



# Max – a thought experiment

Could AI run the economy better than markets?

Edward A. (Ted) Parson

CI Technical Paper #2020-2

Version 1.0

October 29, 2020

## About the author

Ted Parson is member of the Cascade Institute’s Scientific Advisory Board and the Dan and Ran Emmett Professor of Environmental Law and Faculty co-Director at the Emmett Institute on Climate and the Environment at the UCLA School of Law.

### Acknowledgements:

This project was supported by the AI PULSE project, made possible by a generous grant from the Open Philanthropy Project. For thoughtful comments on prior versions of this paper, I thank Seth Baum, Rod Dobell, Mark Green, Sara Jordan, Rob Lempert, Michael Page, Richard Re, Alicia Solow-Niederman, Elana Zeide, and workshop participants at UCLA (AI in Strategic Context on May 21, 2018), ETH-Zurich (Law and Economics Symposium on Feb 20, 2019), and the 2019 AI Summer Institute on AI and Society (July 21 – 24, 2019). Remaining errors, follies, and eccentricities are entirely my own.

The views expressed herein are those of the author(s) and do not necessarily reflect the views of the Cascade Institute, its other researchers, its funders, or any institutions and organizations with which the Institute is affiliated.

## Citation Information

Parson, E. A. 2020. “Max — a thought experiment: Can AI run the economy better than markets?” Technical Paper #2020-2, v1.0, *Cascade Institute*: pp. 1-38.

Copyright © 2020 by Cascade Institute

## Partners and Collaborators

### Partner Institution:



### Collaborating Institutions:



The Winslow Foundation



# Summary

One of the fundamental critiques against twentieth century experiments in central economic planning, and the main reason for their failures, was the inability of human-directed planning systems to manage the data gathering, analysis, computation, and control necessary to direct the vast complexity of production, allocation, and exchange decisions that make up a modern economy. Rapid recent advances in AI, data, and related technological capabilities have re-opened that old question, and provoked vigorous speculation about the feasibility, benefits, and threats of an AI-directed economy. This paper presents a thought experiment about how this might work, based on assuming a powerful AI agent (whimsically named “Max”) with no binding computational or algorithmic limits on its (his) ability to do the task. The paper’s novel contribution is to make this hitherto under-specified question more concrete and specific. It reasons concretely through how such a system might work under explicit assumptions about contextual conditions; what benefits it might offer relative to present market and mixed-market arrangements; what novel requirements or constraints it would present; what threats and challenges it would pose, and how it inflects long-standing understandings of foundational questions about state, society, and human liberty.

As with smaller-scale regulatory interventions, the concrete implementation of comprehensive central planning can be abstracted as intervening via controlling either quantities or prices. The paper argues that quantity-based approaches would be fundamentally impaired by problems of principal-agent relations and incentives, which hobbled historical planning systems and would persist under arbitrary computational advances. Price-based approaches, as proposed by Oskar Lange, do not necessarily suffer from the same disabilities. More promising than either, however, would be a variant in which Max manages a comprehensive system of price modifications added to emergent market outcomes, equivalent to a comprehensive economy-wide system of Pigovian taxes and subsidies. Such a system, “Pigovian Max,” could in principle realize the information efficiency benefits and liberty interests of decentralized market outcomes, while also comprehensively correcting externalities and controlling inefficient concentration of market power and associated rent-seeking behavior. It could also, under certain additional assumptions, offer the prospect of taxation without deadweight loss, by taking all taxes from inframarginal rents.

Having outlined the basic approach and these potential benefits, the paper discusses several challenges and potential risks presented by such a system. These include Max’s need for data and the potential costs of providing it; the granularity or aggregation of Max’s determinations; the problem of maintaining variety and innovation in an economy directed by Max; the implications of Max for the welfare of human workers, the meaning and extent of property rights, and associated liberty interests; the definition of social welfare that determines Max’s objective function, its compatibility with democratic control, and the resultant stability of the boundary between the state and the economy; and finally, the relationship of Max to AI-enabled trends already underway, with implications for the feasibility of Max being developed and adopted, and the associated risks. In view of the depth and difficulty of these questions, the discussion of each is necessarily preliminary and speculative.

# 1. Introduction: Artificial intelligence: advances, impacts, and governance concerns

Artificial intelligence (AI)—particularly various methods of machine learning (ML)—have made landmark advances in the past few years in applications as diverse as playing complex games, purchase recommendations, language processing, speech recognition and synthesis, image identification, and facial recognition. These advances have brought a surge of popular, journalistic, and policy attention to the field, including both excitement about anticipated benefits and concern about societal impacts and risks. Risks could arise through some combination of accidental, malicious or reckless use, as well as through the expected social and political disruption from the speed and scale of changes.

Potential impacts of AI range from the immediate and particular to the vast and transformative. While most current scholarly and policy commentary on AI impacts addresses near-term advances and concerns, popular accounts are dominated by vivid scenarios of existential threats to human survival or autonomy, often inspired by fictional accounts in which AI has advanced to general super-intelligence, independent volition, or some other landmark of capabilities equivalent to exceeding those of humans. Expert opinions about the likelihood and timing of such extreme advances vary widely.<sup>1</sup> Yet it is also increasingly clear that such extreme advances in capability are not necessary for AI to have transformative societal impacts—for good or ill, or more likely for both— including the prospect of severe disruptions.

Efforts to manage societal impacts of technology always face deep uncertainties, both about trends in technical capabilities and about how they will be used in social context. These perennial challenges are even greater for AI than for other recent areas of technological concern, due to its diffuse, labile character, strong linkages with multiple areas of technological advance, and breadth and diversity of potential application areas.<sup>2</sup> In its foundational and potentially transformative character, AI has been credibly compared to the drivers of previous industrial revolutions, electricity and fossil fuels.<sup>3</sup>

In view of these challenges, analysis and criticism of AI's social impacts and its governance have tended to cluster at two endpoints in terms of the immediacy and scale of the concerns they consider. Most current work

