On the Representation Theorems

of Neoclassical Utility Theory: A Comment

Daniel Mahoney, PhD
Atlanta, GA

I would like to thank Robert P. Murphy for many helpful discussions on these issues. This is in no way implies his agreement with my conclusions.

The purpose of this note is to provide some clarification on the so-called representation theorems of neoclassical utility theory. In particular, it is argued that there is some misunderstanding regarding precisely what conclusions may be properly drawn from these theorems, and that there seems to be a tendency among neoclassical economists to misapply these theorems. Reference will be made to an ongoing debate over Austrian-school interpretations of cardinality in neoclassical utility theory.

According to a standard (Mas-Colell et al., 1995, p. 9) theoretical microeconomics text, the definition of a utility representation function is the following:

> A function u:X -> R is a utility function representing preference relation >~ if, for all x, y in X, x >~ y <=> u(x) >= u(y).

(Here R is the set of real numbers, X represents some set of alternatives, x and y are possible alternatives, and x >~ y means that "the consumer" values x at least as much as y.) The various representation theorems deal with proving properties of u given certain properties of the preference relation >~. For example, given so-called rationality conditions on the preference relation (e.g., completeness, transitivity), then it can be proved that a continuous representation function u exists. Further conditions can be used to establish differentiability and other properties.

However, these proofs are beside the point. The *only* thing they establish is that such a function *exists*, not that there is any equivalence between the preference relation and the utility function "representing" it. In other words, they merely permit one to *restate* in mathematical terms the verbal conditions expressing preference. They in no way establish equivalence between the utility function and the preferences, and so any results derived from the mathematical manipulations of the representation function cannot necessarily be applied to the preferences themselves.

In other words, simply because one can *express* a preference relation mathematically, it does not follow that one may act “as if” the preference relation behaves in ways its mathematical proxy does. The proxy serves one purpose; this does not imply it serves others. In particular, one cannot invoke the various representation theorems to claim Austrian criticisms of neoclassical utility theory (specifically, that utility is being treated as cardinal) are mistaken; these theorems do not provide the neoclassical equations with a sound ordinal foundation. If these equations are to have any economic meaning, it can only be if they assume, for example, cardinal utility. (It should be obvious that only cardinal numbers can be arithmetically manipulated.)

The issue has to do with the apparent assumption that every property of this mathematical function u can be attributed to the preference relation. This is entirely unfounded. For example, simply because one can take derivatives of this function in no way implies the resulting quantity has any relation to "marginal utility," the (subjective) value from an additional unit, or anything else of economic importance. In other words, it seems that attempts are being made to derive properties of the preference relation that cannot be derived from the preference relation itself, but only its mathematical representation. As no equivalence has been established, this approach is invalid.

Key to the subsequent development of the standard neoclassical results is the assumption that something real is being said about preferences based on properties of this function u. This assumption is unwarranted. As the Austrians have stated in this debate, there is only one kind of utility (ordinal rankings), and however one may *state* preferences mathematically, one is unjustified in going any further. No such justification is put forward in Mas-Colell's book, to give an example, and it is highly unlikely that one can be given, as the preference relation and the representation function are ontologically different.

For example, one can express a preference for two beers over one by saying u(two beers) = 3 and u(one beer) = 2. But it in no way follows that the marginal utility of the second beer is 1. This function u is useful *only* for mathematically re-stating verbal results; it provides no new information itself. Yet, one must assume such a meaning in order to make sense of equations like MUx/MUy = px/py; i.e., to connect a marginal rate of substitution with utility.

For example, one can *infer* a marginal rate of substitution (MRS) from manipulations of this function u, but it does not then follow that MRS is anything real, as it is not derived from the preference relation itself, only the mathematical representation of preference. (Even if one could derive the notion of MRS from the preference relation, it still would not follow that it was the ratio of partial derivatives of the representation function.) Barring a deductive proof of equivalence (between preference and the function used in a restatement of preference), the only option remaining, it would seem, is to treat utility as some sort of measurable quantity that can be observed empirically.

