The von Neumann-Morgenstern theorem

Martin Peterson

Texas A&M University

Let Z be a finite set of basic prizes, which may include, for instance, a holiday in Malibu, a ticket to a rock concert, or almost any other kind of good. We will assume that:

1. Every basic prize in Z is a lottery.
2. If A and B are lotteries, then so is the prospect of getting A with probability p and B with probability 1-p, for every $0 \leq p \leq 1$.
3. Nothing else is a lottery.

ApB is a lottery in which the agent wins A with probability p and B with probability 1-p. Thus, the second condition stated above could equally well be formulated as follows: If A and B are lotteries, then so is ApB, for every $0 \leq p \leq 1$. Furthermore, because ApB is a lottery it follows that also Cq(ApB), for all $0 \leq q \leq 1$ is a lottery, given that C is a lottery. And so on and so forth. Consider the following axioms:

vNM 1 (Completeness) $A \succ B$ or $A \sim B$ or $B \succ A$

vNM 2 (Transitivity) If $A \succ B$ and $B \succ C$, then $A \succ C$

vNM 3 (Independence) $A \succ B$ if and only if $ApC \succ BpC$

vNM 4 (Continuity) If $A \succ B \succ C$ then there exist some p and q such that $ApC \succ B \succ AqC$

Theorem. The preference relation $\succ$ satisfies vNM 1- 4 if and only if there exists a function u that takes a lottery as its argument and returns a real number between 0 and 1, which has the following properties:

(1) $A \succ B$ if and only if $u(A) > u(B)$.

(2) $u(ApB) = pu(A) + (1-p)u(B)$.

(3) For every other function u' satisfying (1) and (2), there are numbers $c > 0$ and d such that $u' = c \cdot u + d$.

The von Neumann-Morgenstern theorem is an if-and-only-if claim, so we have to prove both directions of the biconditional. We first show that the axioms entail the existence part of the theorem, saying that there exists a utility function satisfying (1) and (2). In the second part of the proof, the uniqueness part, we prove that the utility function is unique in the sense articulated in (3). Finally, in the third part, we prove that if we have a utility function with properties (1)–(3) then the four axioms all hold true.

<u>Part One</u>

Let us start by constructing the utility function u mentioned in the theorem. Because Z is a finite set of basic prizes the completeness axiom entails that Z will contain some optimal element O and some worst element W. This means that O is preferred to or equally as good as every other element in Z, and every element in Z is preferred to or equally as good as W. Furthermore, O and W will also be the optimal and worst elements in all probabilistic mixtures of Z, i.e. in the set of all possible lotteries. This follows from the independence axiom. To grasp why, let A be an arbitrary non-optimal element of Z such that $O \succ A$. Then, by independence, it follows that $OpC \succ ApC$, no matter what C is. Hence, $OpO \succ ApO$, and because OpO is just an alternative way of representing O, it follows that $O \succ ApO$, and this holds for every p. Hence, if we start from an optimal basic prize O, we can never construct any lottery comprising any other non-optimal basic prize A that is strictly preferred to O; this insight can be immediately generalised to hold for any combination of basic prizes.

We now stipulate that:

(i) u(O) = 1 and u(A) = 1 for every A ~ O

(ii) u(W) = 0 and u(A) = 0 for every A ~ W.

The next step is to define utilities for all lotteries A between O and W, i.e. for all lotteries that are neither optimal nor worst-case. The continuity axiom entails that for every A, there exists some p such that

(iii) A ~ OpW

Now, if there is only one such number p we could stipulate that u(A) = p, for every A, $O \succ A \succ W$. In order to see that there is in fact only one p satisfying relation (iii) for every A, suppose for *reductio* that there is some other such number $r \neq p$ such that

(iv) A ~ OrW

From (iii) and (iv) it follows, by transitivity, that OpW ~ OrW. Now, there are two cases to consider, viz. p > r and r> p. First suppose that p > r. OpW and OrW can be represented as in table 1.

	r	**p-r**	**1-p**
OpW	O	O	W
OrW	O	W	W

Table 1.