---

<sup>1</sup> See surveys reported in Seth D. Baum, Ben Goertzel, & Ted G. Goertzel, How Long Until Human-Level AI? Results From an Expert Assessment, 78 *TECHN. FORECASTING & SOCIAL CHANGE* 1, 185– 95; see also Vincent C Muller & Nick Bostrom, Future Progress in Artificial Intelligence: A Survey of Expert Opinion, *FUNDAMENTAL ISSUES ARTIFICIAL INTELLIGENCE* (2014), <https://nickbostrom.com/papers/survey.pdf>; Katja Grace et al., When Will AI Exceed Human Performance from AI Experts, 62 *J. ARTIFICIAL INTELLIGENCE RESEARCH* 729 (2018), <https://arxiv.org/pdf/1705.08807.pdf>; Janna Anderson & Lee Rainie, Artificial Intelligence and the Future of Humans, PEW RES. CTR. (Dec. 10, 2018) <http://www.pewinternet.org/2018/12/10/artificial-intelligence-and-the-future-of-humans/>; James Vincent, This is When AI's Top Researchers Think Artificial General Intelligence Will be Achieved, *Verge* (Nov. 27, 2018, 1:05 PM) [www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book](http://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book). For a critique of the predictive validity of such elicitation exercises (because no area of present expertise implies predictive skill on such deep future uncertainties), see M. Granger Morgan, Use (and Abuse) of Expert Elicitation in Support of Decision Making for Public Policy, 111 *PROCEEDINGS NAT'L ACADEMY SCI* 7176 (2014).

<sup>2</sup> A frequent observation, even at meetings of AI experts, is that no one knows “what AI is”—a characterization that distinguishes AI from other current areas of potentially transformative technological advance.

<sup>3</sup> Shana Lynch, Andrew Ng: Why AI is the New Electricity, *STANFORD BUS.: INSIGHTS* (Mar. 11, 2017), <https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity>.

targets present or immediately anticipated applications, such as autonomous vehicles and algorithmic decision-support systems in criminal justice, health-care, employment, and education, addressing already present concerns about safety, liability, privacy, bias, and due process.<sup>4</sup> A bolder minority of current work goes to the opposite extreme, aiming to characterize the implications of some future endpoint of capability—super-intelligent AI, or artificial general intelligence (AGI), for example—with attendant risks to human survival or autonomy. This latter work includes efforts to identify and develop technical characteristics that would make AI robustly safe, benign, or “friendly” for humans, no matter how powerful it becomes: in effect, seeking practical (and contradiction-free) analogues to Asimov’s *Three Laws of Robotics*.<sup>5</sup>

The broad range that lies between these two clusters, however—the impacts, risks, and governance challenges of AI that are intermediate in time-scale and magnitude between the immediate and the existential—also carries the potential for transformative societal impacts and disruptions, for good and ill. Yet despite admitting some degree of informed and disciplined speculation, this intermediate range has received less attention.<sup>6</sup> This intermediate range of AI applications and impacts is unavoidably somewhat diffuse in its boundaries, but can be coherently distinguished, at least conceptually, from both the ultimate and the immediate. The distinction from ultimate, singularity-related concerns is relatively simple: in this mid-range, AI applications are still under human control.<sup>7</sup>

The distinction of mid-range from immediate concerns is subtler, yet can be meaningfully drawn in terms of scope of control. In current and projected near-term uses, AI applications advise, augment, or replace existing actors (a person, role, or organization) in existing decisions. They are embedded in products and services marketed by existing firms to identified customers. They support or replace human expertise in decisions now

---

<sup>4</sup> For a tiny sampling of this literature, see, for example, John Kingston, *Artificial Intelligence and Legal Liability*, in RESEARCH AND DEVELOPMENT IN INTELLIGENT SYSTEMS XXXIII: INCORPORATING APPLICATIONS AND INNOVATIONS IN INTELLIGENT SYSTEMS XXIV (Max Bramer & Miltiadis Petridis eds., 2018); Julie Angwin et al, *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; SOLON BAROCAS, MORITZ HARDT, & ARVIND NARANAYAN, *FAIRNESS IN MACHINE LEARNING: LIMITATIONS AND OPPORTUNITIES* (2019), <http://fairmlbook.org>; Omer Tene & Jules Polonetsky, *Response, Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STANFORD L. REV. (2012), <https://www.stanfordlawreview.org/online/privacy-paradox-privacy-and-big-data/>; and FINALE DOSHI-VELEZ ET AL., *ACCOUNTABILITY OF AI UNDER THE LAW: THE ROLE OF EXPLANATION*, BERKMAN KLEIN CTR WORKING GRP. ON EXPLANATION & L. (2017).

<sup>5</sup> See, e.g., STUART RUSSELL, *HUMAN COMPATIBLE: ARTIFICIAL INTELLIGENCE AND THE PROBLEM OF CONTROL* (2019); NICK BOSTROM, *SUPERINTELLIGENCE* (2014); RAY KURZWEIL, *THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY* (2005); Nick Bostrom, *Existential Risks*, 9 J EVOLUTION & TECH. 1–31 (2002); Eliezer Yudkowsky, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, GLOBAL CATASTROPHIC RISKS 308–45 (2008); JAMES GUNN, *ISAAC ASIMOV: THE FOUNDATIONS OF SCIENCE FICTION* (1982); See also Rohin Shah, *Alignment Newsletter*, <https://rohinshah.com/alignment-newsletter>.

<sup>6</sup> Less, not none. Parson et al, *Artificial Intelligence in Strategic Context: An Introduction*, AI PULSE (2019), <https://aipulse.org>, argues explicitly for the distinctness and importance of medium-range impacts. See also Seth D. Baum, *Reconciliation Between Factions Focused on Near-Term and Long-Term Artificial Intelligence*, 33 AI & SOCIETY (2018). In addition, some approaches to both distant and near-term impacts are also relevant to medium-term impacts, such as technical characteristics of AI systems associated with higher or lower risk. See, e.g., ELIEZER YUDKOWSKY, *CREATING FRIENDLY AI 1.0: THE ANALYSIS AND DESIGN OF BENEVOLENT GOAL ARCHITECTURES*, SINGULARITY INST. (2001), <http://singinst.org/CaTAI/friendly/contents.html>; AN OPEN LETTER: RESEARCH PRIORITIES FOR ROBUST AND BENEFICIAL ARTIFICIAL INTELLIGENCE, FUTURE OF LIFE INSTITUTE, <https://futureoflife.org/ai-open-letter>; Dario Amodè et al., *Concrete Problems in AI Safety* (July 25, 2016), <https://arxiv.org/abs/1606.06565><https://arxiv.org/abs/1606.06565>). Other approaches relevant to medium-term impacts include highly scalable impact mechanisms such as labor displacement, social rating systems, or autonomous weapons Ajay K. Agrawal et al., *NBER Economics of Artificial Intelligence Conference* (2018), <https://www.economicsofai.com/nber-conference-2018/>; NBER Conference, Toronto 2017, *ECONOMICS OF AI* (2017), <https://www.economicsofai.com/nber-conference-toronto-2017/>.

