A Counter-Example to Nash's Derivation of Utility Theory

Johan E. Gustafsson*

Draft
February 21, 2024

ABSTRACT. I present a counter-example to Nash's derivation of utility theory in 'The Bargaining Problem'.

In his classic paper 'The Bargaining Problem', John Nash claims (without proof) that utility theory can be derived from a set of surprisingly weak axioms. He writes:[1]

> By making the following assumptions we are enabled to develop the utility theory of a single individual:
>
> 1. An individual offered two possible anticipations can decide which is preferable or that they are equally desirable.
>
> 2. The ordering thus produced is transitive; if A is better than B and B is better than C then A is better than C.
>
> 3. Any probability combination of equally desirable states is just as desirable as either.
>
> 4. If A, B, and C are as in assumption (2), then there is a probability combination of A and C which is just as desirable as [B]. This amounts to an assumption of continuity.
>
> 5. If $0 < p < 1$ and A and B are equally desirable, then $pA + (1 - p)C$ and $pB + (1 - p)C$ are equally desirable. Also, if A and B are equally desirable, A may be substituted for B in any desirability ordering relationship satisfied by B.
>
> These assumptions suffice to show the existence of a satisfactory utility function, assigning a real number to each anticipation of an individual. This utility function is not unique, that is, if u is such a function then so also is $au + b$, provided $a > 0$. Letting

* I would be grateful for any thoughts or comments on this paper, which can be sent to me at *johan.eric.gustafsson@gmail.com*.

[1] Nash 1950, pp. 156–7, with the correction to condition (4) noted in Nash 1996, p. 2.

capital letters represent anticipations and small ones real numbers, such a utility function will satisfy the following properties:

(a) $u(A) > u(B)$ is equivalent to A is more desirable than B, etc.

(b) If $0 \precsim p \precsim 1$ then $u[pA + (1-p)B] = pu(A) + (1-p)u(B)$.

Nevertheless, the following is a counter-example:

Suppose there are three exclusive categories of states and anticipations: GOOD, BAD, and NEUTRAL. Any probability combination of states of the same category also belongs to that category. Probability combinations of states of more than one category are NEUTRAL. All anticipations of the same category are equally desirable. And NEUTRAL anticipations are preferred to BAD anticipations, and GOOD anticipations are preferred to NEUTRAL and to BAD anticipations.

These preferences satisfy Nash's axioms but not both (a) and (b). If, for instance, A is GOOD, B is NEUTRAL, and $p = 1/2$, then $u[pA+(1-p)B]$ is NEUTRAL. Then, by (a), it holds that $u(A) > u(B) = u[pA+(1-p)B]$. So it follows that $u[pA+(1-p)B] < pu(A)+(1-p)u(B)$, which violates (b).

Kenneth J. Arrow may have thought of much the same counter-example, since he (discussing very similar assumptions) adds an axiom that rules out this example — namely, that 'there are at least four probability distributions, no two of which are indifferent (or all are indifferent).'[2] But much the same axiom (that there exist at least four distinct indifference sets) was first suggested by Herman Rubin.[3]

Given that we add Rubin's axiom to Nash's, Jacob Marschak likewise claims that we can derive utility theory.[4] Nevertheless, rather than Rubin's axiom, Marschak's proof makes use of the following axiom due to Abraham Wald: There exist two anticipations X and Y, both with a positive probability for all states, and X is preferred to Y.[5]

But, given Nash's axioms, we can derive Wald's axiom from Rubin's. Suppose there are four anticipations such that no two of them are indifferent. By axioms (1), (2), and the latter half of (5), we have a weak ordering of all anticipations. So there must be at least four distinct indifference sets in the ordering. Assume, for proof by contradiction, that all anticipations with a positive probability for all states belong to the same indifference set.

[2] Arrow 1951, p. 425n22. I thank Petter Wakker for this observation.

[3] Marschak 1950, p. 118. Arrow claims that his understanding of utility theory was derived in part from Rubin; see Fishburn and Wakker 1995, p. 1133.

[4] Marschak 1950, pp. 118, 137.

[5] Marschak 1950, p. 118.

Then, since there are at least four distinct indifference sets, there will be an anticipation in an indifference set that is two steps away in the ordering. Then, by axiom (4), another anticipation with a positive probability for all states must belong to the indifference set in between — which contradicts our assumption. So we can conclude that not all anticipations with a positive probability for all states belong to the same indifference set. And then, by axiom (1), Wald's axiom follows.[6]

I wish to thank Kacper Kowalczyk, Wlodek Rabinowicz, Dean Spears, H. Orri Stefánsson, and Peter Wakker for valuable comments.

References

Arrow, Kenneth J. (1951) 'Alternative Approaches to the Theory of Choice in Risk-Taking Situations', *Econometrica* 19 (4): 404–437.

Fishburn, Peter C. and Peter P. Wakker (1995) 'The Invention of the Independence Condition for Preferences', *Management Science* 41 (7): 1130–1144.

Herstein, I. N. and John Milnor (1953) 'An Axiomatic Approach to Measurable Utility', *Econometrica* 21 (2): 291–297.

Marschak, Jakob (1950) 'Rational Behavior, Uncertain Prospects, and Measurable Utility', *Econometrica* 18 (2): 111–141.

Nash, John F. Jr. (1950) 'The Bargaining Problem', *Econometrica* 18 (2): 155–162.

— (1996) 'The Bargaining Problem', in *Essays on Game Theory*, pp. 1–8, Cheltenham: Edward Elgar.

Resnik, Michael D. (1987) *Choices: An Introduction to Decision Theory*, Minneapolis: University of Minnesota Press.

[6] Two alternative paths to repairing Nash's derivation is to strengthen axiom (2) — that is, the continuity axiom — or to strengthen assumption (5) — that is, the independence axiom. See Resnik 1987, pp. 93–6 for a proof with a similar continuity axiom but with a stronger independence axiom. And see Herstein and Milnor 1953, pp. 293–7 for a proof with a similar independence axiom but with a stronger continuity axiom.