What does Harsanyi’s theorem tell us?

Marc Fleurbaey

Summary

- Harsanyi’s claim that his theorem justifies utilitarianism is exaggerated
  “the more complete our factual information and the more completely individualistic our ethics, the more the different individuals' social welfare functions will converge toward the same objective quantity, namely, the unweighted sum (or rather the unweighted arithmetic mean) of all individual utilities” (1955, p. 320)
- The opposite claim that Harsanyi’s theorem has no consequence is also exaggerated
  “If utility only has meaning as a representation of preferences on lotteries… no significance should be attached to the shape of the contours of the social welfare function” (Weymark 1991, p. 305)

- Main references: Weymark 1991, 2005
  + Mongin Ely Lecture (WP 2002)

Harsanyi's utilitarianism

- Pareto + EW → Weighted utilitarianism
- + Anonymity → Unweighted utilitarianism

- A debate on Anonymity
  - Anonymity is usually formulated in a multi-profile context: $f(U_i, U_j, U_{-ij}) = f(U_j, U_i, U_{-ij})$
  - But one can adapt it to single-profile: anonymity over utility *levels* (rather than *functions*)

Sen-Weymark: utility vs VNM

- Utilitarianism on X: $\sum_i U_i(x)$
- Theorem: Pareto + EW → social preferences on X are represented by $\sum_i \varphi_i(U_i(x))$, where $\varphi_i(U_i(x))$ is a VNM function

- Harsanyi (+VNM, Broome, Risse…): VNM functions do represent cardinal intensity of preferences
- Luce-Raiffa, Mongin: one can have declining intensity of preferences and be risk-loving

Utility vs VNM under risk

- What is utilitarianism under risk?
- Ex ante utilitarianism: $\sum_i U_i(p)$
- If $U_i(p)$ respects preferences but is not VNM, then $\sum_i U_i(p)$ is not EW: $\sum_i U_i(p) = \sum_i \varphi_i\left(EV_i(x)\right)$

- Ex post utilitarianism: $E\sum_i U_i(x)$
- If $U_i(x)$ is not VNM, then $E\sum_i U_i(x) = \sum_i EU_i(x)$ does not respect preferences over lotteries

Harsanyi's utilitarianism reformulated

- ***Theorem***: If utilitarianism is EW and respects preferences, then any social ordering that is EW and respects preferences is represented by $\sum_i \alpha_i U_i(p)$, where $U_i(p)$ is the utility function from the utilitarian SWF.
- Proof: Utilitarianism rep. by $\sum_i U_i(p)$, where each $U_i(p)$ is a VNM function

An ordinal formulation

- ***Theorem***: Assume $U_i(x_1) > U_i(x_0)$ for all i.
  If the social ordering is EW and respects preferences, then there is $(\alpha_i)_{i=1,\dots,n}$ such that for all p, all k,
  $$SMRS^0(p_1 / p_k) = \sum_i \alpha_i MRS_i^0(p_1 / p_k).$$

- $SMRS^0(p_1 / p_k) : dp_k, dp_1, dp_0 = -dp_k - dp_1.$

- Note: the weights do not depend on p → they do not depend on how well-off the individuals are.

Proof

- By Harsanyi's theorem, $W(p)=\sum_i \alpha_i \frac{U_i(p)-U_i(x_0)}{U_i(x_1)-U_i(x_0)}$

$SMRS^0(p_1/p_k):$

$dp_k W(x_k)+dp_1 W(x_1)-(dp_k+dp_1)W(x_0)=0$

$\Rightarrow \frac{dp_1}{dp_k}=-\frac{W(x_k)-W(x_0)}{W(x_1)-W(x_0)}=-\frac{W(x_k)}{W(x_1)}.$

$$SMRS^0(p_1/p_k)=-\frac{\sum_i \alpha_i\left(\frac{U_i(x_k)-U_i(x_0)}{U_i(x_1)-U_i(x_0)}\right)}{\sum_i \alpha_i\left(\frac{U_i(x_1)-U_i(x_0)}{U_i(x_1)-U_i(x_0)}\right)}=\frac{\sum_i \alpha_i MRS_i^0(p_1/p_k)}{\sum_i \alpha_i}.$$

Harsanyi's is an impossibility theorem