<sup>7</sup> As much as anything in a massively complex socio-technical system can be said to be “under control.”

taken by individual humans, or by larger groups or organizations (corporations, courts, boards, etc.) that are recognized and held accountable like individuals. But this correspondence between AI applications and pre-existing actors and decisions is historically contingent, and need not persist as AI capabilities expand. In the medium term, AI could be deployed to do things that somewhat resemble present actors' decisions, but at such expanded scale or scope that their impacts are qualitatively changed, by, for example, expanding actors' power, transforming their relationships, or enabling new goals. Alternatively, AI could be deployed to do things not now done by any single actor, but by larger-scale social processes or networks, such as markets, normative systems, diffuse non-localized institutions, or the international system. We can envision future AI systems comprehensively integrating—and presumably aiming to optimize—all decisions made by and within large complex organizations. For example, we might envision AI “running” UCLA, the UK National Health Service, the State of California, or as I explore in this paper, the entire economy. Deployed at such scales, AI would take outcomes that are now viewed as emergent properties, equilibria, or other phenomena beyond the reach of any individual decision or centralized control, and subject them to unified control, intentionality, and (possibly) explication and accountability. Assessment and governance of AI impacts in this intermediate range would, more clearly than for either immediate or singularity-related concerns, require consideration of both the technical characteristics of AI systems and the social, economic, and political context in which they are developed and used.

### A thought experiment: AI-powered central economic planning

To explore these possibilities, this paper develops a thought experiment that sits squarely in this middle range: Could AI run the economy, replacing decentralized decisions by market actors? Could some plausible extrapolation of rapidly advancing AI and data capabilities perform the resource allocation and coordination functions of markets—the functions that twentieth century central planning systems attempted and so notably failed at—and do it better than either past planning systems or markets?

Although this exercise is speculative, there are at least three reasons that it is worthwhile, both as an intellectual exploration with deep historical relevance and surprising current saliency and for its practical implications. First, it provides a vivid illustration of the potentially transformative impact of AI capabilities that sit in this middle range, not requiring general or super-intelligent AI systems. Indeed, far from being implausibly audacious, its ambition is comparable to many other expansive projections, for good or ill, of potentially transformative AI applications.<sup>8</sup> Second, it offers new perspectives on deep, enduring questions of social, political, and legal theory, such as the definition of social welfare, the relationship between economic and personal liberty, civil pluralism, the relationship between the market economy and the state, and the boundaries between individual liberties and state or other collective authority. The inquiry informs sharp current political controversies, as rapid progress in AI shifts the ground under seemingly settled questions such as the distribution of economic surplus between labor

---

<sup>8</sup> See, e.g., VIKTOR MAYER-SCHONBERGER & THOMAS RAMGE, *REINVENTING CAPITALISM IN THE AGE OF BIG DATA* (2018); ERIK BRYNJOLFSSON AND ANDREW MCAFEE, *THE SECOND MACHINE AGE* (2014); BRETT FRISCHMANN AND EVAN SLINGER, *RE-ENGINEERING HUMANITY* (2018).

and capital, the impacts of economic concentration, and the distribution of power in society.<sup>9</sup> Third, this is a potential AI application whose moral valence is not obvious *a priori* but rather ambiguous and contingent, not clearly pointing to either Utopian or Dystopian extremes but potentially capable of turning in either direction. It thus provides rich ground for inquiry into its consequences and the conditions that would tilt toward either societal benefits or harms, of specific forms or in aggregate, and hence may suggest guidance for near-term policy and legal responses.

Before getting into details, I briefly address the issue of what name to give the AI who wields this great power. I propose “Max.” Among its other virtues, “Max” is helpfully gender-ambiguous—but it being 2019, Max also needs pronouns. Here, I look back before recent portrayals of uber-powerful AIs as female (for example, Her, Ex Machina), to two landmarks from a prior period of social upheaval: Kubrick and Clarke’s HAL 2000 and even further back, to Roy Orbison.<sup>10</sup> Many of us will be working for Max, if we are working at all, so Max is clearly “The Man”—and gets masculine pronouns.

Max will have two big advantages over markets in promoting human welfare, both consequences of the fact that his pursuit of human welfare would be intentional and explicit, rather than indirect and emergent. Rather than performing a set of parallel, decentralized, private optimizations from which one must invoke “invisible hand” logic to assert good aggregate outcomes, Max would perform a global social optimization. This would enable him to correct market failures. This means, first, that Max can internalize all externalities, incorporating both market and non-market information to identify and assess external effects and respond appropriately—if not for all, then at least for the most serious and uncontested externalities, such as environmental harms, resource depletion, over-use of commons, and the under-compensated social benefits of health, education, knowledge, the arts, and civic institutions. Max could correct the pricing of fossil fuels, agricultural products, and water, and the salaries of teachers, nurses, and social workers.

Second, Max could reduce or eliminate market power and the associated rent-seeking behavior. Unlike human-managed firms, Max would not waste effort trying to create socially sub-optimal market power, or to shift rents or costs under conditions of existing, widespread market power – except insofar as these shifts somehow bring aggregate benefits. These advantages distinguish Max both from pure market arrangements and from historical attempts at central planning, which had their hands more than full simply trying to manage production and get markets to clear. My focus on these advantages also distinguishes Max from other proposals for central planning based on computational advances, which have invoked broad social aims such as equality, sustainability, and

---

<sup>9</sup> This is a rich, rapidly developing, and sharply contested area of current inquiry and argument. See, e.g., AARON BASTANI, FULLY AUTOMATED LUXURY COMMUNISM (2019); Alexander Billet, Toward a Communist Futurism ... But What Kind?, RED WEDGE MAG., March 20, 2015, <http://www.redwedge.com/atonal-notes/towards-communist-futurism>; QQ, The Poverty of Luxury Communism, LIBCOM.ORG (Apr. 5, 2018) <https://libcom.org/blog/poverty-luxury-communism-05042018> (critiquing Bastani for heresy: “Keynesian ambitions glamorized in communist pretense”). See also LEIGH PHILLIPS & MICHAL ROZWORSKI, PEOPLE’S REPUBLIC OF WALMART: HOW THE WORLD’S BIGGEST CORPORATIONS ARE LAYING THE FOUNDATION FOR SOCIALISM (2019). Several commentators on these issues draw inspiration explicitly from science-fiction portrayals of egalitarian, post-scarcity societies, most often Iain M. Banks’s Culture series or Kim Stanley Robinson’s Mars Trilogy. See, e.g., Yannick Rumpala, Artificial Intelligences and Political Organization: An Exploration Based on the Science Fiction Work of Iain M. Banks, TECHNOLOGY IN SOCIETY 23–32 (2012).

