# The Relation Between Dewey's "A Representation Theorem for Decisions about Causal Models" and the von Neumann-Morgenstern Theorem

Bill Hibbard
22 October 2012

Dewey's forthcoming AGI-12 paper [1] (congratulations Daniel) proves a result about when preferences can be represented by utility functions. His result bears a strong resemblance to the von Neumann-Morgenstern Theorem [2, 3, 4]. But he does not mention that theorem which piqued my curiosity to investigate the relation between these two results. The investigation gave me a clearer understanding of both results, so I am sharing it in this note.

## The von Neumann-Morgenstern Theorem

Define a set of mutually exclusive *outcomes* $O_i$, $i = 1, 2, 3, ...$ and a space $L$ of *lotteries* as sums $\Sigma_i\, p_i\, O_i$ where $\Sigma_i\, p_i = 1$ and each $p_i \geq 0$. Define a *preference relation* $\preccurlyeq$ on $L$ including *indifference* $\approx$ (please see the references [2, 3, 4] for more detailed explanations). The von Neumann-Morgenstern Theorem says that the preference relation can be represented by a utility function $u : L \rightarrow \mathbf{R}$ mapping lotteries to real numbers (the representation is: $l \preccurlyeq m \Leftrightarrow u(l) \leq u(m)$) if and only if a set of four conditions are true. The utility function is linear in the sense that $u(\Sigma_i\, p_i\, O_i) = \Sigma_i\, p_i\, u(O_i)$. The conditions are:

C1 (Completeness) $\forall l, m \in L$ exactly one of $l \prec m$, $m \prec l$ or $l \approx m$ is true.

C2 (Transitivity) $\forall l, m, n \in L.\ l \preccurlyeq m \wedge m \preccurlyeq n \Rightarrow l \preccurlyeq n$.

C3 (Continuity) If $l \preccurlyeq m \preccurlyeq n$ then $\exists p \in [0, 1].\ p\, l + (1\text{-}p)\, n \approx m$.

C4 (Independence) If $l \prec m$ then for any $n$ and $p \in (0, 1]$, $p\, l + (1\text{-}p)\, n \prec p\, m + (1\text{-}p)\, n$.

## Dewey's Representation Theorem

A *causal model* M consists of a set of variables $X$, each defined by a function of other variables $f(Y)$ or by a constant value $x$ (please see Dewey's paper [1] for a more detailed explanation). Let $A$ be a set of *acts* each of the form $\langle M, x \rangle$ where $M$ is a causal model and $x$ is a value for a variable $X$. $M_x \in S$ is a *submodel* resulting from the act $\langle M, x \rangle$ where $X$'s function is replaced by the constant function $X = x$. $S$ is the set of submodels. Define a *preference relation* $\preccurlyeq$ on $A$ including *indifference* $\approx$. Dewey's theorem says that the

preference relation can be represented by a utility function $u : S \rightarrow \mathbf{R}$ mapping submodels to real numbers (the representation is: $\langle M, x\rangle \preccurlyeq \langle M', y\rangle \Leftrightarrow u(M_x) \leq u(M'_y)$) if and only if a set of four conditions are true. The conditions are:

C1' (Completeness) $\forall l, m \in A$ exactly one of $l \prec m$, $m \prec l$ or $l \approx m$ is true.

C2' (Transitivity) $\forall l, m, n \in A.\ l \preccurlyeq m \wedge m \preccurlyeq n \Rightarrow l \preccurlyeq n$.

C3' (Function-independence) $\langle M, x\rangle \approx \langle M_{X=f(Y)}, x\rangle$ where $M_{X=f(Y)}$ is the model derived by replacing $X$'s function with $f$ over values of $Y$ in $M$.

C4' (Variable-independence) $X = x \wedge Y = y$ in M $\Rightarrow \langle M, x\rangle \approx \langle M, y\rangle$.

**The Relation Between the Two Results**

It seems to me that there is a strong resemblance between the two results. Both results say that a preference relation can be represented by a utility function if and only if a set of four conditions on the preference relation hold. Dewey's first two conditions C1' and C2' are essentially identical to the von Neumann-Morgenstern conditions C1 and C2 (merely replacing the set $L$ of lotteries by the set $A$ of acts). The first two conditions are required on the preference relations because these conditions apply to the order relation (i.e., $\leq$) on real numbers that represents the preference relations.

The difference is in the third and fourth conditions and there is no way that C3' and C4' can be derived from C3 and C4. The difference in these conditions is due to the difference in the ways the utility functions are defined in the two cases. In the von Neumann-Morgenstern theorem, the utility function is linear. The preference relation on lotteries must satisfy conditions C3 and C4 in order to be consistent with the linearity of the utility function representing it. In the Dewey theorem there is no linearity condition on the utility function so analogs of C3 and C4 are not needed.

To understand the need for C3' and C4' in the Dewey case consider that the preference relation on acts is represented by a utility function defined on submodels:

$$\langle M, x\rangle \preccurlyeq \langle M', y\rangle \Leftrightarrow u(M_x) \leq u(M'_y))$$

This is different from the von Neumann-Morgenstern case, where the preference relation and the utility function are both defined on lotteries. In this case it is necessary to include a map $w$: $A \rightarrow S$ from acts to submodels, defined by $w(\langle M, x\rangle) = M_x$. Applying $w$ to the representation gives:

$$\langle M, x\rangle \preccurlyeq \langle M', y\rangle \Leftrightarrow u(w(\langle M, x\rangle)) \leq u(w(\langle M', y\rangle))$$

But $w$ may map different acts to the same submodel (C3' and C4' provide ways to construct examples of this). In mathematical terminology, $w$ is not an injection. Conditions C3' and C4' ensure that when two acts map to the same submodel, the preference between the acts is indifferent. That is:

$$w(\langle M, x\rangle) = w(\langle M', y\rangle) \Rightarrow \langle M, x\rangle \approx \langle M', y\rangle$$

Since the conditions C3' and C4' are only needed because the preference relation and the utility function are defined on different sets (i.e., $A$ and $S$), why not represent the preference relation on $A$ with a utility function also defined on $A$ (or a preference relation defined on $S$ with a utility function defined on $S$)? In that case, the only conditions on the preference relation would be C1' and C2'. The representation theorem could say that any preference relation defined on $A$ (or $S$) that is complete and transitive can be represented by a utility function defined on $A$ (or $S$).

**References**


1. Dewey, D. 2012. A Representation Theorem for Decisions about Causal Models. In: Bach, J., Iklé, M., and Goertzel, B. (eds) AGI 2012. LNCS (LNAI). Springer, Heidelberg. http://www.danieldewey.net/representation-theorem-for-decisions-about-causal-models.pdf

2. Levin, J. 2006. Choice under Uncertainty. http://www.stanford.edu/~jdlevin/Econ%20202/Uncertainty.pdf

3. Voorneveld, M. 2010. Mathematical Foundations of Microeconomic Theory: Preference, Utility, Choice. https://studentweb.hhs.se/CourseWeb/CourseWeb/Public/PhD501/1001/micro1.pdf

4. Wikipedia. http://en.wikipedia.org/wiki/Von_Neumann-Morgenstern_utility_theorem