statement of the form "Person i's utility in state x is greater than/equal to/less than a utility level of zero", formally

(ZC) $sign(u_i(x)) = \delta$, where $i \in N$, $x \in X$ and $\delta\in \{-1, 0, 1\}$,

where the *sign*-function is a function $sign : \mathbf{R} \rightarrow \{-1, 0, 1\}$ with the property that, for all $t\in\mathbf{R}$, $sign(t) = -1$ if $t<0$, $sign(t) = 0$ if $t=0$, and $sign(t) = 1$ if $t>0$.

(ZC)-statements are meaningful only if a *utility level of zero* is a meaningful concept. A utility level of zero would have to capture a certain 'dividing line', for instance between 'utility' and 'disutility', or between 'pleasure' and 'pain'. Although (ZC)-statements make explicit reference only to one person, they can be interpreted as a form of *interpersonal* comparisons in that they enable us to make comparisons of utility levels between persons with utility level greater than zero, persons with utility level precisely equal to zero, and persons with utility level less than zero.

of relevant observable linguistic behaviour *and* there is no independent fact of the matter as to which of multiple rival adequate translation schemes is the 'true' one.

[6] Where $u_j(x_2) \neq u_j(y_2)$.

Once we have attributed a profile of utility functions $\{u_i\}_{i\in N}$ to the persons in N, we can make (LC)-, (UC)- and (ZC)- statements *relative to that profile*. Whether these statements can be considered meaningful depends on *how unique* the profile $\{u_i\}_{i\in N}$ is. Assume, for instance, that each utility function u_i is unique only up to positive monotonic transformations.[7] By this assumption, we define two profiles $\{u_i\}_{i\in N}$ and $\{u^*_i\}_{i\in N}$ to be *informationally equivalent* if, for each i, $u^*_i = \phi_i(u_i)$ holds, where ϕ_1, ϕ_2, ..., ϕ_n are positive monotonic transformations, possibly different ones for different persons i.[8] This particular assumption about how unique a profile of utility functions is (stated in terms of the conditions under which two profiles of utility functions are defined to be informationally equivalent) is called *ordinal measurability, no interpersonal comparability of levels or units*, (ONC). (LC)-, (UC)- and (ZC)-statements are not in general invariant under the transformations specified by (ONC). Hence these statements are not considered meaningful – they are not independent of which specific profile we select as a representative from amongst a class of informationally equivalent profiles. We will say that interpersonal comparisons, in the form of (UC)-, (LC)- or (ZC)-statements, are meaningful if and only if they are invariant under the class of transformations up to which $\{u_i\}_{i\in N}$ is unique.

Table 1 shows the relation between different classes of transformations and the invariance (or lack thereof) of (UC)-, (LC)- or (ZC)-statements under these transformations (for a survey of different informational assumptions about measurability and interpersonal comparability and their social-choice-theoretic implications, see d'Aspremont, 1985).[9]

[7] A positive monotonic transformation is a function $\phi : \mathbf{R} \to \mathbf{R}$ with the property that, for any s, $t\in\mathbf{R}$, $s<t$ implies $\phi(s)< \phi(t)$. A positive affine transformation is a function $\phi : \mathbf{R} \to \mathbf{R}$ with the property that there exist a, $b\in\mathbf{R}$ (b>0) such that, for all $t\in\mathbf{R}$, $\phi(t) = a + bt$. A positive linear transformation is a function $\phi : \mathbf{R} \to \mathbf{R}$ with the property that there exists $b\in\mathbf{R}$ (b>0) such that, for all $t\in\mathbf{R}$, $\phi(t) = bt$. A sign-preserving transformation is a function $\phi : \mathbf{R} \to \mathbf{R}$ with the property that, for all $t\in\mathbf{R}$, $sign(\phi(t)) = sign(t)$.

[8] Informational equivalence is an equivalence relation (a reflexive, symmetric and transitive binary relation) on the set of all possible profiles of utility functions. There are different ways of defining informational equivalence, as detailed in table 1, where each such definition corresponds to a different way of partitioning the set of all possible profiles of utility functions into disjoint equivalence classes. Once the equivalence relation of informational equivalence has been defined, this definition then induces a corresponding definition of the *informational content* of a given profile of utility functions. Specifically, whenever two profiles fall into the same equivalence class with respect to informational equivalence, they are taken to have the same informational content. This means that only information shared by *all* profiles *within the same equivalence class* (i.e. statements true of *all* profiles within the same equivalence class) can be considered meaningful.

[9] Table 1 focuses on the implications of the choice of a specific class of transformation for the question of whether (UC)-, (LC)- or (ZC)-statements are invariant under these transformations (and thus "meaningful"). For a more detailed discussion of the logical relation between meaningful statements and classes of admissible transformations, see Bossert and Weymark (1996, section 5).

Condition:	The profiles $\{u_i\}_{i\in N}$ and $\{u^*_i\}_{i\in N}$ are informationally equivalent if ...	(LC)-statements	(UC)-statements	(ZC)-statements
(ONC): Ordinal Measurability, No Interpersonal Comparability of Levels or Units	... for each i, $u^*_i = \phi_i(u_i)$, where $\phi_1, \phi_2, ..., \phi_n : \mathbf{R} \to \mathbf{R}$ are positive monotonic transformations	not invariant	not invariant	not invariant
(ONC+0): Ordinal Measurability with an Interpersonally Significant Zero-Line, No Interpersonal Comparability of Levels or Unity	... for each i, $u^*_i = \phi_i(u_i)$, where $\phi_1, \phi_2, ..., \phi_n : \mathbf{R} \to \mathbf{R}$ are positive monotonic and sign-preserving transformations	not invariant	not invariant	invariant
(CNC): Cardinal Measurability, No Interpersonal Comparability of Levels or Units	... for each i, $u^*_i = \phi_i(u_i)$, where $\phi_1, \phi_2, ..., \phi_n : \mathbf{R} \to \mathbf{R}$ are positive affine transformations	not invariant	not invariant	not invariant
(RNC): Ratio-Scale Measurability, No Interpersonal Comparability of Levels or Units	... for each i, $u^*_i = \phi_i(u_i)$, where $\phi_1, \phi_2, ..., \phi_n : \mathbf{R} \to \mathbf{R}$ are positive linear transformations	not invariant	not invariant	invariant
(OLC): Ordinal Measurability, Interpersonal Comparability of Levels	... for each i, $u^*_i = \phi(u_i)$, where $\phi : \mathbf{R} \to \mathbf{R}$ is a positive monotonic transformation	invariant	not invariant	not invariant
(CUC): Cardinal Measurability, Interpersonal Comparability of Units	... for each i, $u^*_i = a_i+b*u_i$, where $a_1, a_2, ..., a_n \in \mathbf{R}$ and $b\in\mathbf{R}$ with $b>0$	not invariant	invariant	not invariant
(CFC): Cardinal Measurability, Interpersonal Comparability of Levels and Units	... for each i, $u^*_i = \phi(u_i)$, where $\phi: \mathbf{R} \to \mathbf{R}$ is a positive affine transformation	invariant	invariant	not invariant
(RFC): Ratio-Scale Measurability, Interpersonal Comparability of Levels and Units	... for each i, $u^*_i = \phi(u_i)$, where $\phi : \mathbf{R} \to \mathbf{R}$ is a positive linear transformation	invariant	invariant	invariant

The first column states the name of the measurability and interpersonal comparability condition. The second column gives a definition of that condition in terms of the conditions under which two profiles of utility functions are considered to be informationally equivalent. The third, fourth and fifth column state whether, under the given measurability and interpersonal comparability condition, (LC)-, (UC)-, and (ZC)-statements are invariant under all admissible transformations of the profiles.

Table 1

In section 4, I will briefly review the implications of interpersonal comparability of utility for Arrow's impossibility theorem (for a detailed discussion of Arrow’s theorem and the implications of interpersonal comparability of utility, see Bossert and Weymark, 1996). The main argument of the paper can still be followed if section 4 is skipped.