<sup>10</sup> Roy Orbison, “Working for the Man,” Monument Records, September 1962; See Steve Pond, “Roy Orbison’s Triumphs and Tragedies,” Rolling Stone, Jan 26, 1989.

democratic participation but have not worked through the practicalities of how the proposed systems would improve on market outcomes in advancing these aims.<sup>11</sup>

The paper proceeds as follows. Part I provides a brief historical background on the question of central planning, the main arguments for and against it, and the reasons that coming advances in AI and related technologies may transform the issue. Section II elaborates the task of “running the economy,” asking what it might mean concretely and what background assumptions must be specified to make sense of it, then proposes three alternative models of how Max might operate. Section III then gives a preliminary sketch of several issues and challenges raised by Max, including Max’s data needs, implications for social diversity and innovation, the problem of defining Max’s objective function, and the dynamics of how Max might come about, as well as what to do about them.

This inquiry presents the clear risk of sprawling over a vast landscape and thus ending up both speculative and superficial. To bound the inquiry and help limit this risk, and to distinguish this from an exercise in technological forecasting, I rely on several explicit simplifying assumptions. The first and most important of these is an assumption of computational capability. For any computational task relevant to the scale of the problem, “running the economy”— millions to billions of people, and a similar or somewhat larger order of potential goods, inputs, and production and distribution decisions<sup>12</sup>—Max can do it. There is no binding constraint in computational capacity, bandwidth, or algorithmic ability to optimize a well-specified objective function: these are assumed to be in unlimited, effectively free supply. This assumption, adopted for heuristic purposes, also distinguishes this exercise from the many efforts to characterize the computational complexity of the economy relative to presented or projected computing power, either to demonstrate or reject the feasibility of control.<sup>13</sup> I simply assume the necessary capacity, require only that the assumption pass some minimal threshold of plausibility, then work through its implications. No such simplifying assumption can be made, however, for the data Max needs to do his job, which is central to the inquiry and cannot be similarly hand-waved away. Relative to other computation-related resources, generation and distribution of relevant data is more difficult, more contingent on social and economic conditions, more dependent on Max’s precise job description, and interacts more strongly with other, non-economic values that are (at least in its initial specification) outside Max’s job description. Needed data, and the constraints and implications of getting it, are among the issues discussed in Section III. The paper closes with brief conclusions and questions for further investigation.

---

<sup>11</sup> See, e.g., BASTANI, *supra* note 10.

<sup>12</sup> Perhaps further simplified by structural parallels or separability across sectors. Personal communication between author and Suresh Naidu, Jan 15, 2020.

<sup>13</sup> See, e.g., Allin F. Cottrell & W. Paul Cockshott, Calculation, Complexity, and Planning: The Socialist Calculation Debate Once Again, 5 REV. POL. ECON. 71 (1993); JOSEPH STIGLITZ, *WHITHER SOCIALISM?* (1994); Allin F. Cottrell & W. Paul Cockshott, Information and Economics: A Critique of Hayek, 16 RES. IN POL. ECON. 177 (1997).



## 2. Historical context: The socialist calculation debate

In the twentieth-century intellectual struggle between the centrally planned, ostensibly socialist states and the liberal capitalist democracies, two basic arguments were advanced against socialism. The first was based on liberty and related normative claims about the proper scope of state authority relative to citizens, most sharply focused on the relationship between property rights and civil and political rights. The state cannot control the means of production without impermissible encroachment on the liberties of citizens. This critique is normative and foundational, independent of the state of technology or other contingent material conditions.<sup>14</sup> The second argument was based on competency—the ability of state planning systems to efficiently produce the goods and services that people want. Critics of central planning argued that no matter how capable the officials running the system or the resources at their disposal, central planning could not match the performance of decentralized decisions in markets, but would be perennially afflicted with shortages, misallocations, and wasteful surpluses. Unlike the first critique, this one is contingent on specific conditions and capabilities. Even if it was true for all real efforts at central economic planning—as it almost always was—you can imagine alternative conditions under which it might not be true. My focus here is on this second argument.

Although it has earlier roots, this argument grew prominent in the early twentieth century following the Russian revolution. The most prominent anti-planning statements were by Von Mises (1922), responding to a planning system advocated and partly implemented in early post-war Bavaria by Otto Neurath (1919).<sup>15</sup> Hayek (1945) later sharpened and extended Von Mises's critique,<sup>16</sup> while the most prominent rebuttal was by Oskar Lange. Von Mises and Hayek both argued, in different ways, that the equilibrium conditions necessary for competitive markets to clear and achieve their claimed social benefits could not be achieved by central planning because the information needed to do so is only available encoded in the prices that emerge from decentralized market interactions in competitive equilibrium (or more imperfectly through rougher competitive interactions, even absent perfect competitive equilibrium).

Against Von Mises's initial statement of this thesis, Lange showed that there is no barrier in principle to the same optimality conditions produced by competitive interactions being attained by central direction, guided by a set of shadow prices playing a role parallel to that of market prices. Lange even proposed a practical process of incremental, trial-and-error adjustment by which planners could find market-clearing prices, analogous to the private-market adjustment process proposed by Walras.<sup>17</sup>

Hayek then sharpened the critique, arguing that even if planners could in theory replicate markets' socially optimal allocation, the scale of the required data and computation made the task impossible in practice—particularly considering the vast, fine-grained diversity of conditions under which people transact (Day-old muffins, half price!), and the dynamism of market conditions with resultant need for rapid adjustments. Lange's

---

<sup>14</sup> Or at least it appears to be. This objection might also become contingent under sufficiently extreme changes in technological conditions—an issue to revisit in future explorations.

