CHAPTER 7

Are Interpersonal Comparisons of Utility Indeterminate?

Abstract. On the orthodox view in economics, interpersonal comparisons of utility, interpreted as subjectively experienced well-being or happiness, are not empirically meaningful and "hence" impossible. To reassess this view and its implications, this chapter develops the parallel, first hinted at by Davidson, between the problem of interpersonal comparisons of utility and the problem of translation of linguistic meaning, as explored by Quine. The chapter discusses several cases of what the empirical evidence for interpersonal comparisons of utility might be and shows 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 we assign normative significance to certain states of affairs or we accept a condition that posits a fixed connection between certain empirically observable 'proxies' and utility. It is concluded that, even if interpersonal comparisons are not empirically meaningful, they are not impossible, and the identified ways of breaking the underdetermination point towards explanations of the ease with which (what look like) interpersonal comparisons are made in everyday life.

1. Introduction

The present chapter is concerned with our basis for making interpersonal comparisons of utility. By utility I mean a person's subjectively experienced well-being or happiness (as distinct from the concept of utility as preference satisfaction). Examples of interpersonal comparisons of utility are statements of the form "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 of the form "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).

There are many situations in everyday life in which we make what look like interpersonal comparisons of utility. We often seem to attribute certain utility levels or utility gains or losses to people and compare these across different people. Interpersonal comparisons seem to play a great 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 of course far from clear on closer reflection.

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 upon 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 for decades, 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$^1$.

Note that these arguments raise at least two different questions, which are often 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 (to be 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 to be 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 interpersonal comparisons of utility are considered possible has far-reaching implications for utilitarian theories of justice and for welfare economics.

$^1$ 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; Hausman, however, 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.
Arrow's impossibility theorem (1951/1963) confirms the intuition that interpersonal comparisons of utility are of crucial importance to the solubility of certain collective decision problems: Arrow's theorem shows that, if the effects of outcomes on persons are specified only in terms of interpersonally-non-comparable individual (ordinal) utility (or some other interpersonally-non-comparable welfare evaluation standard), there exists no procedure for aggregating such individual utility information into overall collective preference orderings or collective choices in accordance with a rather undemanding set of minimal conditions (to be stated below). If, on the other hand, the effects of outcomes on persons are specified in terms of interpersonally comparable utility (or some other interpersonally comparable welfare evaluation standard), Arrow's impossibility theorem no longer applies, and there are aggregation procedures satisfying Arrow's minimal conditions (see Sen, 1970/1979)\(^2\). We will briefly review these results in section 4 below.

The purpose of this chapter is to reassess the status of interpersonal comparisons of utility. In particular, the chapter develops the parallels, hinted at by Davidson (1974, 1986), between the problem of interpersonal comparisons of utility and the problem of translation of linguistic meaning, as explored by Quine (1960, 1970). I argue that situating the problem of interpersonal comparisons of utility in a Quinean/Davidsonian framework points towards a way of reconciling the main insight underlying the orthodox view on interpersonal comparisons, namely the insight that interpersonal comparisons of utility are empirically underdetermined, with a plausible explanation of the ease with which we make (what look like) interpersonal comparisons in everyday life. I suggest that the empirical underdetermination between different (and, when it comes to interpersonal comparisons, mutually incompatible), yet equally empirically adequate rival attributions of utilities to persons may be broken non-arbitrarily in a way similar to the one in which the empirical underdetermination between mutually incompatible, yet equally empirically adequate rival attributions of meanings to different speakers is broken by long-standing conventions.

\(^2\) 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. Moreover, it should be obvious that 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. The present paper, however, is not committed to any specific view on the question of which evaluation standards are morally relevant and why this is so.
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 turn out to be 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?).

Before turning to the main issues of the chapter, I will briefly introduce the concepts of underdetermination and indeterminacy.

2. Underdetermination and Indeterminacy

To introduce the concepts of underdetermination and indeterminacy, we 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 observations $\Phi$, a theory $T$ is said to be adequate with respect to $\Phi$ if $T$ implies the sentences in $\Phi$. 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, so long as it is adequate, implies the empirical observations, but the empirical observations do not necessarily imply the theory. A theory $T$ (or a specific theoretical statement $\tau$) is said to be determined by a set of observation sentences $\Phi$ if $\Phi$ implies $T$ (or $\tau$). A theory $T$ (or a specific theoretical statement $\tau$) is said to be underdetermined by $\Phi$ if $T$ (or $\tau$) is consistent with, but not determined by, $\Phi$. If $T$ (or $\tau$) is underdetermined by $\Phi$, then there exists an alternative theory $T'$ (or alternative theoretical statement $\tau'$) such that $T'$ (or $\tau'$) is also consistent with $\Phi$, but $T$ and $T'$ ($\tau$ and $\tau'$) are mutually inconsistent (for a more detailed account of the concept of underdetermination, 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 $\Phi$, we are faced with a situation of indeterminacy if each of the given alternative theories is underdetermined by $\Phi$ and there exists no theory-independent fact of the
matter, represented by a larger set of true sentences $\Psi$ (where $\Phi \subset \Psi$), possibly epistemically inaccessible to us, such that one of the alternative theories is determined by $\Psi$.3

3. Profiles of Utility Functions and Interpersonal Comparisons

Let $N = \{1, 2, ..., n\}$ be a set of persons, and let $X = \{x, y, x_1, x_2, y_1, y_2, ...\}$ be 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 $u_i : X \to \mathbb{R}$ 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 $x$.