4. Interpersonal Comparisons and Arrow's Impossibility Theorem

A *social welfare functional* (SWFL) is a function F that maps each profile $\{u_i\}_{i\in N}$ in a given domain to a collective preference ordering R on the set X, where R is reflexive, connected and transitive. xRy is interpreted to mean "x is collectively at least as good as y". R induces a strong ordering on X, defined as follows: for all $x, y\in X$, xPy if and only if xRy and not yRx. Moreover, it is required that F should map informationally equivalent profiles to the same collective preference ordering, formally we have:

INVARIANCE ASSUMPTION (INV). For any $\{u_i\}_{i\in N}$ and $\{u^*_i\}_{i\in N}$ in the domain of F, if $\{u_i\}_{i\in N}$ and $\{u^*_i\}_{i\in N}$ are informationally equivalent, then $F(\{u_i\}_{i\in N}) = F(\{u^*_i\}_{i\in N})$.

The invariance assumption must always be stated with respect to a specific definition of informational equivalence, as given by the measurability and interpersonal comparability assumptions in table 1. The problem addressed by Arrow's theorem is whether there exist SWFLs, F, that satisfy some minimal conditions. Arrow’s conditions are the following:

UNIVERSAL DOMAIN (U). The domain of F is the set of all logically possible profiles of utility functions.

Condition (U) requires that any logically possible profile of utility functions be admissible as input to the aggregation.

WEAK PARETO PRINCIPLE (P). Let $\{u_i\}_{i\in N}$ be any profile in the domain of F, and let $R = F(\{u_i\}_{i\in N})$. For any $x_1, x_2\in X$, we have x_1Px_2 whenever, for all $i\in N$, $u_i(x_1)>u_i(x_2)$.

Condition (P) requires that, if *all* individuals have a greater utility under x_1 than under x_2, then x_1 should be collectively preferred to x_2.

INDEPENDENCE OF IRRELEVANT ALTERNATIVES (I). Let $\{u_i\}_{i\in N}$ and $\{u^*_i\}_{i\in N}$ be any profiles in the domain of F, and let $R = F(\{u_i\}_{i\in N})$ and $R^* = F(\{u^*_i\}_{i\in N})$. For any $x_1, x_2\in X$, if, for all $i\in N$, $u_i(x_1) = u^*_i(x_1)$ and $u_i(x_2) = u^*_i(x_2)$, x_1Rx_2 if and only if $x_1R^*x_2$.

Condition (I) requires that the collective ranking of any pair of alternatives should depend exclusively on the values of the individual utility functions for that pair of alternatives.

NON-DICTATORSHIP (D). F is not dictatorial: there does not exist an $i\in N$ such that, for all $\{u_i\}_{i\in N}$ in the domain of F and any $x_1, x_2\in X$, $u_i(x_1)>u_i(x_2)$ implies x_1Px_2, where $R = F(\{u_i\}_{i\in N})$.

Condition (D) requires that there should not exist one individual, a *dictator*, whose utility function always (except possibly in cases of indifference) determines the collective preference.

Arrow's impossibility theorem states that, given (ONC), there exists no SWFL satisfying these four conditions simultaneously (Arrow, 1951/1963; Sen 1970/1979):

Theorem 1. There exists no SWFL F satisfying [(INV) with respect to (ONC)] and (U), (P), (I), (D).

It is also known that, for suitable other measurability and interpersonal comparability conditions, there exist SWFLs satisfying (U), (P), (I) and (D). Table 2 shows the logical implications of the conditions listed in table 1 for the existence or non-existence of SWFLs satisfying Arrow's minimal conditions (see Sen, 1970/1979, d'Aspremont, 1985; and List, 2001, on the condition (ONC+0)).

Condition:	Which types of statements are invariant under the class of transformations up to which a profile of utility functions is unique?			Do there exist SWFLs satisfying (U), (P), (I) and (D)?
	(LC)-statements	(UC)-statements	(ZC)-statements	
(ONC)	no	no	no	no
(ONC+0)	no	no	yes	yes
(CNC)	no	no	no	no
(RNC)	no	no	yes	yes
(OLC)	yes	no	no	yes
(CUC)	no	yes	no	yes
(CFC)	yes	yes	no	yes
(RFC)	yes	yes	yes	yes

Table 2

We observe that, given a choice between the alternative conditions listed in table 1, there exist SWFLs satisfying Arrow's minimal conditions *if and only if* at least one of the three identified types of interpersonal comparisons ((LC)-, (UC)- or (ZC)-statements) are meaningful, i.e. invariant under the class of transformations up to which a profile of utility functions is unique. Viewed in this light, the meaningfulness of interpersonal comparisons of utility is a necessary and sufficient condition for the existence of aggregation procedures satisfying all of Arrow's minimal conditions simultaneously.

5. The Parallel between Translation of Meaning and Interpersonal Comparisons of Utility

To identify a parallel between the problem of interpersonal comparisons of utility and the problem of translation of meaning, as explored by Quine (1960), it is useful to recall the characterization of utility given in section 1 and to compare it with Quine's conception of meaning.[10] Utility and meaning differ with respect to property (i): utility does, whereas meaning does not, capture some form of *welfare*. But with respect to property (ii), both utility and meaning are something we *attribute* to a person. With respect to property (iii), both utility and meaning are

[10] The present argument does not depend on the defensibility of Quine's theory of language and its underlying assumptions. Much of modern linguistics has departed from Quine's account (e.g. Chomsky, 1969). Rather, given the parallels between the *structure* of Quine's conception of meaning and the *structure* of the conceptions of utility addressed here, the present argument seeks to draw on Quine's insights on what the *implications of this structure* are. This is independent from the question of whether an account of language or utility based on this structure is defensible.

something that can be *experienced* (if at all) only from a first-person perspective, although Quine himself, as a behaviourist, might be reluctant to speak of *experiencing* meaning.[11] Finally, with regard to property (iv), both utility and meaning *surface observably* in the form of certain behavioural and/or other observable proxies.

Translation involves attributing linguistic meanings to different speakers (property (ii)). Suppose I observe that a speaker of a different language assents to the observation sentence "Gavagai!" in precisely the same empirical conditions in which I assent to the English observation sentence "Rabbit!". Or suppose I observe that another speaker of *English* assents to the sentence "Rabbit!" in precisely the same empirical conditions in which I assent to this sentence. Then I am inclined to infer that the sentences "Gavagai!" for the foreign language speaker, "Rabbit!" for the *other* English speaker, and "Rabbit!" for me all have the same meaning. On Quine's account, our sole basis for making such judgments of interpersonal sameness of meaning lies in our empirical observations of people's linguistic behaviour (property (iv)), since we have no introspective first-person access to other persons' minds (property (iii)). According to Quine's indeterminacy of translation thesis, even the totality of empirical evidence about a person's linguistic behaviour underdetermines the attribution of meanings to that person. Given suitable adjustments in the translation of other sentences, the rival hypotheses that "Gavagai!" for the foreign language speaker (or "Rabbit!" for the other English speaker) means "Undetached rabbit part!" or "Temporal rabbit stage!" rather than "Rabbit!" are equally compatible with our empirical observations of the foreign speaker's (or the other English speaker's) linguistic behaviour. Which translation of “Gavagai!” we adopt has potentially far-reaching repercussions for the translation of more theoretical sentences.[12] Although the non-standard translations seem less parsimonious *from the perspective of our own English language*, there is, on Quine's account, not even in principle any evidence that would break the underdetermination between different such rival translations. And since Quine holds that positing *in principle inaccessible* facts of the

[11] It should be emphasized that Quine’s position is quite radical. While the orthodox account of utility denies the existence of direct *third-person* access to a subject’s utilities, the account still assumes that the subject him- or herself has *first-person* access to his or her own utilities. Quine’s position, in its radical form, is not only that there is no direct *third-person* access to what a speaker means, but also that not even the speaker him- or herself has *first-person* access to his or her own meanings.