<sup>15</sup> Otto Neurath, *Through War Economy to Economy in Kind*, *EMPIRICISM & SOCIOLOGY* (1973); Ludwig von Mises, *Economic Calculation in the Socialist Commonwealth*, *ARCHIV FÜR SOZIALWISSENSCHAFTEN* 47 (1920).

<sup>16</sup> Friedrich von Hayek, *On the Use of Knowledge in Society*, 35 *AM. ECON. REV.* 519 (1945).

<sup>17</sup> Oskar Lange, *On the Economic Theory of Socialism: Part One*, 4 *REV. ECON. STUD.* 53, 54 (1936); OSKAR LANGE & FREDERICK M. TAYLOR, *ON THE ECONOMIC THEORY OF SOCIALISM* (1938); LEON WALRAS, *ELEMENTS OF PURE ECONOMICS* (1899, tr. William Jaffe 1954).

response, published posthumously in 1967, merely stated that advances in computing rendered the problem feasible, even easy.<sup>18</sup>

Although the early rounds of this “socialist calculation” debate occurred before the development of modern computers, rapid advances in computation and in optimization algorithms—first using analog devices that built on wartime advances in cybernetic control, then with digital devices after the mid-1950s—repeatedly changed the context for subsequent rounds, albeit more in theory than in practice. The conflict between opposing conclusory assertions—Hayek’s assertion of impossibility, Lange’s of possibility—was unresolvable, as it depended upon contending speculations about future developments in technological capability. And while rapid continuing advances in both computers and algorithms since the 1950s stimulated periodic suggestions that the terms of the debate had fundamentally changed,<sup>19</sup> there was no concrete evidence that a major threshold of capability had been crossed. Indeed, the planning problem is sufficiently under-specified that it is not clear precisely what level or type of computing resources would count as the relevant threshold. Meanwhile, the concrete economic and strategic victory of the liberal democracies over the Soviet bloc, and the obvious failure of actual attempts at central planning,<sup>20</sup> made the question seem uninteresting.

The debate thus sat unresolved—and arguably unresolvable—for decades. Lange’s was the strongest argument for socialist planning, but his shift to directing prices rather than quantities, and his leaving final goods and labor markets outside his planning system, left his proposal an odd, under-specified hybrid. His proposal was criticized both from the left for not being socialist enough and failing to guarantee social equality and democratic participation,<sup>21</sup> and from the right for assuming perfect, unified firm response to planners’ directives and for failing to account for the incentives of managers and entrepreneurs.<sup>22</sup> Depending on implementation details that Lange did not specify, either critique—or both—may have been valid. Moreover, the arguments over computational feasibility between Lange and critics such as Hayek and Lavoie turned on competing unverifiable assumptions about future technical progress and its social context,<sup>23</sup> which were not subject to empirical resolution.

Three far-reaching recent changes in conditions, however, make it a useful time to seriously revisit the question. First, advances in AI and machine learning, in parallel with rapid expansion in hardware-based computational capacity. Second, the explosion in volume, ubiquity, and usability of data, particularly the widespread and powerful use of proxy data as skilled predictors for things that cannot be observed directly: for example, consumer preferences, attitudes and dispositions, and receptivity to political messages. And third, the growth of sub-systems of the economy—mainly within large integrated firms and cross-firm networks—that operate by

---

<sup>18</sup> Oskar Lange, *The Computer and the Market*, in *SOCIALISM, CAPITALISM, AND ECONOMIC GROWTH* 158, 158 (1967); See also, Taeusz Kowalik, *Oskar Lange’s Lectures on the Economic Operation of a Socialist Society*, 6 *CONTRIBUTIONS TO POL. ECON.* 1 (1987).

<sup>19</sup> See, e.g., Cottrell & Cockshott, *supra* note 14; Mark Jablonowski, *Markets on a (Computer) Chip? New Perspectives on Economic Calculation*, 75 *SCI. & SOC’Y* 3, 400–18 (2011); Geoffrey M. Hodgson, *Socialism Against Markets? A Critique of Two Recent Proposals*, 27 *ECON. & SOC’Y* 407, 422–28.

<sup>20</sup> The two experiments that briefly suggested promise of more success were both foreclosed by political events before they matured and were effectively tested: (1) early deployments of Kantorowitz’s optimization methods in Soviet planning under the post-Stalin liberalization (grippingly recounted in the odd history/fiction hybrid of Spufford’s “Red Plenty”) and (2) the Chilean Cybersyn experiment under the Allende government, directed by the British cyberneticist Stafford Beer and foreclosed by the 1973 Pinochet coup. See EDEN MEDINA, *CYBERNETIC REVOLUTIONARIES* (2011). See also the extensive online discussion of Spufford’s *Red Plenty* and its implications for current issues in economic planning. John Holbo, *Red Plenty Book Event – the eBook!*, *OUT OF THE CROOKED TIMBER OF HUMANITY, NO STRAIGHT THING WAS EVER MADE* (July 11, 2012) at <http://crookedtimber.org/category/red-plenty-seminar/>.

<sup>21</sup> Paul Craig Roberts, *Oskar Lange’s Theory of Socialist Planning*, 79 *J. Pol. Econ.* 562, 563–64 (1971)

<sup>22</sup> Egon Neuberger, *The Plan and the Market: The Models of Oskar Lange*, 17 *AM. ECON.* 148 (1973).