By applying the independence axiom from right to left, i.e. by deleting the rightmost column, we find that there exists some probability s > 0 (never mind what it is) such that O ~ OsW. However, we showed above that it holds for every p that $O \succ ApO$ given that $O \succ A$; hence, $O \succ OsW$ and we have a contradiction. The analogous contradiction arises if we assume that r > q.

We have now shown that there exists a function u that assigns a number between 0 and 1 to every lottery, such that u(O) = 1, u(w) = 0, and for every other $A \neq O$, W, u(A) = p iff A ~

OpW. The next step is to verify that the utility function we have constructed satisfies properties (1) and (2). We start with (1), i.e. the claim that

(v) $A \succ B$ if and only if u(A) > u(B).

We prove both directions of the biconditional simultaneously. By stipulation u(A) = p iff A ~ OpW, so it therefore holds that A ~ Ou(A)W. For the same reason, B ~ Ou(B)W. Hence, $A \succ B$ iff Ou(A)W $\succ$ Ou(B)W. By applying the independence axiom it follows that $A \succ B$ iff u(A) > u(B). To see this, first suppose that u(A) = u(B). It then follows that Ou(A)W $\succ$ Ou(A)W, and by the independence axiom we have $O \succ O$, which violates the asymmetry condition derived from the completeness and transitivity axioms above. Next suppose that u(A) < u(B). This is also inconsistent with the observation that $A \succ B$ iff Ou(A)W $\succ$ Ou(B)W. To see this, we repeat a manoeuvre that should be familiar by now. Please take a look at table 2.

	u(A)	**u(B)-u(A)**	**1-u(B)**
Ou(A)W	O	W	W
Ou(B)W	O	O	W

Table 2.

The headings of each column denote probabilities, not utilities. Because of the independence axiom, the rightmost column can be deleted. It then follows that there is some probability s > 0 (never mind what it is) such that OsW $\succ$ O. However, it was shown above that $O \succ ApO$ for all p and A, which means that $O \succ OsW$, and we have a contradiction.

We shall now verify property (2) of the utility function, i.e. we wish to show that

(vi) $u(ApB) = pu(A) + (1-p)u(B)$

To start with, note that the independence axiom in conjunction with the ordering axiom guarantees that A ~ B iff ApC ~ BpC. (Because if A ~ B then neither $ApC \succ BpC$ nor $BpC \succ ApC$ can hold, so it has to be the case that ApC ~ BpC; the proof of the other direction of the biconditional is analogous.) It follows that

(vii) If A ~ BpC, then

AqD ~ (BpC)qD, and

DqA ~ Dq(BpC).

Recall that because of the way we constructed the utility function it holds by definition that:

A ~ Ou(A)W

B ~ Ou(B)W

ApB ~ Ou(ApB)W

Hence, by substituting Ou(A)W for A and Ou(B)W for B we get

(viii) ApB ~ Ou([Ou(A)W]p[Ou(B)W])W

This expression might look a bit complicated, but it can fortunately be reduced to a lottery comprising only basic prizes. This is because we assumed that the probability calculus applies to lotteries, i.e. that if pq + (1-p)r = s, then (AqB)p(ArB) ~ AsB. Hence, Ou([Ou(A)W]p[Ou(B)W])W ~ OsW, where s = pu(A) + (1-p)u(B). Because of transitivity, ApB ~ OsW, which entails that

(ix) Ou(ApB)W ~ OsW

Hence,

(x) Ou(ApB)W ~ O[pu(A) + (1-p)u(B)]W

By eliminating the probabilities connecting O and W we get

(xi) u(ApB) = pu(A) + (1-p)u(B)

This complets Part One of the proof.