[12] Strictly speaking (and as Quine himself points out), the "Gavagai!"-example by itself illustrates only indeterminacy of reference, not indeterminacy of translation. Indeterminacy of translation requires that there exist sentences which can be adequately translated not only in two or more *different* ways (like the sentence "Gavagai!"), but also in *logically incompatible* ways (unlike the sentence "Gavagai!", whose rival translations are different, but *not* logically incompatible – in Quine's terms, they are holophrastically equivalent). The arguments of the present paper, however, are not dependent on the indeterminacy of translation thesis. The present analysis of the problem of interpersonal comparisons of utility can equally be developed on the basis of a parallel between the problem of attributing utility and the problem of attributing reference.

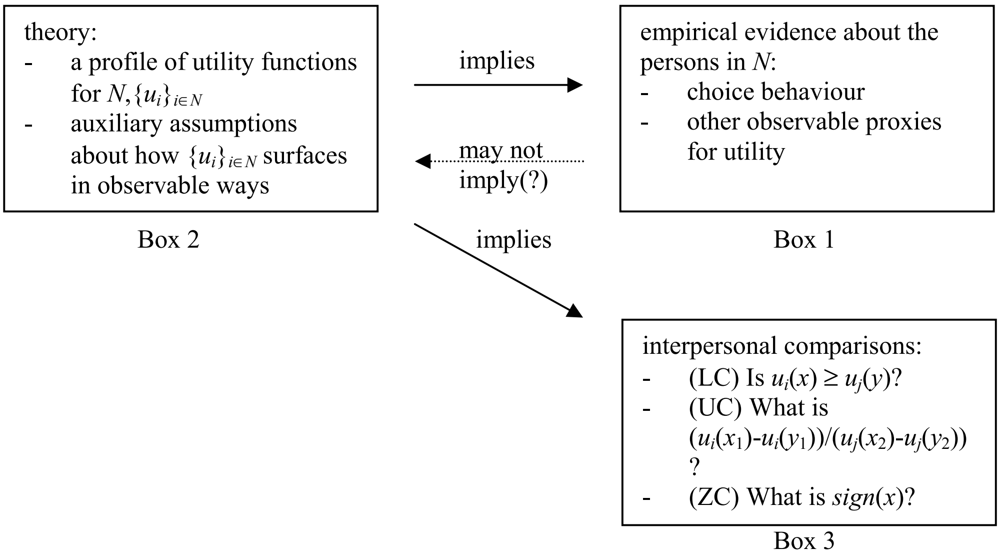
matter is methodologically unacceptable, he concludes that translation and, more generally, judgments of interpersonal sameness of meaning are indeterminate.

Similarly, the utilities experienced by another person are not directly observable by us (property (iii)). We can only observe the behaviour of the person (property (iv)), including their choice behaviour and their verbal expressions, and possibly other physiological proxies for utility. These proxies might range from the person's facial expressions on the folk-psychological side to a measurement of their neural activity on the high-tech-psychological side. Like the attribution of meanings to a speaker on the basis of the speaker's linguistic behaviour, the attribution of utility to a person (property (ii)) involves theorizing on the basis of whatever evidence about this person is accessible from an external third-person perspective. Therefore, in making interpersonal comparisons of utility, we must rely on whatever attribution of utilities to the relevant persons most adequately covers the available empirical evidence, given certain background assumptions about how utility surfaces in observable ways. On the orthodox account, even the totality of such evidence underdetermines interpersonal comparisons of utility: "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 difference. Every mind is thus inscrutable to every other mind, and no common denominator of feeling seems to be possible ... the motive in one mind is weighed only against other motives in the same mind, never against the motives in other minds" (Jevons, 1911, p. 14). As in the case of translation, if all relevant observable behaviour and, possibly, other relevant observable physiological responses of two persons are identical, then we are inclined to attribute identical utilities to these persons. Call this the standard hypothesis. But consider the rival hypothesis that one of the two persons, the 'utility monster', is one thousand times more susceptible to pleasure and pain than the other. If Jevons's argument is correct, this rival hypothesis, while apparently less parsimonious than the standard hypothesis, is equally compatible with all available empirical evidence. *If* we believe that Jevons's argument is correct and that we have not left out any relevant empirical evidence, we are forced to conclude that interpersonal comparisons of utility are underdetermined by the totality of empirical evidence. *If* we believe in addition that there is no independent fact of the matter to break the underdetermination, we are forced to conclude that interpersonal comparisons of utility are indeterminate.

Such, in short, is the parallel between the problem of translation of meaning and the problem of interpersonal comparisons of utility. I will now turn to a more detailed discussion of the latter.

6. Attributing Utility to Persons on the Basis of Empirical Evidence

As noted, the utilities experienced by the persons in *N* under the options in *X* cannot be directly observed. Like the attribution of meanings to a set of speakers, the attribution of utilities to a set of persons involves building a theory on the basis of the available evidence. This theory consists of a profile of utility functions $\{u_i\}_{i\in N}$ and some auxiliary assumptions about how utility surfaces in observable ways. A typical, but not uncontroversial, such assumption is that, if $u_i(x)>u_i(y)$, then person *i* would, in normal circumstances, choose *x* over *y*. It is only *after* attributing a profile of utility functions to the persons in *N* and specifying relevant auxiliary assumptions that we can make interpersonal comparisons of utility. Thus making such comparisons is a two-step process. In a first step, we attribute to the persons in *N* a profile of utility functions $\{u_i\}_{i\in N}$ such that $\{u_i\}_{i\in N}$ (jointly with the relevant auxiliary assumptions) is adequate with respect to the available evidence. In a second step, we use the attributed profile of utility functions to make interpersonal comparisons of utility.[13] Schematically, the logical relation between empirical evidence (box 1), an attributed profile of utility functions (box 2) and interpersonal comparisons (box 3) is as shown in table 3:



theory:
- a profile of utility functions for $N, \{u_i\}_{i\in N}$
- auxiliary assumptions about how $\{u_i\}_{i\in N}$ surfaces in observable ways

Box 2

empirical evidence about the persons in *N*:
- choice behaviour
- other observable proxies for utility

Box 1

interpersonal comparisons:
- (LC) Is $u_i(x) \geq u_j(y)$?
- (UC) What is $(u_i(x_1)\text{-}u_i(y_1))/(u_j(x_2)\text{-}u_j(y_2))$?
- (ZC) What is *sign*(*x*)?

Box 3

Table 3

[13] The two-step model is an idealization. The main aim of that model is to emphasize that making interpersonal comparisons involves the attribution of a profile of utility functions to the set of persons. As an anonymous reviewer has pointed out, in real life, the two-step process may not always take place in the way suggested by the model; in particular, the direction of the two-step process may sometimes even be reversed: our attribution of a profile of utility functions to a set of persons may sometimes be informed by the interpersonal comparisons we are inclined to make, not the other way round. For example, we may sometimes attribute intense preferences to a person not on the basis of this person's choices, but because we have found such intense preferences in others (as judged by *their* choices) and because we assume different persons to be similar in their psychology.

The relation between box 2 and box 3 is one of logical implication: Given a profile of utility functions, we can make (LC)-, (UC)- and (ZC)-statements *relative to that profile*. This means that, if we can be sure that we have filled box 2 'correctly', i.e. that we have found the 'correct' profile of utility functions, we will have found a basis for interpersonal comparisons. Or, to be more precise, we will have found such a basis if we can be sure that the content of box 2 is unique up to a sufficiently small class of transformations for (LC)-, (UC)- or (ZC)-statements to be invariant under these transformations. The central question we have to address is therefore whether the empirical evidence in box 1 determines the theory in box 2 sufficiently uniquely.