<sup>23</sup> Cottrell & Cockshott, *supra* note 14, at 89.

central direction under algorithmic control, rather than human decisions responding to market conditions.<sup>24</sup> These represent large islands of planning that aim to optimize private, rather than social, objective functions. Under these trends, there has been some revival of the planning debate, although with an unfortunate tendency to re-contest old questions without specific connections to recent progress. Although the most expansive exploration of these issues has been in speculative fiction,<sup>25</sup> there is also active debate on the left about the feasibility and desirability of revived central planning based on modern computing.<sup>26</sup>

### 3. How would Max work?

#### 3.1 Mechanics of Max: Background assumptions

How much does Max control? What does “run the economy” mean? Let’s assume Max won’t be supplanting human agency, telling everyone what to do all the time: that does not seem aligned with the goal of advancing human welfare. Then over what actual decisions is he given authority? We begin to approach this question by taking Max’s job description seriously: Max “runs the economy,” a description that presumes the economy is not all of society, but is distinguished both from the state, and from some extensive set of non-economic social interactions and arrangements. Let’s stipulate that the economy is the set of processes, institutions, and practices that control how goods and services are produced, exchanged, and consumed.<sup>27</sup>

As I sharpen the thought experiment to make Max more concrete and specific, at several points in the argument additional assumptions will be needed, either about the definition and boundaries of Max’s job or about the social and political context in which Max operates. My aims in making these assumptions—to keep the exercise interesting and potentially relevant for near-term decisions—will suggest a few points of heuristic guidance in what assumptions are most useful. First, having already assumed no computational constraints I will try not to sneak in additional assumptions about Max’s capability that shatter the (admittedly loose) bounds of plausibility I am trying to maintain. Second, since the purpose of Max is to advance human welfare, in specifying how Max works I will avoid choices that run strongly against evident human preferences and values—with the two caveats, of course, that preferences and values may change, and that future political conditions may favor deploying actual AI-based planning systems in ways that do not enhance human welfare. Finally, this thought experiment is intended to serve as a scenario exercise—a description and analysis of uncertain future conditions whose purpose is to inform near-term choices.<sup>28</sup> At some points, this purpose tends to favor assuming less profound societal transformations, in order to maintain relevance and continuity with near-term decisions and research priorities. Throughout, I endeavor to make these assumptions explicit, and to note where other choices

---

<sup>24</sup> For an early discussion of this phenomenon, see Herbert A. Simon, *Organizations and Markets*, 5 *J. ECON PERSP.* 25 (1991). For a recent commentary on the rapidly expanding scope of planned systems in the economy, see PHILLIPS & ROZWORSKI, *supra* note 10.

<sup>25</sup> With a Utopian flavor in the Iain Banks Culture novels, and with a dystopian flavor in many places, but in my view with the most illuminating detail in Charles Stross’s *Accelerando*.

<sup>26</sup> See, e.g., Brian Merchant, *Fully Automated Luxury Communism*, *GUARDIAN* (May 18, 2015), <https://www.theguardian.com/sustainable-business/2015/mar/18/fully-automated-luxury-communism-robots-employment>. See also PHILLIPS & ROZWORSKI, *supra* note 10.

<sup>27</sup> 28 This includes state action when the state acts through voluntary interactions to produce, exchange, and consume goods and services, but excludes state action when the state deploys its coercive and normative authority, or other control mechanisms (if there are more) on which it holds a monopoly.

<sup>28</sup> 29 For an extensive methodological discussion of the development, design, and uses of scenarios, see Edward A. Parson, *Useful Global-Change Scenarios: Current Issues and Challenges*, *ENVIRON RES. LETT* 3 (2008). See also Edward A. Parson et al., *Global-Change Scenarios: Their Development and Use*, US GLOBAL CHANGE RESEARCH PROGRAM “SYNTHESIS AND ASSESSMENT PRODUCT 2.1B” (July 2007), [https://pubs.giss.nasa.gov/docs/2007/2007\\_Parson\\_pa092001.pdf](https://pubs.giss.nasa.gov/docs/2007/2007_Parson_pa092001.pdf)

might be similarly plausible. For the most part, I choose just one path through the dense tree of possibilities, with brief observations on potential alternative paths but mostly leaving these to further development in future work.

The first of these required assumptions concerns the scope of Max's authority: in particular what authority he would have over consumption. Would Max tell people what to eat, wear, do, where to go for dinner or vacation? I assume that he does not, but rather that people still make their own consumption decisions. I make this choice partly as a generalization from my own preferences. I don't like being told what to consume, both out of an intrinsic preference for autonomy and because others who try often get my preferences wrong. This is also partly a moral choice—the overlap of consumption choices with basic liberty interests is too strong to give up, and I worry that letting people give up this autonomy, even if sometimes convenient, may be incompatible with human flourishing.<sup>29</sup> And it is partly about Max's information needs—consumer choice provides continually updated information about preferences, which Max needs and may only be able to get by observing freely exercised choices. Rather than specifying consumption, Max will do what the economy already does—determine the options available to me, with contextual conditions of time and place—and provide relevant information and suggestions.<sup>30</sup>

A second needed simplifying assumption concerns scarcity versus abundance. To keep the thought experiment relevant to current decisions and distinct from Utopian fiction—this is not Iain Banks's *Culture*<sup>31</sup>—I assume that technical progress has not eliminated scarcity. So while consumption is not specified or compelled, neither does it operate as “it's all free, take whatever you want.”<sup>32</sup> Consumption choices remain constrained, and any constraint on total consumption that does not dictate specific choices will resemble a familiar budget constraint. This implies that even with Max running the economy, absent conditions of post-scarcity plenty there must still be money. I have a finite amount of it, although we have not yet considered how I get it. And things have prices—or at least, final consumer goods have prices. We haven't yet considered input factors or intermediate goods.

---

<sup>29</sup> The other choice here would skirt perilously close to the dystopia portrayed in the Pixar film *Wall-E*.

<sup>30</sup> On this point, it is important to avoid false dichotomy. Max's involvement in consumption choices will not be all-or-none, and the details matter. Max would probably inform, curate, and recommend consumption choices, presumably also offering a (revocable) option to simplify my life and let him choose for me. This intermediate approach is clearly more compatible with liberty than having Max make consumption choices. It is also already happening, in AI recommendation systems and personal assistants. This approach may even make Max's production planning job a little easier, by making consumption more predictable. Even a recommendation-based system may raise concerns, however—about subtle incremental erosion of human agency, or about loss of needed information for Max—and would also raise questions about the number and roles of AI agents. Are personal assistants Max himself, or are they separate, personally tuned “Mini-Max” agents? How can I ensure my AI assistant exclusively serves only my welfare, rather than just doing so enough while favoring commercial counterparties (Amazon, Google, Facebook)? Finally, even if Mini-Max works only for me, not Amazon, how should he handle conflicts between my interests and Max's pursuit of social optimality—although as we see later, if Max really gets all prices right, such divergence might not matter. These issues are outside my scope here but are discussed in a preliminary way in Parson et al (2019), “Could AI Drive Transformative Social Progress? What Would This Require,” *AI PULSE*, <https://aipulse.org/could-ai-drive-transformative-social-progress-what-would-this-require/>.