An interpersonal comparison of utility levels is a statement of the form

\[(LC) \quad u_i(x) \geq u_j(y), \text{ where } i, j \in N, x, y \in X, i \neq j;\]

and an interpersonal comparison of utility units is a statement of the form

\[(UC) \quad \frac{u_i(x_1) - u_i(y_1)}{u_i(x_2) - u_i(y_2)} = \lambda, \text{ where } i, j \in N, x_1, y_1, x_2, y_2 \in X, i \neq j, u_i(x_2) \neq u_i(y_2) \text{ and } \lambda \in \mathbb{R}.\]

We shall add to these two commonly discussed types of interpersonal comparisons a third, less commonly acknowledged one. 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

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

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

(ZC)-statements are meaningful only if a utility level of zero can be shown to be a meaningful concept. A utility level of zero would somehow capture the 'dividing line' between 'utility' and 'disutility' or between 'pleasure' and 'pain'. Although (ZC)-statements make explicit reference only

\[3 \text{ On Quine's account, physical theories are underdetermined by the totality of observable evidence without being indeterminate, while translation is indeterminate, in so far as translation schemes are underdetermined by the totality of relevant observable linguistic behaviour and there is no theory-independent fact of the matter (given a naturalistic ontology) that would break the underdetermination between rival adequate translation schemes (for a more detailed account of the relation between underdetermination, indeterminacy and facts of the matter, see Gibson, 1986).} \]
to one person, they can be interpreted as a form of *interpersonal* comparisons in so far as they enable us to make comparisons of utility levels between persons in situations with utility level greater than zero, persons in situations with utility level precisely equal to zero, and persons in situations with utility level less than zero.

Once we have attributed a profile of utility functions $\{u_i\}_{i \in N}$ to the persons in $N$, we can of course make (LC)-, (UC)- and (ZC)- statements *relative to that profile*. However, whether these statements can be regarded as meaningful in a broader sense depends crucially on *how unique* the profile $\{u_i\}_{i \in N}$ is. Suppose, for instance, that each utility function $u_i$ in that profile is unique only up to positive monotonic transformations $^4$. In other words, suppose that we consider the profiles $\{u_i\}_{i \in N}$ and $\{u^*_i\}_{i \in N}$ to be *informationally equivalent* whenever, for each $i$, $u^*_i = \phi_i(u_i)$, where $\phi_1, \phi_2, \ldots, \phi_n$ are positive monotonic transformations (possibly different ones for different $i$). Then (LC)-, (UC)- and (ZC)-statements are not in general invariant under these transformations, and hence they are not considered meaningful, in that they are not independent of which specific profile we select as a representative from amongst a large class of informationally equivalent profiles. The present assumption about how unique a profile of utility functions is (stated in terms of the conditions under which two profiles of utility functions are informationally equivalent) is called *ordinal measurability, no interpersonal comparability of levels or units* (ONC). We will say that interpersonal comparisons, in the form of (UC)-, (LC)- or (ZC)-statements, are meaningful only if they are invariant under the class of transformations up to which $\{u_i\}_{i \in N}$ is unique.

The following table shows the relation between different classes of transformations and the invariance (or lack thereof) of (UC)-, (LC)- or (ZC)-statements under these transformations (a survey of different informational assumptions about measurability and interpersonal comparability and their social-choice-theoretic implications can be found in d’Aspremont, 1985).

\footnote{A positive monotonic transformation is a function $\phi : \mathbb{R} \to \mathbb{R}$ with the property that, for any $s, t \in \mathbb{R}$, $s < t$ implies $\phi(s) < \phi(t)$. A positive affine transformation is a function $\phi : \mathbb{R} \to \mathbb{R}$ with the property that there exist $a, b \in \mathbb{R}$ ($b > 0$) such that, for all $t \in \mathbb{R}$, $\phi(t) = a + b*t$.}
### Table 1