The onus of argument on the proponent of the *empirical meaningfulness* of interpersonal comparisons of utility is to show that the empirical evidence in box 1 determines a profile of utility functions in box 2 uniquely up to a sufficiently small class of transformations. The onus of argument on the proponent of the *possibility* of interpersonal comparisons of utility is slightly weaker: it is to show that, even if there is no straightforward relation of logical implication leading from box 1 to box 2, there are other, *possibly non-empirical*, considerations over and above the evidence in box 1 which enable us to select a profile of utility functions in box 2 uniquely up to a sufficiently small class of transformations.

The rest of section 6 is mainly concerned with the former question about empirical meaningfulness, section 7 mainly with the latter one about possibility. In subsection 6.1, I will introduce several different cases of what the relevant empirical evidence might be. In subsection 6.2, I will then explore the implications of the various cases.

6.1. Different Types of Empirical Evidence

I will now present several cases of what the empirical evidence for utility might be. Each case represents an idealized limiting case, positing a body of evidence that is richer than what we realistically expect to find empirically. This is not harmful for the present argument. If there are underdetermination problems even in a utopia of unrealistically rich empirical evidence, then, *a fortiori*, these problems will occur in more realistic circumstances of sparse evidence. Whether any of the discussed types of evidence are really evidence for utility is a philosophical question this paper cannot resolve.[14] The formal conditions stated in table 4 will be discussed more

[14] As indicated above, interpreting a body of empirical observations as evidence for utility requires certain auxiliary assumptions about how utility surfaces in observable ways. Amongst these auxiliary assumptions are relatively common ones such as the assumption (mentioned above) that, if $u_i(x)>u_i(y)$, then person *i* would, under normal circumstances, choose *x* over *y*, as well as more contestable ones such as condition (N1 a/b/c) introduced below. Whether or not commonly made such assumptions are defensible is left open here. These open questions, however, reinforce rather than weaken the central point of the paper, namely that attributing utilities to people on the basis of

informally below. The argument can be informally followed even if the technical details of table 4 are skipped.

Ranking Evidence – Options (RankEv).
The evidence includes all true statements of the form
- xR_iy, where $x, y \in X$, $i \in N$ and xR_iy means "person i weakly prefers x to y"[15], satisfying
(P1) ("ordering") for each $i \in N$, R_i is a reflexive, transitive and complete binary relation.

A *binary lottery* is an option of the form $p{*}x{+}(1{-}p){*}y$, where $x, y \in X$, $p \in [0,1]$. $p{*}x{+}(1{-}p){*}y$ means "with probability p, get x; with probability $1{-}p$, get y". Given a set of options X, let $L(X)$ be the set of all binary lotteries in X. Note that $X \subseteq L(X)$, since each $x \in X$ can be interpreted as a binary lottery $1{*}x{+}0{*}y \in L(X)$ (where $y \neq x$). For each $i \in N$, R_i induces a strong ordering P_i, defined as follows: xP_iy if and only if xR_iy and not yR_ix.

Ranking Evidence – Options and Binary Lotteries (RankEvLot).
The evidence includes all true statements of the form
- xR_iy, where $x, y \in L(X)$, $i \in N$ and xR_iy means "person i weakly prefers x to y", satisfying
(P1) ("ordering") for each $i \in N$, R_i is a reflexive, transitive and complete binary relation on $L(X)$;
(P2) ("Archimedean property") for each $i \in N$ and all $x, y, z \in X$, if xP_iy and yP_iz, then there exist $\lambda, \mu \in (0,1)$ such that $(\lambda{*}x{+}(1{-}\lambda){*}z)P_iy$ and $yP_i(\mu{*}x{+}(1{-}\mu){*}z)$;
(P3) ("independence") for each $i \in N$, all $x, y, z \in X$ and all $\lambda \in (0,1]$, xR_iy if and only if $(\lambda{*}x{+}(1{-}\lambda){*}z)R_i(\lambda{*}y{+}(1{-}\lambda){*}z)$.

Additional Proxies for Utility – Case a (Prox-a).
The evidence includes all true statements of the form
- $f_i(x) = \lambda$, where $x \in X$, $i \in N$, $\lambda \in \mathbf{R}$ and $f_i : X \rightarrow \mathbf{R}$ is some real-valued observable proxy for person i's utility in response to options in X, satisfying (given that we also have (RankEv))
(P4 a) ("consistency of f-response with preference") there exist a positive monotonic transformation $\phi : \mathbf{R} \rightarrow \mathbf{R}$ and *some* profile of utility functions $\{u^*_i\}_{i \in N}$ representing $\{R_i\}_{i \in N}$ (according to theorem 2 below) such that, for all $i \in N$ and all $x, y \in X$, $f_i(x) = \phi(u^*_i(x))$.

Additional Proxies for Utility – Case b (Prox-b).
The evidence includes all true statements of the form
- $g_i(x, y) = \lambda$, where $x, y \in X$, $i \in N$, $\lambda \in \mathbf{R}$ and $g_i : X^2 \rightarrow \mathbf{R}$ is some real-valued observable proxy for person i's utility in response to differences/switches between options in X, satisfying (given that we also have (RankEvLot))
(P4 b) ("consistency of g-response with preference") there exist a positive monotonic transformation $\psi : \mathbf{R} \rightarrow \mathbf{R}$ and *some* profile of utility functions $\{u^*_i\}_{i \in N}$ representing $\{R_i\}_{i \in N}$ (according to theorem 3 below) such that, for all $i \in N$ and all $x, y \in X$, $g_i(x, y) = \psi(u^*_i(x) - u^*_i(y))$.

Additional Proxies for Utility – Case c (Prox-c).
The evidence includes all true statements of the form
- $h_i(x) = \delta$, where $x \in X$, $i \in N$, $\delta \in \{-1,0,1\}$ and $f_i : X \rightarrow \mathbf{R}$ is some -1/0/1-valued observable proxy for person i's utility in response to options in X (taking values 1='positive utility', 0='zero utility', -1='negative utility'), satisfying (given that we also have (RankEv))
(P4 c) ("consistency of h-response with preference") there exists *some* profile of utility functions $\{u^*_i\}_{i \in N}$ representing $\{R_i\}_{i \in N}$ (according to theorem 2 below) such that, for all $i \in N$ and all $x, y \in X$, $h_i(x) = sign(u^*_i(x))$.

Table 4

empirical evidence involves a substantial act of theorizing that may suffer from underdetermination and possibly indeterminacy problems.

[15] This definition of the evidence entails that, for any $x, y \in X$ and any $i \in N$, the negation of xR_iy is true *if and only if* xR_iy is *not* included in the evidence.

In terms of the conditions stated in table 4, we consider the following cases.

Case 1: We have (RankEv): Each person's utility surfaces only in the form of the revealed preference ordering R_i over the options in X. This means, given an apple, an orange and a banana, we can determine each person's preference ordering over these three fruits.

Case 2: We have (RankEvLot): Each person's utility surfaces only in the form of the revealed preference ordering R_i over the options and binary lotteries in X. This means we can determine not only whether a person prefers an orange to a banana to an apple, but also whether, for any given probability p, the person prefers a guaranteed banana to a lottery whose prize would be *either* an orange *or* an apple with associated probabilities p and $1-p$, respectively.

Case 3a: We have (RankEv) and (Prox-a);

Case 3b: We have (RankEvLot) and (Prox-b);

Case 3c: We have (RankEv) and (Prox-c):

("utopian best case scenarios") Each person's utility surfaces in the form of the revealed preference ordering R_i (over the options – and, in case 3b, binary lotteries – in X) and some other observable proxies for utility, formalized here by the functions f_i, g_i or h_i.[16] These proxy functions could measure such characteristics as a person's observable facial expression of pleasure or pain, a person's verbal expressions, a person's heartbeat or body temperature, a person's relevant neural activity, in response to the options[17] or in response to switches between options[18]. Or they could measure a person's spontaneity of choosing one option over another in a forced-choice situation[19] (see Waldner, 1972), assuming that a greater such spontaneity corresponds to a greater utility gain. Or they could measure something else that might be thought of as a proxy for a person's utility. I am here making no claims as to whether such additional observable proxies for utility exist. The point is only to identify the logical implications of such a utopian best case scenario for the problem of interpersonal comparisons of utility.