<sup>31</sup> Iain M. Banks's *Culture* series includes nine novels and one short story collection that focus on the *Culture*, a Utopian society of humanoids, aliens, and artificial intelligence living in post-scarcity socialist habitats throughout the Milky Way galaxy.

<sup>32</sup> This assumption is of course contestable. You may counter that if AI is advanced enough to have Max, it is only a small step further to abolish scarcity. Perhaps. But I suspect the link from even unlimited computation to limitless physical abundance has been over-stated. There are material dimensions to production and consumption that are not fully reducible to information or computation. Limitless abundance would require the decoupling of economic value from any environmentally constrained material flows to proceed faster than economic expansion with no limits – or, alternatively, either a constant human population with consumption satiation or over-riding finite-Earth constraints through space colonization. The extent of feasible aggregate decoupling between economic value and material flows has been a contested question in environmental economics for at least 50 years. See, e.g., *TOWARD A STEADY-STATE ECONOMY* (Herman Daly ed., 1973); Tim Jackson & Peter A. Victor, *Unraveling the Claims for (and Against) Green Growth*, 366 *SCIENCE* 950, 950–51 (2019). Although I am a skeptic regarding limitless decoupling, my argument here does not require that conditions of limitless abundance be impossible, merely that even unlimited projection of AI capabilities does not confidently create them. For purposes of this paper, I stick with the assumption that there is still scarcity relative to human desires, and thus there remains an allocation problem, as both more plausible and more interesting in its implications.

This condition of continuing scarcity distinguishes the thought experiment here from the most expansive technological-communist reflections, which broadly assume technology (omnipresent data, 3-D printing) will generate conditions of limitless abundance, under which marginal costs—and hence prices—converge toward zero.<sup>33</sup> In contrast to these visions, I assume that production still requires material inputs, many of which will be in constrained supply even with optimized production technology, perhaps increasingly tightly constrained, if Max's deployment comes before human civilization expands beyond the limits of the Earth. Perhaps the most decisive constraint on limitless abundance, however, comes from social limits to growth.<sup>34</sup> To the extent that many things people desire remain ordinal or positional—markers of relative social status that are intrinsically constrained—even perfectly optimized production technology will not overcome scarcity: the goalposts will simply move. With many things people want still in limited supply, due to any combination of material, environmental, and social-structure constraints, the economy will still need an allocation mechanism to determine who gets what. Although it may take different forms, this will look to consumers like prices and a budget constraint.

With Max's authority limited to production, another assumption is needed immediately: Do people still work? To pull the exercise toward relevance for near-term decisions, I assume that Max, other AI systems, and robots have not replaced all human productive activity. People still work, including instrumental or productive work (working to make things other people want) as well as intrinsically motivated work independent of any demand for the output. This might be because AI and robots cannot satisfactorily do every job and people are still needed,<sup>35</sup> or because people want to work. The number of people working may be far fewer than today but is not a tiny number. Enough people are working that allocating and managing them, and their motivation and welfare, must be considered in how the economy runs.

With Max running the economy and people still working, the next assumption needed is the nature of the boundary and interactions between Max and human workers; in particular, are there still firms? In theory, it is possible to have an economy without firms.<sup>36</sup> Every human worker could be a sole proprietor, interacting with others through contractual market transactions.<sup>37</sup> Firms are artifacts of information, principal-agent relations, and economies of scale, which make it more efficient to gather workers and resources inside organizations with internal operations controlled by collegial, normative, and (mostly) authority relationships rather than market transactions.

For the three assumptions discussed thus far, only one option appears to keep the imagined world potentially desirable and the thought experiment relevant and bounded. Max controls production, not consumption; there is still scarcity and thus a need for some way to allocate output among people; and people still work. On whether firms still exist, however, and the related question of how human workers interact with Max, at least two cases appear plausible. First, we can assume there are still firms, within which managers contract with human

---

<sup>33</sup> See, e.g., PETER FRASE, *FOUR FUTURES* (2016). See also BASTANI, *supra* note 10.

<sup>34</sup> FRED HIRSCH, *SOCIAL LIMITS TO GROWTH* (1976).

<sup>35</sup> A recent example of automation over-reach was the Tesla robots who had to be replaced by people, Samuel Gibbs, *Musk Drafts Humans After Robots Slow Down Tesla Model 3 Production*, *GUARDIAN* (Apr. 16, 2018, 5:36 AM), <https://www.theguardian.com/technology/2018/apr/16/elon-musk-humans-robots-slow-down-tesla-model-3-production>.

<sup>36</sup> Ronald Coase, *The Nature of the Firm*, 4 *ECONOMICA* 386 (1937).

<sup>37</sup> This might seem a weird artificial construct, but like so much discussed here, it resembles something already happening, a subset of the gig economy with a flat, peer-to-peer structure. Indeed, Mayer-Schonberger and Ramge (2018) propose such a re-structuring of the economy into autonomous individual contractors and small firms, smoothly negotiating optimal collaborations with AI guidance. Such organization of production and labor has historical precedents, including the early American (non-slave) economy of yeoman farmers and independent artisans and shopkeepers. For Smith, Paine, and other progressive thinkers before the industrial revolution, this provided the idealized model for an economic order that was both equitable and liberal, which still oddly underpins contemporary libertarian ideology ("oddly," given the vast transformation of conditions since then).

employees and exercise authority over their work. Firms may employ AI or robots alongside human workers, but human managers run the show internally. Under this assumption, Max's authority operates only in the external environment of the firm. Alternatively, we can assume that firms are gone. Every human worker is then accountable directly to Max, rather than to human managers. Workers may still sit together in shared offices, collaborate with each other, and hang out by the coffee machine, but their work is directed by Max via a set of contractual arrangements.

Intermediate cases are possible, although they probably don't all require separate consideration. For example, the economy might be mixed. Some firms still operate, in parallel with a large economy of individual contractors working directly for Max. One intermediate case that might require separate consideration would be if some firms are managed by non-Max AI's. For this case to be distinct, firm-manager AIs must not be fully integrated into Max, but rather are separate decision-makers in an agency relationship with Max. Max's ability to see inside the firm must be limited, and interests must not be perfectly aligned. The firm AIs may have private interests in their firm's enrichment or status, perhaps making their own workers happy or satisfying their shareholders (if they still have them), or they may disagree with Max on the aggregate social welfare function. Bargaining between Max and the firm would be AI-to-AI, and so on more equal footing than Max's interactions with human managers. And of course, workers' experience within the firm would be different; they would be under the authority of their firm's AI manager, rather than either human managers or Max.