<table>
<thead>
<tr>
<th>Case:</th>
<th>The profiles ( { u_i }_{i \in N} ) and ( { u^*<em>i }</em>{i \in N} ) are informationally equivalent if ...</th>
<th>(LC)-statements</th>
<th>(UC)-statements</th>
<th>(ZC)-statements</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(ONC):</strong> Ordinal Measurability, No Interpersonal Comparability of Levels or Units</td>
<td>( u^*_i = \phi(u_i) ), where ( \phi_1, \phi_2, ..., \phi_n : \mathbb{R} \to \mathbb{R} ) are positive monotonic transformations</td>
<td>not invariant</td>
<td>not invariant</td>
<td>not invariant</td>
</tr>
<tr>
<td><strong>(ONC+0):</strong> Ordinal Measurability with an Interpersonally Significant Zero-Line, No Interpersonal Comparability of Levels or Unity</td>
<td>( u^*_i = \phi(u_i) ), where ( \phi_1, \phi_2, ..., \phi_n : \mathbb{R} \to \mathbb{R} ) are positive monotonic and sign-preserving transformations</td>
<td>not invariant</td>
<td>not invariant</td>
<td>invariant</td>
</tr>
<tr>
<td><strong>(CNC):</strong> Cardinal Measurability, No Interpersonal Comparability of Levels or Units</td>
<td>( u^*_i = \phi(u_i) ), where ( \phi_1, \phi_2, ..., \phi_n : \mathbb{R} \to \mathbb{R} ) are positive affine transformations</td>
<td>not invariant</td>
<td>not invariant</td>
<td>not invariant</td>
</tr>
<tr>
<td><strong>(RNC):</strong> Ratio-Scale Measurability, No Interpersonal Comparability of Levels or Units</td>
<td>( u^*_i = \phi(u_i) ), where ( \phi_1, \phi_2, ..., \phi_n : \mathbb{R} \to \mathbb{R} ) are positive linear transformations</td>
<td>not invariant</td>
<td>not invariant</td>
<td>invariant</td>
</tr>
<tr>
<td><strong>(OLC):</strong> Ordinal Measurability, Interpersonal Comparability of Levels</td>
<td>( u^*_i = \phi(u_i) ), where ( \phi : \mathbb{R} \to \mathbb{R} ) is a positive monotonic transformation</td>
<td>invariant</td>
<td>not invariant</td>
<td>not invariant</td>
</tr>
<tr>
<td><strong>(CUC):</strong> Cardinal Measurability, Interpersonal Comparability of Units</td>
<td>( u^<em>_i = a_i + b^</em> u_i ), where ( a_1, a_2, ..., a_n \in \mathbb{R} ) and ( b \in \mathbb{R} ) with ( b &gt; 0 )</td>
<td>not invariant</td>
<td>invariant</td>
<td>not invariant</td>
</tr>
<tr>
<td><strong>(CFC):</strong> Cardinal Measurability, Interpersonal Comparability of Levels and Units</td>
<td>( u^*_i = \phi(u_i) ), where ( \phi : \mathbb{R} \to \mathbb{R} ) is a positive affine transformation</td>
<td>invariant</td>
<td>invariant</td>
<td>not invariant</td>
</tr>
<tr>
<td><strong>(RFC):</strong> Ratio-Scale Measurability, Interpersonal Comparability of Levels and Units</td>
<td>( u^*_i = \phi(u_i) ), where ( \phi : \mathbb{R} \to \mathbb{R} ) is a positive linear transformation</td>
<td>invariant</td>
<td>invariant</td>
<td>invariant</td>
</tr>
</tbody>
</table>

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.

Before turning to the empirical basis for interpersonal comparisons in more detail, we shall briefly review the implications of interpersonal comparability of utility for Arrow's impossibility theorem. The main argument of the chapter 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 aggregates each profile $\{u_i\}_{i \in N}$ in a given domain into 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 $y$". $R$ also induces a strong ordering on $X$, defined by, 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, i.e. $F(\{u_i\}_{i \in N}) = F(\{u^*_i\}_{i \in N})$ whenever $\{u_i\}_{i \in N}$ and $\{u^*_i\}_{i \in N}$ are informationally equivalent.

Arrow's minimal conditions are the following:

**Universal Domain (U).** The domain of $F$ is the set of all logically possible profiles of utility functions.

**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 $i \in N$, $u_i(x_1) > u_i(x_2)$.

**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 $x_1R^*x_2$.

**Non-Dictatorship (D).** $F$ must not be dictatorial: there must 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 $R = F(\{u_i\}_{i \in N})$.

Now 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.** Given (ONC), there exists no SWFL $F$ satisfying (U), (P), (I) and (D).

The following table shows the logical implications of the cases in table 1 for the existence or non-existence of SWFLs satisfying Arrow's minimal conditions (with regard to the entries of the right-most column, see d’Aspremont, 1985; and List, 2001).
Table 2

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

The crucial observation to make is that 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

Translation involves the attribution of linguistic meanings to different speakers. When 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 when 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, since we have no introspective first-person access to other minds. According to Quine's famous 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; and which of these we adopt has potentially far-reaching
repercussions for the translation of more theoretical sentences. Apart from the fact that these alternative translations may strike us as less parsimonious and less intuitively compelling from the perspective of our own 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 positing any in principle inaccessible facts of the matter is at odds with Quine's naturalistic ontology, Quine's conclusion is that translation and, more generally, judgments of interpersonal sameness of meaning are indeterminate.

