Kenneth Arrow's Last Theorem

Paul Milgrom[1]
May 15, 2024

***Abstract**. In Kenneth Arrow's last week of life at age 95, he reported that "I began my research career with an impossibility theorem. If I had time now, my last theorem would be an impossibility theorem about social choice for environmental policy." This paper completes the formalization, proof, and discussion of the theorem that Arrow then described.*

In his final days, Kenneth Arrow's interest in economic scholarship never waned. When I visited him, he wanted to discuss two topics: income equality and climate change. In one conversation, he proposed an impossibility theorem that applies especially to collective choice for environmental policies. Below, I introduce a complete mathematical model based on his description.

The finite set of agents $n \in N$ consume environmental amenities over time periods $t = 0,1,2$ …. We may interpret each agent as a clan or dynasty, which cares for its living and future members. The *state* of the system at time t is a positive integer $s_t \in S$, where S may be finite or infinite. In each period, the collective observes the state and takes an *action* $a_t \in \{1, \dots, A\}$, which governs the next state according to the transition probabilities $p(s_{t+1}|s_t, a_t)$.

A *policy* is a function that describes a pure or mixed action $\pi(s)$ for each state s. In each period, agent n earns flow utility $u_n(s_t, a_t) \in [0,1]$ that depends on the realized state and action. Let Π denote the set of policies. Each agent evaluates its future utility starting in state s as the discounted sum of future flow utilities, $U_{nt}^{\pi}(s) = \mathbb{E}_{\pi}\left[\sum_{\tau=t}^{\infty} \delta_n^{\tau-t} u_n\big(s_\tau, \pi(s_\tau)\big) \,|s_t = s\right]$, in which the policy determines the probability distribution over future states and actions and δ_n is n's discount factor. A policy π is Pareto dominated in state s by another policy or a mixture of policies π' if for all agents n, $U_{n0}^{\pi}(s) \le U_{n0}^{\pi'}(s)$ and for some n, the inequality is strict. A policy π is *efficient* if it is not Pareto dominated in any state.

Arrow proposed a single axiom to govern the collective choice: the chosen policy should be efficient. Using this one axiom, the simplest version of Arrow's Last Theorem is the following.

Theorem. Suppose that $p(\cdot \mid \cdot) > 0$ and that for all $n \neq m$, $\delta_n \neq \delta_m$. Then for any efficient policy π^*, there is an m (the "dictator") such that in every state, π^* is a most preferred policy for m.

Proof. Let $U_t^{\pi}(s) = \big(U_t^{\pi}(s)\big)_{n\in N}$ denote the payoff vector beginning at time t in state s with policy π and let $F(s) \in [0,1]^N$ denote the set of payoff vectors corresponding to feasible policies. Since policies may be randomized, F is convex and for any efficient policy π^*, $U_0^{\pi^*}(s)$ is on the boundary of $F(s)$. So $U_0^{\pi^*}(s)$

[1] I gratefully acknowledge support from the National Science Foundation, grant SES1947514. This report is also based in part upon work supported by the National Science Foundation under Grant No. DMS-1928930 and by the Alfred P. Sloan Foundation under grant G-2021-16778, while the author was in residence at the Simons Laufer Mathematical Sciences Institute (formerly MSRI) in Berkeley, California, during the Fall 2023 semester. I thank Marion Ott for her careful reading, Michael Crystal for his research assistance, Ilya Segal for pointing out the related paper by Jackson and Yariv, and a referee for suggesting improvements to the proof and its interpretation.

it is supported by a hyperplane. This means that for some nonzero vector $\lambda_s \in \mathbb{R}_+^N$ with $\sum_n \lambda_{ns} = 1$ and all $u \in F(s)$, $\lambda \cdot u \leq \lambda_s \cdot U_0^{\pi^*}$. In terms of policies, this inequality means:

$$(\forall s \in S)\ \pi^* \in \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_\pi \left[\sum_{\tau=0}^{\infty} \sum_{n \in N} \lambda_{ns} \delta_n^\tau u_n\big(s_\tau, \pi(s_\tau)\big) \middle| \, s_0 = s \right]$$

Fix any state s and let $m := \arg \max_{n:\lambda_{ns}>0} \delta_n$, that is, m is the agent with the highest discount factor among *influential* agents (agents with positive welfare weights). Then, π^* must also solve the continuation problem of maximizing the payoff after the first period given that a transition to state $s_1 = s'$ has occurred.

$$\pi^* \in \underset{\pi \in \Pi}{\operatorname{argmax}} \sum_{s'} p\big(s' \big| s, \pi^*(s)\big)\, \delta_m \mathbb{E}_\pi \left[\sum_{\tau=1}^{\infty} \sum_{n \in N} \frac{\delta_n}{\delta_m}\, \lambda_{ns} \delta_n^{\tau-1} u_n\big(s_\tau, \pi(s_\tau)\big) \middle| \, s_1 = s' \right]$$