[16] At first sight, the proxy functions f_i, g_i and h_i (particularly f_i and g_i) may raise similar problems of measurability and uniqueness as the utility functions u_i themselves. But even if there is no unique privileged scale for measuring f_i and g_i, we will assume that what makes f_i and g_i *observable* is that, whatever scale of measurement we choose, this scale is a *common* one for all persons in N. It is thus crucial that f_i and g_i are unique up to identical transformations (say positive affine ones) for every person. The proxies are to be interpreted, using Elster and Roemer's phrase (1991, introduction, p. 10), as "objective proxies for subjective well-being", not as suggesting an "objective conception of well-being".

[17] In case 3a, real-valued; in case 3c, -1/0/1-valued.

[18] In case 3b, real-valued.

[19] Also in case 3b.

6.2. Implications

To determine the implications of cases 1, 2 and 3a/b/c for the problem of interpersonal comparability, we will use two standard representation theorems. The argument can be informally followed even if the technical details of the theorems are skipped.

Theorem 2. (Debreu, 1954) For each $i \in N$, the following holds: Given that X is finite or denumerable, R_i satisfies (P1) if and only if there exists a utility function $u_i : X \rightarrow \mathbf{R}$ such that, for all x, $y \in X$, xR_iy if and only if $u_i(x) \geq u_i(y)$. Moreover, if u_i has this property, then so does $\phi_i(u_i)$, where $\phi_i : \mathbf{R} \rightarrow \mathbf{R}$ is any positive monotonic transformation.

Theorem 2 states that any preference ordering satisfying condition (P1) in table 4 can be represented by a utility function that is unique up to positive monotonic transformations.

Theorem 3. (von Neumann and Morgenstern, 1944) For each $i \in N$, the following holds: R_i satisfies (P1), (P2) and (P3) if and only if there exists a utility function $u_i : L(X) \rightarrow \mathbf{R}$ such that (i) for all x, $y \in X$, xR_iy if and only if $u_i(x) \geq u_i(y)$ and (ii) for all x, $y \in X$ and all $p \in [0,1]$, $u_i(p^*x+(1-p)^*y)= p^*u_i(x)+(1-p)^*u_i(y)$. Moreover, if u_i has this property, then so does $\phi_i(u_i)$, where $\phi_i : \mathbf{R} \rightarrow \mathbf{R}$ is any positive affine transformation.

Theorem 3 states that any preference ordering satisfying conditions (P1), (P2) and (P3) in table 4 can be represented by a utility function that is unique up to positive affine transformations.

6.2.1. Using Only Ranking Evidence

Cases 1 and 2. In case 1, by theorem 2, each person's utility function is determined uniquely only up to positive monotonic transformations. In case 2, by theorem 3, each person's utility function is determined uniquely only up to positive affine transformations. Thus cases 1 and 2 generate, respectively, conditions (ONC) and (CNC) in table 1, and therefore leave (UC)-, (LC)- and (ZC)-statements underdetermined.

If we hold in addition that there is no independent fact of the matter about what the 'true' interpersonal comparisons of utility are, interpersonal comparisons of utility are indeterminate.

6.2.2. Using Additional Proxies for Utility

The situation changes if we use the additional evidence (Prox-a), (Prox-b) or (Prox-c). This means that we can use not only the persons' revealed preferences, but also the other observable proxies for utility as a potential basis for interpersonal comparisons. The conditions (P4 a/b/c) have two implications. First, a utility function we attribute to a person on the basis of revealed preferences is consistent with what the other proxy functions, f_i, g_i or h_i would lead us to infer about this utility function: f_i strictly increases with an increase in utility; g_i strictly increases with an increase in the utility gain a person experiences as a result of a switch from one option to another; h_i is weakly monotonic in utility. In particular, we can formalize the functional relation between each person's utility function u_i and the proxy functions f_i, g_i or h_i in terms of a suitable transformation, where the transformation describes how utility surfaces in the form of the proxy functions f_i, g_i or h_i.[20] Second, it is possible to attribute a profile of utility functions to the persons in such a way that these transformations are the *same* for all persons (namely by setting $\{u_i\}_{i\in N} :=$ $\{u^*_i\}_{i\in N}$, where $\{u^*_i\}_{i\in N}$ is as in conditions (P4 a/b/c)).

Now much depends on whether or not we accept the following (non-empirical) conditions:

(N1 a) ("interpersonal sameness of the conversion of utility into the proxy functions")
There exists a positive monotonic transformation $\phi: \mathbf{R} \rightarrow \mathbf{R}$ such that, for all $i\in N$ and all $x, y\in X, f_i(x) = \phi(u_i(x))$.

(N1 b) ("interpersonal sameness of the conversion of utility into the proxy functions")
There exists a positive monotonic transformation $\psi: \mathbf{R} \rightarrow \mathbf{R}$ such that, for all $i\in N$ and all $x, y\in X$, $g_i(x, y) = \psi(u_i(x)\text{-}u_i(y))$.

(N1 c) ("interpersonal sameness of the conversion of utility into the proxy functions")
For all $i\in N$ and all $x\in X$, $h_i(x) = sign(u_i(x))$.

Conditions (N1 a/b/c) state that a profile of utility functions is adequate *only if*, according to that profile, all persons have identical transformations of utility into the observable proxies f_i, g_i or h_i. In cases 3a/b, conditions (N1 a/b) rule out the possibility that different persons exhibit identical f_i or g_i values and yet their underlying utilities are different. In case 3c, condition (N1 c) rules out

[20] In cases 3a and 3b, the transformation is a positive monotonic transformation; in case 3c, the transformation is the sign-function.

the possibility that different persons exhibit identical h_i values and yet they are not in the same one of the three states 'positive utility', 'zero utility', 'negative utility'. We will now see that cases 3a, 3b and 3c, jointly with conditions (N1 a), (N1 b), (N1 c), generate conditions (OLC), (CUC) and (ONC+0) in table 1, respectively.

Case 3a with (Prox-a). If we accept condition (N1 a), we are no longer free to apply *different* positive monotonic transformations to the utility functions of different persons without undermining the adequacy of the resulting profile. Suppose we apply a positive monotonic transformation to one person's utility function, i.e. for *some* $i \in N$, we replace u_i with $\theta(u_i)$, where $\theta: \mathbf{R} \to \mathbf{R}$ is a positive monotonic transformation. Then we are also forced to replace ϕ with ϕ^*, where ϕ is the transformation in (N1 a), defining ϕ^* as follows: for all $t \in \mathbf{R}$, $\phi^*(t) = \phi(\theta^{-1}(t))$ (θ^{-1} is the inverse transformation of θ). Consequently, we are forced to replace u_i with $\theta(u_i)$ for *every* $i \in N$. Hence a profile of utility functions is determined uniquely up to *identical* positive monotonic transformations *for every person*. This is condition (OLC) in table 1, determining (LC)-statements.

Case 3b with (Prox-b). If we accept condition (N1 b), the situation is similar. Suppose we apply a positive affine transformation to one person's utility function, i.e. for *some* $i \in N$, we replace u_i with $a_i + bu_i$. Then we are also forced to replace ψ with ψ^*, where ψ is the transformation in (N1 b), defining ψ^* as follows: for all $t \in \mathbf{R}$, $\psi^*(t) = \psi(t/b)$. Consequently, we are forced to replace u_i with $a_i + bu_i$ for *every* $i \in N$. Note that, while the a_i may be different for different persons i, b must be the same for all persons. This is condition (CUC) in table 1, determining (UC)-statements.