The issue is not whether preferences are known in totality or not, or whether assumptions of continuity are realistic or not, or whether the relevant equations are invariant up to a monotonic transformation of the representation function. The issue is whether this mathematical representation of preference has any logical (realistic or not) connection with preference itself, beyond a mere restatement of preference. The point of the Austrians is that preferences do not possess the "desirable" properties of the functions used to represent them, so any implications so drawn are questionable. Put another way, it does not follow that reality has all the properties of a mathematical model of reality.

Now it may be conceded that neoclassicals are in fact fully aware of these issues, and that they do not consider the representation function to attribute any kind of cardinality or measurability to utility (as defined in terms of preference). In this case, though, it would seem that they would have to admit that their resulting equations do not have any kind of economic content at all. An honest neoclassical would have to say: "the expression u(x) > u(y) means that x is preferred to y. We are now going to divorce this function u from its previous economic meaning and perform a series of mathematical manipulations on it. In a different context this function u can be used to *state* or express the condition of preference. It is not clear, though, what the mathematical results really imply." However, I am unaware that such disclaimer is ever made.

Indeed, in Mas-Colell's textbook (p. 46), the following statement is made: "For analytical purposes, it is very helpful if we can summarize the consumer's preferences by means of a utility function because mathematical programming techniques can then be used to solve the consumer's problem." These authors then go on to prove the *existence* of such a utility function under certain conditions, but nowhere establish that mathematical manipulations of this function provide any kind of new information about the preferences. The presumed suitability of using "mathematical programming techniques" (e.g. constrained optimization) merely begs the question of why one would think they can be used to solve the "consumer's problem."

Perhaps a comparison with other branches of science will be of use. Neoclassical economics seems to have adopted uncritically the methods of mathematics and the physical sciences, so it would be instructive to briefly consider how those disciplines go about establishing knowledge. In mathematics, a relation between (abstract) objects is made (that is, proved from some definitions, axioms, and/or postulates), and then the consequences of this relation are traced out. These consequences are valid if the premises can be established as true (and of course the logic and argumentation is not faulty). In the physical sciences, a relation between objects is hypothesized, and the (mathematical) consequences of this relation can be subjected to empirical testing. This approach presumes the existence of time-invariant causes and the results derived are only tentative, pending a correct judgment on the suitability of their application.

In the case of the representation theorems under discussion, neither approach (or any at all) has been undertaken to establish a connection deeper than mere restatement. It has neither been proven rigorously or axiomatically that the representation functions which *re-express* preference are equivalent to the underlying preferences, so that these functions can reveal information that the preferences themselves do not, nor has this deeper connection been established experimentally. Indeed, the latter approach would evidently require some sort of assumption of measurable utility in order to take place, and of course would also presume the existence of causality regarding human action, a very dubious proposition (Hoppe, 1993, ch. 7).

If the neoclassical equations are to be taken seriously as statements of economic reality, then Austrians seem to be justified in criticizing neoclassicals for treating utility as cardinal. If the equations are not to be so interpreted, one wonders what the point of the neoclassical research program is. In this case the Austrian "error" would seem to be in taking these equations more seriously than the neoclassicals themselves do.

To summarize, the representation theorems of neoclassical utility theory merely guarantee that a mathematical expression of (ordinal) utility is possible; they do NOT allow one to infer any information about the preferences not revealed by the preferences themselves. This is an additional and, it must be said, largely glossed-over step. Therefore, any mathematical manipulations of these functions (e.g., through maximization or computation of "marginal rates of substitution"), no matter how successfully they avoid recourse to a conception of utility as cardinal, cannot be said to reveal any kind of economic information. The mathematics is simply mathematics, using expressions which have economic meaning in a different context. It is another example of the danger of using mathematics in the social sciences.

References

Hoppe, Hans-Hermann, 1993, "The Economics and Ethics of Private Property," Kluwer Academic Publishers.

Mas-Colell, Andreu, Whinston, Michael D., and Green, Jerry R., 1995, "Mircroeconomic Theory," Oxford University Press.