Similarly, if utility (at least on the definition of utility as subjectively experienced well-being or happiness) can be experienced only from a first-person perspective, the utilities experienced by another person are not directly observable by us. Rather, we can only observe the behaviour of this person, including their choice behaviour and their verbal expressions, and possibly other physiological responses, ranging from their 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 this speaker's linguistic behaviour, the attribution of utility to a person involves theorizing on the basis of whatever facts about this person are accessible from our 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. The argument that even the totality of such evidence underdetermines interpersonal comparisons of utility can be summarized quite succinctly in the words of Jevons (1911, p. 14): "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." 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 at first sight inclined to attribute identical utilities to these persons. But, if Jevons's argument is correct, the rival hypothesis that one of the two persons, the 'utility monster', is one thousand

---

5 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 in fact 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.
times more susceptible to pleasure and pain than the other, while in some sense less parsimonious
and intuitive, is equally compatible with all available empirical evidence. Now 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 in the previous section, the utilities subjectively experienced by the persons in \( N \) in
response to 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 empirical evidence. The theory to be built consists of a profile of utility functions
\( \{u_i\}_{i \in N} \) and some auxiliary assumptions about how subjectively experienced utility surfaces in
observable ways. It is only after we have attributed a profile of utility functions to the persons in
\( N \) that we can make comparisons between the utilities of different persons. Thus making
interpersonal comparisons of utility is a two-step process. In a first step, we attribute to the set of
persons \( N \) 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 then use the
attributed profile of utility functions to make interpersonal comparisons of utility.

Schematically, the logical relation between empirical evidence (box 1), an attributed profile of
utility functions (box 2) and interpersonal comparisons is as follows (box 3):
Box 2

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

implies

may not imply (?)

Box 1

empirical evidence about the persons in $N$:
- choice behaviour
- verbal expressions
- other measurable indices of pleasure or pain

implies

Box 3

interpersonal comparisons:
- (LC) Is $u(x) \geq u(y)$?
- (UC) What is $(u(x_1)-u(y_1))/(u(x_2)-u(y_2))$?
- (ZC) What is sign($x$)?

Table 3

The relation between box 2 and box 3 is one of logical implication: As we have seen, 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: the onus 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 2 which enable us to determine a profile of utility function in box 2 uniquely up to a sufficiently small class of transformations.

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

6.1. Different Types of Empirical Evidence

As will become clear, each of the presently discussed cases of what the empirical evidence might be represents an idealized limiting case, positing a body of evidence that is richer than what we realistically expect to find empirically. However, this is not harmful in the context of the present argument. If we are faced with 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. The formal conditions stated in table 4 will be discussed more informally below. The argument can be informally followed even if the technical details of table 4 are skipped.

---

6 Whether any of the introduced types of evidence do actually represent evidence for subjectively experienced utility is a philosophical question the present paper does not propose to resolve. As indicated above, interpreting a body of empirical observations as evidence for subjectively experienced utility requires certain auxiliary assumptions about how subjectively experienced utility surfaces in observable ways. Amongst these auxiliary assumptions are relatively common ones such as the assumption 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 in the next sub-section 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 that attributing utilities to people on the basis of empirical evidence involves a substantial act of theorizing that may suffer from underdetermination and possibly indeterminacy problems.
Given a set of options $X$, let $L(X)$ denote the set of all binary lotteries in $X$, where a binary lottery is an option of the form $p^*x+(1-p)^*y$, where $x, y \in X$, $p \in [0,1]$ and $p^*x+(1-p)^*y$ is interpreted to mean "get either $x$ or $y$ with associated probabilities $p$ and $1-p$, respectively". Note that $X \subseteq L(X)$, since each option $x \in X$ can be interpreted as a binary lottery $1^*x+0^*y \in L(X)$, where $y$ is any option. Further, note that $P$, the strong ordering induced by the weak ordering $R_p$, is defined as follows: for each $i \in N$, $x \in R_p y$ if and only if $xR_p y$ and not $yR_p x$.

(E1) The evidence includes all true statements of the form
- $xR_p y$, where $x, y \in X$, $i \in N$ and $xR_p y$ is interpreted to mean "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 (i.e. a weak linear ordering) on $X$.

(E1') The evidence includes all true statements of the form
- $xR_p y$, where $x, y \in L(X)$, $i \in N$ and $xR_p y$ is interpreted to mean "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 $xP_i y$ and $yP_i z$, then there exist $\lambda, \mu \in (0,1)$ such that $(\lambda^*x+(1-\lambda)^*y)P_i (\mu^*x+(1-\mu)^*z)$;

(P3) ("independence") for each $i \in N$, all $x, y, z \in X$ and all $\lambda \in (0,1)$, $xR_p y$ if and only if $(\lambda^*x+(1-\lambda)^*y)R_i (\lambda^*x+(1-\lambda)^*z)$.