Case 3c with (Prox-c). If we accept condition (N1 c), the only positive monotonic transformations we can apply to each u_i without undermining the adequacy of the resulting profile are sign-preserving ones. The reason is that, if we replace u_i with $\theta(u_i)$, where θ is *not* sign-preserving, then it may no longer be true that, for all $x \in X$, $h_i(x) = sign(\theta(u_i(x)))$ as required by condition (N1 c). This is condition (ONC+0) in table 1, determining (ZC)-statements.

If we do not accept conditions (N1 a/b/c), on the other hand, we are free to apply *different* transformations to the utility functions of different persons. In cases (3 a/b), we then have to admit the possibility that different persons exhibit identical f_i or g_i values and yet their underlying utilities are different. In case (3 c), we have to admit the possibility that different persons exhibit identical h_i values and yet they are not in the same one of the three states 'positive utility', 'zero utility', 'negative utility'. For instance, if we multiply person 1's utility function by a factor of 10 while leaving all other utility functions unchanged, we must also accept that person 1's 'rate' of converting utility into observable f_i or g_i values is divided by a factor of 10. If we are prepared to make such adjustments (and an opponent of interpersonal comparisons of utility would indeed ask, why not?), we are back to the conditions (ONC) or (CNC) in table 1. (UC)-, (LC)- and (ZC)-statements then remain underdetermined. Again, if we hold that there is no independent fact of the matter about what the 'true' interpersonal comparisons of utility are, interpersonal comparisons of utility are indeterminate.

6.2.3. Interpretation

There are at least three different views one might take on the status of conditions (N1 a/b/c). On the first (realist) view, conditions (N1 a/b/c) are held to be true in a realist sense: the functional relation between the *real* utilities experienced by the persons and the observable proxies f_i, g_i or h_i is the same for all persons. One possible source of this realist position might be the view that utility is systematically reducible to, or in an interpersonally identical way correlated with, certain observable physiological states. On such a view, identical physiological states of the relevant kind – expressed in terms of the proxy functions f_i, g_i or h_i – indicate identical utilities.[21]

On the second (pragmatic) view, which Davidson (1986) attributes to Harsanyi (1955) and Waldner (1972), conditions (N1 a/b/c) are regarded not as stating a truth about *real* utilities, but as a requirement of good scientific methodology: in the absence of any observable differences between different persons, it is bad methodology to attribute different utilities[22] to them; good methodology requires us to attribute identical utilities[23] if the observable proxies are identical. Davidson summarizes this view as follows: "[I]t does not make sense to say that two people are alike in all relevant observable respects but have different thoughts and feelings. Or perhaps it makes sense, but it is bad science." Harsanyi offers the following defence: "If two objects or 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 a completely gratuitous hypothesis and one contrary to sound scientific method. ... Thus in the case of persons with similar preferences and expressive reactions we are fully entitled to assume that they derive the same utilities from similar situations." (Harsanyi, 1955, p. 279) We will return to this pragmatic view in section 7.2.

On the third (sceptical) view, instead of "not postulating any differences unless there is some reason to do so" (Waldner, 1972, p. 102), it is held that "there is no scientific reason to postulate anything at all" (Davidson, 1986, p. 202), and conditions (N1 a/b/c) are therefore rejected.

[21] Or, in the case of h_i, identical interpersonally significant states ‘positive utility’, ‘zero utility’ and ‘negative utility’.

[22] Or different interpersonally significant states 'pleasure', 'zero utility', 'pain'.

[23] Or identical states 'pleasure', 'zero utility', 'pain'.

6.2.4. Summary

Table 5 shows the logical relation between the different types of evidence introduced above, conditions (N1 a/b/c) and the conditions listed in table 1.

If we have								Then we have
(RankEv)	(RankEv Lot)	(Prox-a)	(Prox-b)	(Prox-c)	(N1 a)	(N1 b)	(N1 c)	
yes	no	no or yes	no	no or yes	no	no	no	(ONC)
yes	no	no or yes	no	yes	no	no	yes	(ONC+0)
yes	yes	no or yes	no or yes	no or yes	no	no	no	(CNC)
yes	yes	no or yes	no or yes	yes	no	no	yes	(RNC)
yes	no	yes	no	no or yes	yes	no	no	(OLC)
yes	yes	no or yes	yes	no or yes	no	yes	no	(CUC)
yes	yes	yes	no or yes	no or yes	yes	no or yes	no	(CFC)
yes	yes	yes	no or yes	yes	yes	no or yes	yes	(RFC)

Table 5

Comparing tables 2 and 5, we observe that the conditions which are sufficient for the existence of aggregation procedures satisfying all of Arrow's conditions simultaneously are precisely the ones in which at least one of conditions (N1 a/b/c) is accepted. This highlights the significance of conditions (N1 a/b/c) not only for the question of whether interpersonal comparisons of utility are determined by the available empirical evidence, but also for the solubility of Arrowian collective decision problems.

7. Yet Another Impossibility Argument?

Schematically, the argument of the present paper can be summarized as follows:

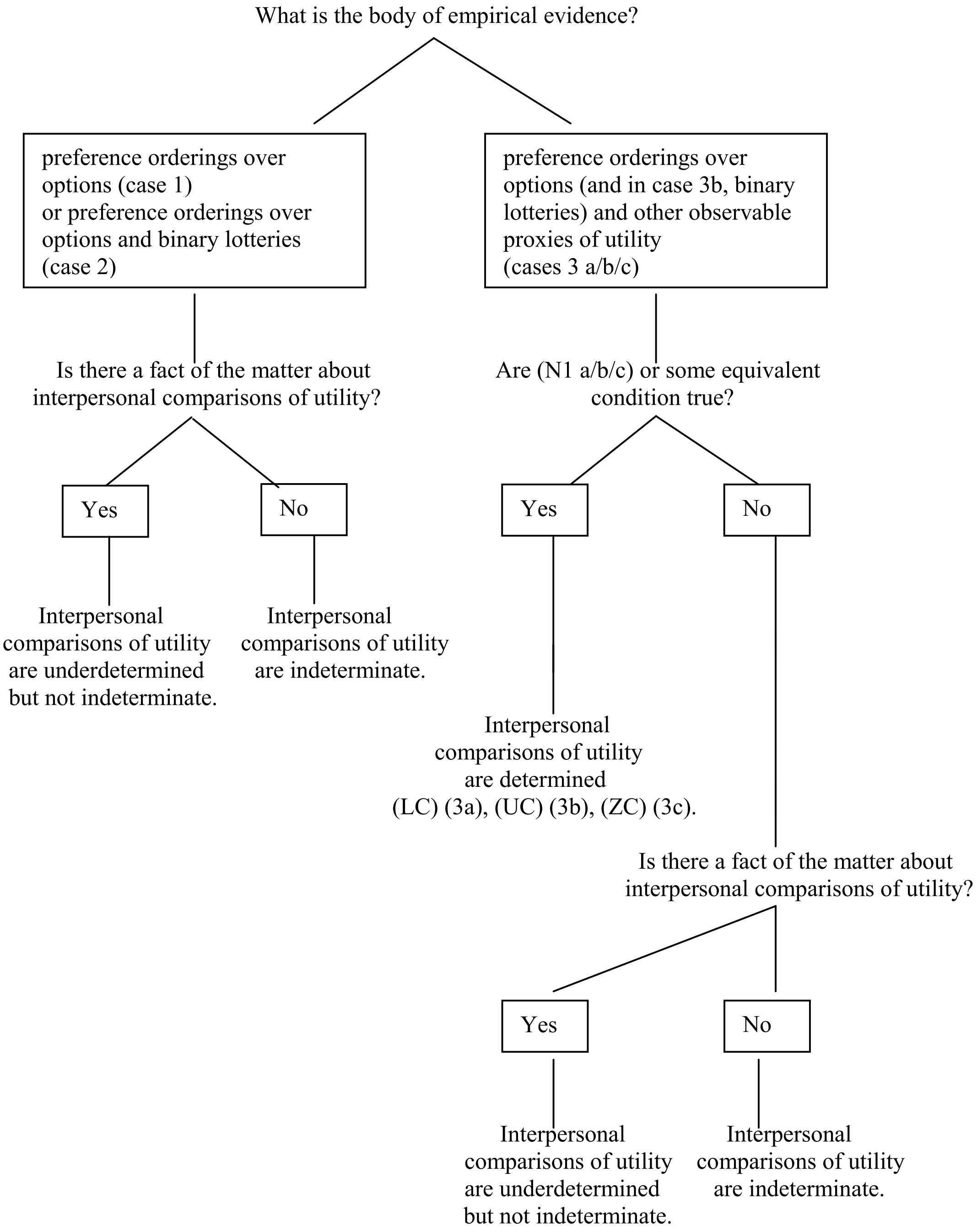

What is the body of empirical evidence?
preference orderings over options (case 1) or preference orderings over options and binary lotteries (case 2)
preference orderings over options (and in case 3b, binary lotteries) and other observable proxies of utility (cases 3 a/b/c)
Is there a fact of the matter about interpersonal comparisons of utility?
Yes
No
Interpersonal comparisons of utility are underdetermined but not indeterminate.
Interpersonal comparisons of utility are indeterminate.
Are (N1 a/b/c) or some equivalent condition true?
Yes
No
Interpersonal comparisons of utility are determined (LC) (3a), (UC) (3b), (ZC) (3c).
Is there a fact of the matter about interpersonal comparisons of utility?
Yes
No
Interpersonal comparisons of utility are underdetermined but not indeterminate.
Interpersonal comparisons of utility are indeterminate.

Table 6

In short, unless we have the rich evidence of cases 3a, 3b or 3c (or, to be more precise, a sufficiently large subset of such evidence) *and* we accept at least one of the corresponding conditions (N1 a/b/c) (or some equivalent condition) as true, interpersonal comparisons of utility are underdetermined and, if we also believe that there is no independent fact of the matter about what the 'true' interpersonal comparisons of utility are, indeterminate.

Is this yet another version of the argument that interpersonal comparisons of utility are impossible? Does the present argument once again make a mystery of the apparent ease with which we make (what look like) interpersonal comparisons of utility?

I believe not. Underdetermination and even indeterminacy do *not* imply impossibility. As Quine stresses in the context of translation, there *do* exist adequate translation schemes. As soon as we select one such scheme, questions of interpersonal sameness of meaning have well-defined, though translation-scheme-dependent, answers. Quine's point is not that adequate translation is impossible. Rather, it is that no adequate translation scheme is determined *uniquely* by the available evidence. The underdetermination between alternative adequate translation schemes can be broken only by *non-empirical* considerations, such as conventions or considerations of parsimony.[24] In the case of the attribution of meanings to another speaker of my *own* language, for example, the *homophonic* translation scheme – which translates "Rabbit!" for the other English speaker into "Rabbit!" for me –, while empirically underdetermined, seems more parsimonious than the non-standard translation scheme which translates "Rabbit!" for the other English speaker into "Undetached rabbit part!" for me.

Similarly, to defend the possibility of interpersonal comparisons of utility, even on the view that such comparisons are indeterminate, we require an explanation of how the underdetermination between rival attributions of utilities to persons can be broken in a non-arbitrary way. I believe that the present account points towards at least two possible such explanations, independent from each other. The first one, assigning normative significance to certain states of affairs, is compatible even with the narrow evidence of cases 1 and 2 above. The second one, positing a fixed connection between certain empirically observable 'proxies' and utility, requires the richer evidence of cases 3a/b/c.

7.1. Assigning Normative Significance to Certain Options or States of Affairs

Given the weak evidence of cases 1 and 2, we cannot use constraints like (N1 a/b/c) for breaking the underdetermination between rival attributions of utilities to persons. But suppose that we identify some fixed options (or states of affairs) x_0 and/or y_0 in X as normatively significant. We might interpret these options, respectively, as 'deprivation' and 'saturation' consumption bundles of goods/resources (or as the state of consuming these bundles). And suppose further that we impose (some of) the following additional conditions on the attribution of utilities to persons:

[24] When asked why mutually incompatible, yet equally empirically adequate translation schemes never seem to occur in practice, Quine responds that the terrain has already been conquered by existing translation schemes and certain long-standing conventions.

(N2 a) ("options/states x_0 and y_0 each generate the same utility level for all persons")
$u_1(x_0) = u_2(x_0) = ... = u_n(x_0)$, and $u_1(y_0) = u_2(y_0) = ... = u_n(y_0)$.

(N2 b) ("a switch from option/state x_0 to option/state y_0 generates the same welfare gain/loss for all persons")
$u_1(y_0)-u_1(x_0) = u_2(y_0)-u_2(x_0) = ... = u_n(y_0)-u_n(x_0)$, where, for each $i \in N$, person i prefers y_0 to x_0.

(N2 c) ("option x_0 generates the same interpersonally significant norm level of utility for all persons")
$u_1(x_0) = u_2(x_0) = ... = u_n(x_0) = \alpha_0$, where α_0 is a fixed real number, in particular $\alpha_0 = 0$.

If we identify a single option x_0 (e.g. a 'deprivation' consumption bundle) as normatively significant and impose condition (N2 c), then the evidence of case 1 generates (a version of) condition (ONC+0) in table 1, determining (ZC)-statements. This, in turn, is sufficient for the existence of aggregation procedures satisfying all of Arrow's conditions simultaneously (List, 2001).

If we identify two distinct options x_0 and y_0 (e.g. a 'deprivation' consumption bundle and a 'saturation' consumption bundle, respectively) as normatively significant and impose condition (N2 b), then the evidence of case 2 generates condition (CUC) in table 1, determining (UC)-statements. If we identify two such options x_0 and y_0 and impose condition (N2 a) (which implies (N2 b)), then the evidence of case 2 generates condition (CFC) in table 1, determining both (LC)- and (UC)-statements. Either of these cases is sufficient for the existence of aggregation procedures satisfying all of Arrow's conditions simultaneously (Sen, 1970/1979).

More generally, the following table shows the logical relation between the types of evidence introduced above, conditions (N2 a/b/c) and the condition listed in table 1.

If we have					Then we have
(RankEv)	(RankEvLot)	(N2 a)	(N2 b)	(N2 c)	
yes	no	no	no	no	(ONC)
yes	no	no	no	yes	(ONC+0)
yes	yes	no	no	no	(CNC)
yes	yes	no	no	yes	(RNC)
not possible to generate here					(OLC)
yes	yes	no	yes	no	(CUC)
yes	yes	yes	yes	no	(CFC)
yes	yes	yes	yes	yes	(RFC)

Table 7

We often make the (normative) assumption that, for sufficiently similar people, similar states of affairs (e.g. "two people live in similar environments, both have a happy family and many friends, and both have similar jobs, etc.") generate similar levels of utility.[25] Such a (normative) assumption is effectively an informal instance of what conditions (N2 a), (N2 b) and (N2 c) capture in more formal terms. As we have seen, even if only as few as one or two such normatively distinguished options or states of affairs are identified, the underdetermination between rival attributions of utilities to persons can be broken – of course, in a *non-empirical* way, but nonetheless, by stipulation, in a *normatively significant* one.

7.2. Positing a fixed connection between empirically observable proxies and utility

The evidence of cases 3a/b/c may seem unrealistically rich. But, on closer inspection, the evidence of cases 1 and 2 may seem unrealistically sparse, and cases 3a/b/c may seem a better description of the types of evidence we use when we make (what look like) interpersonal comparisons of utility in everyday life. In making such comparisons, we seem to rely on evidence over and above people's revealed preference orderings. In particular, we seem to rely on a range of behavioural and physiological proxies for utility, such as a person's facial expression and other gestures, body language, the sound of a person's voice and a person's verbal self-description of his or her level of pleasure or pain. This body of evidence might be seen as an informal instance of what cases 3a/b/c describe in an idealized form.