Part Two

The aim of Part Two is to show that for every other function u' satisfying (1) and (2), there exist numbers $c > 0$ and d such that $u'(x) = c \cdot u(x) + d$. It is worth keeping in mind that u and

u' are two different utility scales assigning numbers to the same set of objects, and that we already know the value of each object on the u scale. Thus, let t be a function that transforms the u scale into the u' scale. The transformation performed by t can be conceived of as a two-step process: For every number x on the u scale, t first picks a lottery A such that u(A) = x, i.e. $u^{-1}(x) = A$. Then, in the second step, t assigns a new real number y = $u'(A)$ to lottery A. So, by definition

(xii) $t(x) = u'[u^{-1}(x)] = y$

Suppose that i and j are two arbitrary numbers on the u scale such that $i \geq j$, and note that for every k, $1 \geq k \geq 0$, it holds that $i \geq ki + (1 - k)j \geq j$. This means that the number ki + (1 - k)j is also on the u scale. Hence, by substitution

(xiii) $t(ki + (1 - k)j) = u'[u^{-1}(ki + (1 - k)j)] = y$

Now, note that $u^{-1}[(ki + (1 - k)j)]$ is a lottery according to the definition above, and the utility of this lottery is ki + (1 - k)j. Because i and j are also numbers on the u scale, these utilities also correspond to some lotteries, A and B, respectively, so i = u(A) and j = u(B). Hence,

(xiv) $t(ki + (1 - k)j) = u'[u^{-1}(ku(A) + (1 - k)u(B))]$

Because of property (2) of the theorem, it follows that

(xv) $t(ki + (1 - k)j) = u'(AkB)$

From Part One we know that u' also satisfies (2). Hence, $t(ki + (1 - k)j) = k u'(A) + (1\text{-}k) u'(B)$, from which it follows that

(xvi) $t(ki + (1 - k)j) = k u'(A) + (1\text{-}k) u'(B)$

We noted above that $i = u(A)$ and $j = u(B)$, and it follows from this that $t(i) = u'(A)$ and $t(j) = u'(B)$. By inserting this into equation (xvi), we get

(xvii) $t(ki + (1 - k)j) = kt(i) + (1\text{-}k)t(j)$

Now, $u'(x) = t[u(x)]$, and the right-hand side of this can be rewritten as $u'(x) = t(u(x)1 + [1 – u(x)]0)$. By applying this to equation (xvii) we get

(xviii) $u'(x) = u(x)t(1) + [1\text{-}u(x)]t(0) = u(x)[t(1) - t(0)] + t(0)$

Now we are almost done. We stipulate that $c = t(1) - t(0)$ and that $d = t(0)$, because we thereby get $u'(x) = cu(x)+ d$. All that remains to show is that $c > 0$. This is equivalent to showing that $t(1) > t(0)$. By definition, $t(1) = u'(u^{-1}(1)) = u'(O)$, and $t(0) = u'(u^{-1}(0)) = u'(W)$. Because $O \succ W$, it follows from property (1) of the theorem that $u'(O) > u'(W)$.

Part Three

The aim of this part is to show that if we have a utility function that satisfies (1)–(3), then the four axioms hold true. Property (1) directly entails the completeness axiom. If u(A) > u(B), then A $\succ$ B, and if u(B) > u(A) then B $\succ$ A, and if u(A) = u(B) then A ~ B. Property (1) also entails the transitivity axiom. If u(A) > u(B) and u(B) > u(C), then u(A) > u(C). Hence, if A $\succ$ B and B $\succ$ C, then A $\succ$ C. To verify the independence axiom, holding that $A \succ B$ iff $ApC \succ BpC$, we need to use property (2), i.e. the fact that $u(ApB) = pu(A) + (1-p)u(B)$. Because u(ApC) = pu(A) + (1 - p)u(C) and u(BpC) = pu(B) + (1 - p)u(C), all we have to show is that

(xix) u(A) > u(B) if and only if pu(A) + (1 - p)u(C) > pu(B) + (1 - p)u(C)