(E2 a) The evidence includes all true statements of the form
- $f(x) = \lambda$, where $x \in X$, $i \in N$, $\lambda \in \mathbb{R}$ and $f : X \rightarrow \mathbb{R}$ is some real-valued measurable index of person $i$'s expression of pleasure or pain in response to options in $X$ (e.g. a person's facial expression, heart beat, body temperature, or relevant neural activity) satisfying (given that we also have (E1))

(P4 a) ("consistency of $f$-response with preference") there exist a positive monotonic transformation $\phi : \mathbb{R} \rightarrow \mathbb{R}$ and some profile of von Neumann-Morgenstern utility functions $\{u_i^*\}_{i \in N}$ representing $\{R_i\}_{i \in N}$ such that, for all $i \in N$ and all $x, y \in X$, $f(x) = \phi(u_i^*(x))$.

(E2 b) The evidence includes all true statements of the form
- $g(x, y) = \lambda$, where $x, y \in X$, $i \in N$, $\lambda \in \mathbb{R}$ and $g_i : X \rightarrow \mathbb{R}$ is some real-valued measurable index of person $i$'s expression of pleasure or pain in response to differences/switches between options in $X$ (e.g. a person's spontaneity of choosing $x$ over $y$) satisfying (given that we also have (E1))

(P4 b) ("consistency of $g$-response with preference") there exist a positive monotonic transformation $\psi : \mathbb{R} \rightarrow \mathbb{R}$ and some profile of von Neumann-Morgenstern utility functions $\{u_i^*\}_{i \in N}$ representing $\{R_i\}_{i \in N}$ such that, for all $i \in N$ and all $x, y \in X$, $g(x, y) = \psi(u_i^*(x)-u_i^*(y))$.

(E2 c) The evidence includes all true statements of the form
- $h(x) = \delta$, where $x \in X$, $i \in N$, $\delta \in \{-1,0,1\}$ and $h_i : X \rightarrow \mathbb{R}$ is some -1/0/1-valued pleasure/pain function (taking values 1='pleasure', 0='neutral', -1='pain') in response to options in $X$ satisfying (given that we also have (E1))

(P4 c) ("consistency of $h$-response with preference") there exists some profile of von Neumann-Morgenstern utility functions $\{u_i^*\}_{i \in N}$ representing $\{R_i\}_{i \in N}$ such that, for all $i \in N$ and all $x, y \in X$, $h_i(x) = \text{sign}(u_i^*(x))$.

Table 4
In terms of the conditions stated in table 4, we will focus on the following cases.

Case 1. We have (E1).
Case 2. We have (E1').
Case 3a. We have (E1) and (E2 a).
Case 3b. We have (E1') and (E2 b).
Case 3c. We have (E1) and (E2 c).

In case 1, each person's utility function surfaces only in the form of this person's revealed preference ordering $R_i$ over the options in $X$. That is, given an apple, an orange and a banana, for instance, we can determine each person's preference ordering over these three fruits.

In case 2, each person's utility function surfaces only in the form of this person's revealed preference ordering $R_i$ over all options and binary lotteries in $X$. That is, we can determine, for instance, not only that a person prefers an orange to a banana to an apple, but also whether or not, for any given probability $p$, the person would prefer a guaranteed banana to a lottery whose prize would be either an orange or an apple with associated probabilities $p$ and $1-p$, respectively.

In cases 3a, 3b, 3c, each a rather utopian best case scenario, each person's utility function surfaces in the form of this person's revealed preference ordering $R_i$ over all options (and, in case 3b, all binary lotteries) in $X$ and also in the form of some other measurable indices of this person's expression of pleasure or pain in response to options in $X$, formalized here by the functions $f_i$, $g_i$, or $h_i$. These indices are to be interpreted as further proxies of a person's utility in addition to revealed preference. They 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 (in case 3a, real-valued; in case 3c, -1/0/1-valued) or in response to switches between options (in case 3b, also real-valued), or a person's spontaneity of choosing one option over another in a forced-choice situation (also in case 3b) (see Waldner, 1972) (assuming that a greater such spontaneity corresponds to a greater utility gain), or something else that might be thought of as a proxy for a person's utility. No claims are made as to whether it is at all realistic to think that such additional

---

7 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 shall assume that what makes $f_i$ and $g_i$ observable in the relevant sense 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.
observable proxies of utility are available. The point is only to ask what the logical implications of such a utopian best case scenario for the problem of interpersonal comparisons of utility would be.

6.2. Implications

The following two standard representation theorems show that cases 1 and 2 generate, respectively, cases (ONC) and (CNC) in table 1, and therefore leave (UC)-, (LC)- and (ZC)-statements underdetermined. If we believe in addition that there is no theory-independent fact of the matter about what the 'true' interpersonal comparisons of utility are, interpersonal comparisons of utility are indeterminate.