As soon as we use evidence as described by cases 3a/b/c, there are non-arbitrary ways of breaking the underdetermination between rival attributions of utility to a set of persons. Even if we do not accept one of conditions (N1 a/b/c) as *true* in a realist sense, we can adopt what we described as the 'pragmatic' view in section 6.2.3 and accept one of conditions (N1 a/b/c) as a *principle of parsimony*. On such a view, conditions (N1 a/b/c) are analogous to the convention in translation to give priority to homophonic translation schemes over non-standard translation

[25] Conditions (N2 a), (N2 b), (N2 c) can be replaced with the following more refined conditions that allow the identification of person-specific normatively significant options/states (thereby acknowledging, for example, the possibility that different persons have different 'deprivation' or 'saturation' consumption bundles):

(N2' a) $u_1(x_{01}) = u_2(x_{02}) = ... = u_n(x_{0n})$, and $u_1(y_{01}) = u_2(y_{02}) = ... = u_n(y_{0n})$;

(N2' b) $u_1(y_{01})\text{-}u_1(x_{01}) = u_2(y_{02})\text{-}u_2(x_{02}) = ... = u_n(y_{0n})\text{-}u_n(x_{0n})$,
where, for each $i \in N$, person i prefers y_{0i} to x_{0i};

(N2' c) $u_1(x_{01}) = u_2(x_{02}) = ... = u_n(x_{0n}) = \alpha_0$,
where α_0 is a fixed real number, in particular $\alpha_0 = 0$;

where, for each $i \in N$, x_{0i} and y_{0i} are the options in X identified as normatively significant for person i.

schemes, provided that we have no empirical reason to reject a homophonic translation scheme in favour of a non-standard one.

The argument for the possibility of interpersonal comparisons of utility, then, is the following. First, we use evidence as described by cases 3a/b/c and, second, so long as empirical adequacy permits, we accept conditions (N1 a/b/c) for breaking the underdetermination between rival attributions of utility to persons. As we have seen above, cases 3a, 3b or 3c, jointly with conditions (N1 a), (N1 b) or (N1 c), respectively, are sufficient not only for determining interpersonal comparisons of utility, but also for the existence of aggregation procedures satisfying Arrow's minimal conditions.

7.3. Concluding Remarks

We have seen that, even if interpersonal comparisons are not determined by the available evidence, the underdetermination can be broken non-arbitrarily (though not purely empirically) if (i) we assign normative significance to certain states of affairs or (ii) we posit a fixed connection between certain empirically observable 'proxies' and utility.

We can speculate whether the present account of the possibility of interpersonal comparisons of utility captures the actual mechanisms by which we make (what look like) interpersonal comparisons in everyday life. As suggested above, in everyday life, we may be inclined to attribute similar utility levels two different persons for similar options or similar states of affairs, so long as these persons are sufficiently similar. This is in essence an instance of (i). If we make interpersonal comparisons in this fashion, the underlying mechanism might be an informal instance of the account given in section 7.1. Alternatively, suppose that the evidence we actually use in attributing utilities to persons is richer than the sparse evidence of cases 1 and 2. And suppose in particular that we do rely on (more informal versions of) the kinds of proxies described by cases 3a/b/c, attributing similar utilities for similar observable proxies. This is in essence an instance of (ii). If we make interpersonal comparisons in this fashion, the underlying mechanism might be an informal instance of the account given in section 7.2.

A psychological account of how we actually make (what look like) interpersonal comparisons of utility in everyday life is beyond the scope of this paper. Rather, the argument of this paper might be interpreted as an existence argument, showing that interpersonal comparisons of utility are not in principle impossible, and clarifying the logical structure that evidence and auxiliary assumptions must have in order to provide a basis for interpersonal comparisons of utility. Indeterminacy does not imply impossibility, and even if we hold that interpersonal comparisons of utility are indeterminate, we do not need to claim that such comparisons cannot in principle be made.

Acknowledgements

Work on this paper was begun while the author was a Harsanyi Program Visitor in the Social and Political Theory Program at the Research School of Social Sciences, Australian National University, and completed while he was a Visiting Scholar at the Center for Basic Research in the Social Sciences, Harvard University, and the Department of Linguistics and Philosophy, MIT. Previous versions of the paper were presented at a seminar at the ANU and at the 2000 conference of the Australasian Association Association of Philosophy, held at the University of Queensland, July 2000. The author wishes to express his gratitude to Geoffrey Brennan, Campbell Brown, Robert Goodin, Daniel Harbour, Karen Jones, Daniel Nolan, Philip Pettit, Michael Wagner, and two anonymous reviewers for helpful comments, and to the German National Merit Foundation and the Australian National University for financial support.

Nuffield College
Oxford OX1 1NF, U.K.
christian.list@nuffield.oxford.ac.uk

References

Arrow, K.: 1951/1963, Social Choice and Individual Values, Wiley, New York.

d'Aspremont, C.: 1985, ‘Axioms for social welfare orderings’, in Hurwicz, L., Schmeidler, D., and Sonnenschein, H. (eds.): Social Goals and Social Organization, Cambridge University Press, Cambridge, 19-76.

Bossert, W., and Weymark, J. A.: 1996, ‘Utility in social choice’, forthcoming in Barberà, S., Hammond, P. J., and Seidel, C. (eds.): Handbook of Utility theory, Volume 2, Kluwer, Boston.

Chomsky, N.: 1969, ‘Quine’s Empirical Assumptions’, in Davidson, D., and Hintikka, J. (eds.): Words and Objections: Essays on the Work of W. V. Quine, D. Reidel, Dordrecht, 53-68.

Davidson, D.: 1974, ‘Belief and the Basis of Meaning’, Synthese 27, 309-323.

Davidson, D.: 1986, ‘Judging interpersonal interests’, in Elster, J., and Hylland, A. (eds.): Foundations of Social Choice Theory, Cambridge University Press, Cambridge.

Debreu, G.: 1954, ‘Representation of a Preference Ordering by a Numerical Function’, in Thrall, Coombs and Davies (eds.): Decision Processes, Wiley, New York.

Elster, J., and Roemer, J. E. (eds.): 1991, Interpersonal Comparisons of Well-Being, Cambridge University Press, Cambridge.

Gibson, R.: 1986, ‘Translation, Physics, and Facts of the Matter’, in Hahn, L. E., and Schilpp, P. A. (eds.): The Philosophy of W. V. Quine, Open Court, La Salle, Ill.

Harsanyi, J.: 1955, ‘Cardinal welfare, individualistic ethics and interpersonal comparisons of utility’, Journal of Political Economy 63, 309-321.

Hausman, D.: 1995, ‘The Impossibility of Interpersonal Utility Comparisons’, Mind 104, 473-490.

Jevons, S.: 1911, The Theory of Political Economy (4th ed.), Macmillan, London.

List, C.: 1999, ‘Craig's Theorem and the Empirical Underdetermination Thesis Reassessed’, Disputatio 7, 28-39.

List, C.: 2001, ‘A Note on Introducing a 'Zero-Line' of Welfare as an Escape-Route from Arrow's Theorem’, Pacific Economic Review 6(2), special section in honour of Amartya Sen, 223-238.

von Neumann, J., and Morgenstern, O.: 1944, Theory of Games and Economic Behavior, Princeton University Press, Princeton.

Quine, W. V.: 1960, Word and Object, MIT Press, Cambridge, MA.

Quine, W. V.: 1970, ‘On the Reasons for Indeterminacy of Translation’, Journal of Philosophy 67, 178-183.

Robbins, L.: 1932, An Essay on the Nature and Significance of Economic Science, Macmillan, London.

Sen, A. K.: 1970/1979, Collective Choice and Social Welfare, Holden-Day, San Franscisco (1970); North Holland, Amsterdam (1979).

Waldner, I.: 1972, ‘The Empirical Meaningfulness of Interpersonal Utility Comparisons’, Journal of Philosophy 69, 87-103.