

Anurag Behar: Don't let the 'alignment problem' of our systems fester

Anurag Behar | 4 min read | 05 May 2025, 12:00 PM IST



The human system alignment problem doesn't fit neatly into any single discipline or field of action or responsibility.

SUMMARY

There's a disconnect between the welfare of society and what institutions and individuals do. If we don't fix our systems, they will eventually collapse under the weight of their own contradictions.

In every sphere of human endeavour—be it business, politics, administration, academia or civil society—we operate under a quiet but pervasive assumption: those who rise to the top do so because they are the most competent, the most suited for leadership. This belief is so deeply embedded in our collective consciousness that we seldom pause to question it. Yet, a dispassionate examination reveals that the upper echelons of any field are not exclusively populated by the best and brightest. Instead, they mirror the broader population—some are exceptional, many are

mediocre and a troubling number are woefully inadequate or even harmful. **My last column** was about this.

This phenomenon is not merely a matter of individual shortcomings; it is a systemic failure. At the heart of this dysfunction lies what might be termed the 'alignment problem of human systems' —a chronic disconnect between the broader good and welfare of society on one hand, and the goals, actions and behaviours of individuals as well as institutions on the other.

Political systems, for instance, are designed around the mechanics of election and re-election, not necessarily the demands of competent governance or the long-term health of society. **Markets**, left to their own devices, chase short-term profits at the expense of the environment and society, fuelling climate catastrophe and grotesque inequality.

Civil society, which ought to act as a counterbalance to power, finds itself dependent on funding from the very entities it seeks to hold accountable. Some media organizations, in their pursuit of audience engagement, are seen to sacrifice their fundamental duty to truth and public accountability.

Even academia, in its unreflective pursuit of knowledge, often loses sight of the larger human good it is meant to serve.

None of these systems is entirely broken. There are always exceptions among individuals and institutions that rise above the fray. But the structural drift and problems are unmistakable.

The reasons for this misalignment are manifold. One is the classic principal-agent problem, where those entrusted with authority—such as CEOs, politicians and bureaucrats—end up serving their own interests rather than those they were meant to represent. Another is the perverse incentive structure that rewards behaviours antithetical to the system's intended purpose, such as **healthcare systems** that prioritize billing over patient outcomes.

Then there is the slow erosion of institutional values, where organizations stray over time from their founding principles, succumbing to corruption, sclerosis or cultural decay. And finally, there are coordination failures, where individual rationality leads to collective disaster, as seen in the over-exploitation of shared resources despite wide awareness of the consequences.

What is striking is how little attention this fundamental problem receives. The tech world is consumed by debates over AI alignment—how to ensure artificial intelligence remains tethered to human values—yet we scarcely acknowledge the same existential crisis in our political, economic and social systems.

Part of the reason is that human system alignment is a meta-problem, one that doesn't fit neatly into any single discipline or field of action or responsibility. For example, economists study incentives, political scientists examine governance and sociologists analyze institutions, but few take on the challenge of realigning entire human systems.

Historically, this has not been an existential issue. When societies were smaller and less interconnected, the consequences of misalignment were limited. Today, in a world where every nation, every human field of action and each individual is bound in a tight web of interdependence, the stakes could not be higher. Climate collapse, spiralling inequality and poor and arbitrary governance are not isolated crises—they are symptoms of a deeper, systemic misalignment.

Solutions, though imperfect, do exist. We might begin by redesigning institutions to better align incentives with societal goals. We need tight regulation and high carbon taxes to combat climate change, for example, and must invest in public health systems for healthier societies. Stronger checks and balances—-independent oversight and the separation of powers—could mitigate the principal-agent problem.

Culturally, we need a shift towards long-term thinking, with an intergenerational ethical perspective in policymaking. We also need education systems that foster civic responsibility and ethical leadership. Governance, too, must become more adaptive, incorporating sharper and quicker feedback loops that allow for course correction and decentralization to prevent the concentration of unaccountable power. These are not new ideas. But we will surely take them more seriously if we acknowledge that misalignment is not an anomaly but a fundamental flaw in the architecture of our systems. And then have the humility to accept that no single set of actors, no matter how powerful, can realign these structures on its own. If we don't do this, the systems meant to serve us will continue to fail—until one day, they'll collapse under the weight of their own contradictions, and we with them.

The author is CEO of Azim Premji Foundation.