**Theorem 2.** (standard) 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 $u_i : X \to \mathbb{R}$ such that, for all $x, y \in X$, $x R_i y$ if and only if $u_i(x) \geq u_i(y)$. Moreover, if $u_i$ has this property, then so does $\phi(u_i)$, where $\phi : \mathbb{R} \to \mathbb{R}$ is any positive monotonic transformation.

**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 $u_i : L(X) \to \mathbb{R}$ such that (i) for all $x, y \in X$, $x R_i y$ 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 $\phi(u_i)$, where $\phi : \mathbb{R} \to \mathbb{R}$ is any positive affine transformation.

In cases 3a and 3c, theorem 1 also applies and shows that, if we use only evidence of type (E1) or (E1') but not of types (E2 a/c), each person's utility function is determined uniquely only up to positive monotonic transformations. Once again, this is case (ONC) in table 1, which leaves (UC)-, (LC)- and (ZC)-statements underdetermined.

In case 3b, theorem 2 applies and shows that, if we use only evidence of type (E1) but not of type (E2 b), each person's utility function is determined uniquely only up to positive affine transformations. This is case (CNC) in table 1, which also leaves (UC)-, (LC)- and (ZC)-statements underdetermined.

However, the situation may change once we use the additional evidence of types (E2 a), (E2 b) or (E2 c). The availability of this additional evidence means that we can use not only the persons'
revealed preferences, but also the other observable proxies as a potential basis for interpersonal comparisons of utility. In particular, the conditions (P4 a/b/c) imply two things. 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. Second, it is possible (but of course by no means necessary) to choose a profile of utility functions for the persons in such a way (namely \( \{u_i\}_{i \in N} = \{u^*_i\}_{i \in N} \)) that the transformations describing the functional relation between the utility function \( u_i \) and the proxy functions \( f_i, g_i \), or \( h_i \) (i.e. the transformations describing how utility surfaces in the form of the proxy functions \( f_i, g_i \), or \( h_i \)) are the same for all persons (in cases 3a and 3b, the transformation is a positive monotonic transformation; in case 3c, the transformation is the sign-function).

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") a profile of utility functions \( \{u_i\}_{i \in N} \) is adequate only if there exists a positive monotonic transformation \( \phi: \mathbb{R} \to \mathbb{R} \) such that, for all \( i \in N \) and all \( x, y \in X \), \( f(x) = \phi(u_i(x)) \).

- **(N1 b)** ("interpersonal sameness of the conversion of utility into the proxy functions") a profile of utility functions \( \{u_i\}_{i \in N} \) is adequate only if there exists a positive monotonic transformation \( \psi: \mathbb{R} \to \mathbb{R} \) such that, for all \( i \in N \) and all \( x, y \in X \), \( g(x, y) = \psi(u_i(x) - u_i(y)) \).

- **(N1 c)** ("interpersonal sameness of the conversion of utility into the proxy functions") a profile of utility functions \( \{u_i\}_{i \in N} \) is adequate only if, for all \( i \in N \) and all \( x \in X \), \( h(x) = \text{sign}(u_i(x)) \).

Conditions (N1 a/b/c) state that a profile of utility functions is adequate only if, according to that profile, all persons have identical ways of converting utility into the observable proxies \( f_i, g_i \), or \( h_i \). This rules out the possibility, in cases (3 a/b) with (N1 a/b), that different persons exhibit identical \( f_i \) or \( g_i \) values and yet their underlying utilities are supposed to be different, or, in case (3 c) with (N1 c), that they exhibit identical \( h_i \) values and yet they are not in the same one of the three states ‘pleasure’, ‘neutral’, ‘pain’.
As we will now see, cases 3a, 3b and 3c, jointly with conditions (N1 a), (N1 b), (N1 c), generate, respectively, cases (OLC), (CUC) and (ONC+0) in table 1, determining, respectively, (LC)-, (UC)- and (ZC)-statements.

In case 3a, 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 affecting the adequacy of the resulting profile. Rather, if we choose to apply a positive monotonic transformation to one person’s utility function, i.e. if, for some \(i \in N\), we replace \(u_i\) with \(\theta(u_i)\), where \(\theta: \mathbb{R} \rightarrow \mathbb{R}\) is a positive monotonic transformation, then we are also forced to replace \(\phi\) with \(\phi^*\), where, for all \(i \in \mathbb{R}\), \(\phi^*(t) = \phi(\theta^{-1}(t))\), where \(\theta^{-1}\) is the inverse transformation of \(\theta\), and 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, and we have case (OLC) in table 1, determining (LC)-statements.

In case 3b, if we accept condition (N1 b), similarly, if we apply a positive affine transformation to one person’s utility function, i.e. if, for some \(i \in N\), we replace \(u_i\) with \(a_i + b^* u_i\), then we are also forced to replace \(\psi\) with \(\psi^*\), where, for all \(t \in \mathbb{R}\), \(\psi^*(t) = \psi(t/b)\), and consequently we are forced to replace \(u_i\) with \(a_i + b^* u_i\) for every \(i \in N\). Note that, while the \(a_i\) may be different for different persons, \(b\) must be the same for all persons. Hence we have case (CUC) in table 1, determining (UC)-statements.

