Institutions for the Future

Ryan Doody

April 26, 2022

Safeguarding Humanity

Ord says that we face unprecedented risks, resulting in three challenges:

1. *Institutional Problem.* We can't rely on our current intuitions and institutions (which have evolved to deal only with small- to medium-scale risks).

2. *No Safety Net.* We cannot afford to fail *even once* (which means that trial-and-error is not a viable strategy).

3. *Epistemic Problem.* How are we to predict, quantify, or understand risks that have never transpired?

To help with the Epistemic Problem, we can fund more research into existential risks. What else can be done?

Question 1: Should we try to colonize other distant planets?

Question 2: Should we try to *slow* the rate of technological progress?

Question 3: Should we try to *speed up* the rate of technological progress?

It's not obvious what our plan should be. But even if it was, there are further questions about the feasibility of implementing it.

International (and Intergenerational) Coordination

Safeguarding humanity's future is an *intergenerational global public good*.

As a result, each nation (and generation) is inadequately incentivized to reduce and avoid existential risk. So, risk-reduction will be undersupplied—unless steps are taken to reform our institutions so that their incentives align with the goal of safeguarding humanity's future.

We need *international* (and *intergenerational*) *coordination*. But how can it be achieved?

Question 4: Should we try to move toward a World Government?

Question 5: Would internationally binding constraints (e.g., international treaties, laws, organizations, etc.) work?

Toby Ord's Grand Strategy for Humanity:

1. Reaching Existential Security
2. The Long Reflection
3. Achieving Our Potential

We are going to focus on what can be done to achieve the first step.

We are in a better position regarding *natural* risks than *anthropomorphic* ones. We can use what we've learned from near misses.

We also must be mindful to account for the risk of error in estimating the risks.

It's non-excludable, non-rival, and its pool of beneficiaries span the entire globe throughout the entire future.

Free-rider and *Assurance* Problems.

Question 6: Is democracy compatible with our obligations to the future of humanity?

Longtermist Institutions

The current practice of most governments are at odds with longtermism. John & MacAskill identify three sources of political short-termism:

1. *Epistemic.* Lack of knowledge about the future; irrational discounting stemming from cognitive bias.

2. *Motivational.* Positive rate of pure time preference; self-interest and partiality; cognitive biases.

3. *Institutional.* Elections motivate politicians to "prioritize policy which results in very near-term, visible benefits for which they can publicly take credit, while hiding or deferring costs"; economically dependent on actors with short-term interests; 24-hour media cycle; political polarization; omnibus bills; collective action problems . . .

To respond to these problems, J& M advocate four reforms:

1. **In-government Research Institutes and Archivists**

    Tasked with producing periodic, public reports that: (1) *chronicle* long-term trends, (2) *summarize* extant research to improve accessibility, (3) *analyze* expected impacts of policy, and (4) *identify* matters of long-term importance.

2. **Futures Assemblies**

    Permanent citizens' assemblies—deliberative bodies of randomly selected citizens who provide non-binding advice—with the explicit mandate to represent the interests of future generations.

3. **Posterity Impact Assessments**

    Require Posterity Impact Statements—an extension of the environmental impact statements already required by many—on all proposed legislation with significant effects occurring beyond 2-4 years.

4. **Legislative Houses for the Future**

    An upper house in the legislative branch devoted exclusively to representing the well-being of future generations. Legislators could be randomly selected from among voting-eligible population (with some seats reserved for experts).

How might these reforms help? Can you think of others? How realistic are they?

Problem for Democracy. Democracy is rule of the people *by* the people. If some matter substantially affects your interests, you should have a voice in how it's decided. Much (if not all) of what we decide to do substantially affects the interests of future people. But future people—in virtue of the fact that they do not yet exist—cannot make their voices heard in how those matters are decided.

Are there other ways to understand *democracy?*
   Typically, we think of *elections* are playing an essential role in a democracy. They are the mechanism by which the voice of the people are heard. But, because future people (by necessity) cannot vote, does this show that elections are the problem (not democracy itself)?

The first three reforms are moderate, "soft-power" reforms: they rely only on informal punitive and reward mechanisms (e.g., public outcry, embarrassment). The fourth suggestion is more extreme.

Examples: Singapore's Centre for Strategic Futures, The U.S.'s Office of Technology Assessment

Examples: Ireland's 100-member Citizens' Assembly, The UK government's Select Committees

Examples: The UK's 2020 Well-being of Future Generations Bill

Examples: Hungary's (2008-2012) and Israel's (2001-2006) Commissioners for Future Generations

Next Time: How can we represent the interests of future people? We will look at two recommendations: *age-weighted* voting and *youth quotas.*
   Both presume that the young are best suited to represent the interests of the future. Is that assumption warranted?