So,

$$\pi^* \in \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_\pi \left[\sum_{\tau=0}^{\infty} \sum_{n \in N} \left(\lambda_{ns} \frac{\delta_n}{\delta_m} \right) \delta_n^\tau u_n\big(s_\tau, \pi(s_\tau)\big) \middle| \, s_0 = s \right]$$

Comparing this to the original maximization problem, the coefficients $\lambda_{ns} \frac{\delta_n}{\delta_m}$ have replaced the coefficients λ_{ns} of the original problem. Iterating this coefficient replacement step t times leads to:

$$\pi^* \in \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_\pi \left[\sum_{\tau=0}^{\infty} \sum_{n \in N} \left(\lambda_{ns} \left(\frac{\delta_n}{\delta_m} \right)^t \right) \delta_n^\tau u_n\big(s_\tau, \pi(s_\tau)\big) \middle| \, s_0 = s \right]$$

If $n \neq m$, then $\delta_n < \delta_m$. So, for all $n \neq m$, $\lim_{t \to \infty} \lambda_{ns} \left(\frac{\delta_n}{\delta_m} \right)^t = 0$, leading to:

$$\pi^* \in \underset{\pi \in \Pi}{\operatorname{argmax}}\, \mathbb{E}_\pi \left[\sum_{\tau=0}^{\infty} \lambda_{ms} \delta_m^\tau u_m\big(s_\tau, \pi(s_\tau)\big) \middle| \, s_0 = s \right].$$

So, π^* is a most preferred policy for m. ■

Discussion

The essence of the proof is to note that if π^* is an efficient policy starting in some state s today, that implies a maximization using welfare weights for agents that must be unchanging over time. However, efficiency requires that decisions in the distant future must give relatively greater weight to the preferences of more patient individuals, at least among agents with non-zero weights. Those two things can be consistent only if there is just one agent with a non-zero weight.

The dictatorship conclusion in Arrow's Last Theorem conjures memories of his famous Impossibility Theorem, but this new theorem differs in some important ways. The original Impossibility Theorem asserts that any *social welfare function* satisfying three reasonable axioms must reflect the preferences of just one agent, who is called the dictator. Similarly, Arrow's Last Theorem introduces an axiom – efficiency – and concludes that the socially chosen policy must be a most preferred policy of a single agent. However, the conclusion of Arrow's Last Theorem is different from his first impossibility theorem because the definitions of dictatorship are different: the agent that dictates the policy for one value profile in Arrow's Last Theorem may not be a dictator for the social welfare function.

To illustrate the difference, consider the social welfare function that works as follows. First, any given preference profile is mapped into an ordered list of agents. Then, the first-ranked agent identifies her most preferred policy or policies. If there are multiple most preferred policies, then the second-ranked agent picks her most preferred policies among those, and so on in a process of serial dictatorship until some policy is chosen. For *every* function that maps preference profiles to an ordered list of agents, the social welfare function represented by this two-step construction selects an efficient policy and for every profile, that policy is the most preferred one for some agent, but in contrast to Arrow's original impossibility theorem, different profiles may identify different dictators.

Arrow's Last Theorem is examines a property of a policy rather than of a social welfare functions. Consequently, my preferred interpretation is that it asserts that, when its conditions apply, it is impossible for a policy to efficiently *compromise* the interests of agents with different time preferences.[2] One may wonder: what rules out the possibility that the policy chooses a preferred action for less patient agents at early dates and for more patient agents at later dates? The answer is that the assumption that transition probabilities $p(\cdot \mid \cdot)$ are all strictly positive rules out such a policy because it makes it impossible to keep track of the date t by encoding it in the state.

If some transitions were allowed to have zero probability, we could construct the state space to be the union of two non-communicating sets of states, which allows that an efficient policy can have different dictators in the two non-communicating sets. Or, it could be possible that the transitions ensure that no state can occur more than once. For example, let $S_t = \{(t-1)S+1, \ldots, Kt\}$ and suppose that that each state in S_t can transition only to a state in S_{t+1}. That allows the state to encode the date and allows efficient policies that increasingly favor more patient agents at later dates. Restricting transition probabilities to be positive both eliminates those possibilities. It ensures the unity of the model (just one set of communicating states) and enforces a perspective that Arrow championed: that the date should be irrelevant for setting or evaluating environmental policy beginning today.

Finally, the theorem also assumes that different agents have different rates of time discount. Omitting that assumption, a similar proof implies that any efficient policy maximizes a weighted sum of utilities of a *dictatorial group* of agents who all have the same discount factor.

Kenneth Arrow told me that this result would be his final theorem, and throughout Arrow's career, he encouraged intelligent debate. On his behalf, I have completed the formalization and proof of his theorem and tried to provide the kind of challenging remarks that he would have welcomed.

References

Jackson, Matthew and Leeat Yariv (2015), "Collective Dynamic Choice: The Necessity of Time Inconsistency," *American Economic Journal: Microeconomics*, 7(4): 150–178. http://dx.doi.org/10.1257/mic.20140161

[2] A similar point was made by Jackson and Yariv (2015), although their formulation and framing of the problem were different.