In case 3c, if we accept condition (N1 c), the only positive monotonic transformations we can apply to each \(u_i\) without affecting the adequacy of the resulting profile are sign-preserving ones, for if we replace \(u_i\) with \(\theta(u_i)\), where \(\theta\) is not sign-preserving, the requirement that, for all \(x \in X\), \(h_i(x) = \text{sign}(u_i(x))\) may be violated. Hence we have case (ONC+0) in table 1, determining (ZC)-statements.

Note that each of cases 3a, 3b and 3c, jointly with conditions (N1 a), (N1 b) and (N1 c), respectively, is sufficient for the existence of aggregation procedures satisfying all of Arrow's conditions simultaneously.

If we do not accept conditions (N1 a/b/c), there are no similar restrictions on the transformations that may be applied to the utility functions of different persons. We would then have to admit the possibility, in cases (3 a/b), that different persons exhibit identical \(f_i\) or \(g_i\) values and yet their underlying utilities are different, or, in case (3 c), that they exhibit identical \(h_i\) values and yet they
are not in the same one of the three states ‘pleasure’, ‘neutral’, ‘pain’.

If we were to multiply person 1’s utility function by a factor of 10 while leaving all other utility functions unchanged (or to apply a non-sign-preserving transformation to person 1’s utility function), for instance, we would also have to accept that person 1’s ‘rate’ of converting utility into observable $f_i$ or $g_i$ values is divided by a factor of 10 (or that, for person 1, $h_i$ values of 1, 0 and -1 do not correspond to the interpersonally significant states ‘pleasure’, ‘neutral’ and ‘pain’, respectively). 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 cases (ONC) or (CNC) in table 1, and (UC)-, (LC)- and (ZC)-statements remain underdetermined. Once again, if we believe in addition that there is no theory-independent fact of the matter about what the ‘true’ interpersonal comparisons of utility are, interpersonal comparisons of utility are indeterminate.

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

<table>
<thead>
<tr>
<th>If we have</th>
<th>Then we have</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E1)</td>
<td>(E1')</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>no or yes</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>yes</td>
<td>no or yes</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 5

Comparing tables 2 and 5, an important observation to make is that those cases 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 certain collective decision problems.

There are at least three different views one might take on the status of conditions (N1 a/b/c):

On the first (and strongest) 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 source of this view might be the view that utility is either identical with, or systematically reducible to, or in an interpersonally identical way correlated with, certain (neuro-)physiological states. On such a view, identical (neuro-)physiological states of the relevant kind, expressed -- in the present framework -- in terms of the proxy functions $f_i$, $g_i$, or $h_i$, would then, literally, indicate identical utilities (or, in the case of $h_i$, identical interpersonally significant states ‘pleasure’, ‘neutral’ and ‘pain’).

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 would be bad methodology to attribute different utilities (or different interpersonally significant states ‘pleasure’, ‘neutral’, ‘pain’) to them; good methodology would require us to attribute identical utilities (identical states 'pleasure', 'neutral', 'pain') 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 himself 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)

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.
7. Yet Another Impossibility Argument?

Schematically, the argument of the present chapter 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)

Is there a theory-independent fact of the matter about utility?

Yes
Interpersonal comparisons of utility are underdetermined but not indeterminate.

No
Interpersonal comparisons of utility are indeterminate.

preference orderings over options (and in case 3b, binary lotteries) and other measurable indices of pleasure or pain (cases 3 a/b/c)

Are (N1 a/b/c) or some equivalent condition true?

Yes
Interpersonal comparisons of utility are determined (LC) (3a), (UC) (3b), (ZC) (3c).

No

Is there a theory-independent fact of the matter about utility?

Yes
Interpersonal comparisons of utility are underdetermined but not indeterminate.

No
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 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 theory-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 not once again make a mystery of the ease with which people make, and reach intersubjective agreement on, (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, and 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, but rather that no adequate translation scheme is determined uniquely by the available evidence, and that the underdetermination between alternative adequate translation schemes can be broken, if at all, only by non-empirical considerations, such as conventions or considerations of parsimony. 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 vastly more parsimonious than the non-standard translation scheme which translates "Rabbit!" for the other English speaker into "Undetached rabbit part!" for me.

Similarly, to explain how people make (what look like) 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 plausible such explanations, independent from each other. The first one, assigning normative significance to certain options or states of affairs, is compatible even with the narrow evidence of cases 1 and 2 above. The second one, invoking one of conditions (N1 a/b/c) as a principle of parsimony, requires the richer evidence described by cases 3a/b/c.

---

8 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.
7.1. Assigning Normative Significance to Certain Options or States of Affairs

The evidence of cases 1 and 2 is by itself too weak for the use of constraints like (N1 a/b/c) for breaking the underdetermination between rival attributions to 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 (for instance, by interpreting them, respectively, as 'deprivation' and 'saturation' consumption bundles of goods/resources) and impose (some of) the following additional conditions on the attribution of utilities to persons:

(N2 a) ("options/states $x_0$ and $y_0$ each generate the same utility level for all persons")

$$u_1(x_0) = u_2(x_0) = \ldots = u_n(x_0), \text{ and } u_1(y_0) = u_2(y_0) = \ldots = u_n(y_0).$$

(N2 b) ("a switch from option/state $x_0$ to option/state $y_0$ generates the same welfare gain/loss for all persons")

$$u_1(y_0) - u_1(x_0) = u_2(y_0) - u_2(x_0) = \ldots = u_n(y_0) - u_n(x_0), \text{ where everyone prefers } y_0 \text{ to } x_0.$$

(N2 c) ("option $x_0$ generates the same interpersonally significant norm level of utility for all persons")

$$u_1(x_0) = u_2(x_0) = \ldots = u_n(x_0) = \alpha_0, \text{ where } \alpha_0 \text{ is a fixed real number, in particular } \alpha_0 = 0.$$

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 case (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).

Further, 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 case (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 case (CFC) in table 1, determining both (LC)- and (UC)-statements. Once again, 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 cases listed in table 1.
If we have

<table>
<thead>
<tr>
<th>(E1)</th>
<th>(E1')</th>
<th>(N2 a)</th>
<th>(N2 b)</th>
<th>(N2 c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>No</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>No</td>
<td>yes</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>No</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>No</td>
<td>yes</td>
</tr>
</tbody>
</table>

not possible to generate here

| yes  | yes   | no     | yes    | no     |
| yes  | yes   | yes    | yes    | no     |
| yes  | yes   | yes    | yes    | yes    |

Table 7

People often seem to make the (normative) assumption that, for sufficiently similar people\(^9\), similar states of affairs (e.g. "two people are both healthy, both have a happy family and many friends, and both have a good job, etc.") generate (or should be taken to generate) similar levels of utility. Such a (normative) assumption is effectively an informal instance of what conditions (N1 a), (N2 b) and (N2 c) capture in more formal terms. And 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. Invoking Conditions (N1 a/b/c) as Principles of Parsimony

At first sight, the evidence of cases 3a/b/c may seem unrealistically rich. But, on reflection, the evidence of cases 1 and 2 may actually seem unrealistically sparse, and cases 3a/b/c may actually seem a better description of the types of evidence we use in making interpersonal comparisons of utility in a folk-psychological manner. In making such comparisons, we usually 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.

\(^9\) 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):

\[(N2' a) \quad u_1(x_{01}) = u_2(x_{02}) = \ldots = u_n(x_{0n}), \text{ and } u_1(y_{01}) = u_2(y_{02}) = \ldots = u_n(y_{0n});\]

\[(N2' b) \quad u_1(y_{01}) - u_i(x_{01}) = u_2(y_{02}) - u_i(x_{02}) = \ldots = u_n(y_{0n}) - u_i(x_{0n}),\]

where, for each \(i \in N\), person \(i\) prefers \(y_{0i}\) to \(x_{0i}\);

\[(N2' c) \quad u_1(x_{01}) = u_2(x_{02}) = \ldots = u_n(x_{0n}) = \alpha_0,\]

where \(\alpha_0\) 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 \(i\).
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 can be seen as an informal instance of what cases 3a/b/c describe in an idealized manner.

But, as soon as we use what is essentially an instance of the evidence of cases 3a/b/c, there are non-arbitrary ways of breaking the underdetermination between rival attributions of utility to a set of persons that would all be consistent with these persons' revealed preferences. Even if we do not accept one of conditions (N1 a/b/c) as true in a realist sense, we can take what we described as the 'pragmatic' view in the previous section and accept one of conditions (N1 a/b/c) as a principle of parsimony. On such a view, conditions (N1 a/b/c) can be regarded as analogous to the convention in translation to give priority to homophonic translation schemes over non-standard translation schemes, so long as we have no empirical reason to reject a homophonic translation scheme in favour of a non-standard one.

On the present account, the reason why we easily make, and agree on, (what look like) interpersonal comparisons of utility is the following. First, we use (a more informal version of) the type of evidence described (in an idealized manner) by cases 3a/b/c and, second, so long as empirical adequacy permits, we are inclined to accept (more informal versions of) constraints like conditions (N1 a/b/c) for breaking the underdetermination between rival attributions of utility to persons, just as we are inclined to give priority to homophonic translation schemes. And 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.

In conclusion, even if we hold the view that interpersonal comparisons of utility are indeterminate, there are non-arbitrary ways of breaking the underdetermination between rival attributions of utilities to persons. Thus indeterminacy does not imply impossibility, and it is perfectly possible to hold that interpersonal comparisons of utility are indeterminate without making a mystery of the ease with which such comparisons are made in everyday life.