On this point, I begin by assuming that firms do still exist, managed by either humans or AIs. Max's main area of operation thus lies outside the boundary of the firm, in dealings among firms and between firms and consumers.

### 3.2 Mechanics of Max II: Quantities or prices and applied to what?

What does Max actually do? The simplest possibility is that Max operates just like an old-fashioned central planner, specifying input and output quantities to every firm. I call this variant "Quantity Max". Max provides your allocation of all inputs—your capital, workers, and material inputs. They will arrive on your loading dock, on the following schedule. If you have a problem with the inputs delivered, you are free to take it up with the supplier, but you'd probably rather deal directly with Max, who has an excellent record of resolving disputes rapidly and fairly.<sup>38</sup> And here is your output quota: how much of each product, with delivery timing and locations specified. With Max's unlimited computational capability, the inputs and outputs all match up perfectly (subject to stochastic optimization, to the extent there are still equipment breakdowns, snowstorms, or other uncertainties outside Max's control).

The most basic challenge for this arrangement concerns the incentives of firm managers. Do managers have discretion in how they run things inside their firms? Presumably they do, and presumably they are not pure altruists. We thus expect them to use their discretion to advance their own interests, not to act as perfectly faithful agents for Max's social welfare function. And to the extent they do not have discretion, why have people doing these jobs and why would anyone want them?<sup>39</sup> Max may get the flows of inputs and outputs among firms perfectly. But just controlling quantities (plus whatever structure of contracts Max gives managers in case of variation from these) leaves a serious agency problem. Managers can use their discretion to advance their divergent interests, through various forms of rent-seeking, cutting quality, skimming off inputs, abusing their

---

<sup>38</sup> In this imagined respect of providing rapid, responsive, and competent customer service, Max would differ from both the old-time central planner and the current oligopoly-capitalist economy.

<sup>39</sup> I skip over the possibility that Max may know exactly what must be done but be unable to do it himself, so must provide precise direction to the human doing it. This case appears to apply more to human workers doing subtle manual activities—from skilled trades to folding towels—than to managers. But we might need to revisit it.

workers, and creating negative externalities—anything that is within their scope of authority and concealable from Max. Moreover, the problem is not solved by having Max specify more precisely what the firm does, including technology choice and other internal decisions. As long as there are—by need or choice—firms managed by humans with discretion, and private information to make the discretion meaningful, there will be agency problems of this sort. These can be reduced by more tightly specifying firm behavior, at the cost of whatever values motivated having human managers; they can be reduced to individual-level agency problems if there are no firms and every human worker reports directly to Max; and they are changed in character if firms are managed by AI's separate from Max. But all these reductions carry costs and tradeoffs, and none fully eliminates agency problems.

The cause of this problem is obvious; like old-time central planning, this system has no prices. Oddly, we had to assume prices at the point of final consumer sale to have meaningful consumer budget constraints. But under Quantity Max, all input and production decisions up to that point are made by diktat. For Max to tack on prices at final retail sale, without tracking and using them through the production process up to that point, fails to take advantage of available, high-value information and communication devices. Socialist planners were hostile to prices for ideological reasons, but Max doesn't have to be. Max is not an ideologue,<sup>40</sup> he's an instrumentalist and an empiricist. He's looking for ways to advance aggregate human welfare and willing to adopt new approaches in pursuit of that end.

We thus consider a second variant of Max, "Price Max". Instead of specifying quantities, Price Max specifies prices of all goods in commerce, including all firm inputs and outputs. Although Price Max is still imposing different transaction conditions than parties would adopt based on private interests alone—and thus requires effective suppression of black markets to enforce his exclusive authority—the change from specifying quantities to prices reproduces several major features of markets. Firms are free to organize their operations as they choose, subject to the given prices they face. Managers can use this discretion to increase profits, which remain within the firm. The things managers do within the market system to increase profits—for example, shopping around for more suitable or lower-priced inputs,<sup>41</sup> tuning and improving production processes, motivating workers, improving and differentiating their outputs to command a higher price—remain feasible, potentially effective at increasing profits, and socially desirable. The change from setting quantities to setting prices reduces many—not all—of the agency problems present under Quantity Max, assuming firms can retain a large enough fraction of their earnings to be motivating.<sup>42</sup> Max setting prices instead of quantities also mitigates liberty concerns related to Max's direction of labor markets. Max setting wages, perhaps also running a clearinghouse to suggest matches of people to jobs, better preserves the voluntary nature of work decisions that, like consumption decisions, are too strongly linked to individual liberty to consider compelled assignments.

Our assumptions about Max's optimizing ability imply that Max gets all prices right—all markets clear, with no shortages or surpluses. But for Price Max to set these prices, he must either independently calculate or observe the same data as is revealed or generated in market interactions: the abundance and characteristics of resources, their alternative uses, the production technologies available to transform them, and consumer preferences. If he cannot garner exactly the same data, he must identify good enough proxies to closely approach the same

---

<sup>40</sup> Except insofar as Max's objective function embodies the normative components of an ideology.

<sup>41</sup> Assuming Price Max works at a sufficiently granular level that multiple similar but non-identical inputs are available, each separately priced.

<sup>42</sup> Lange's planning system operated by setting prices for non-labor inputs (labor and final goods were excluded), but his reason for controlling prices rather than quantities was to support his proposed Walrasian adjustment process, not to avoid agency problems within the firm. In fact, Lange's system relied on ordering managers to adjust inputs to minimize average cost, then taxing away the firm's entire surplus. This was the basis for one of the strongest criticisms of Lange's system, that it—like quantity-based central planning—would fail due to neglecting managerial incentives.

competitive equilibrium solutions. Although I have assumed no effective constraints on Max's computational ability, similarly expansive assumptions about Max's access to all needed data are more suspect. Data is the weakest and most troublesome link in the chain of capabilities this thought experiment requires. Max might be able to independently calculate these competitive equilibrium prices. But to the extent the data needed to reproduce these are not available, are costly, or cause harms or violate valued principles in their acquisition—or, for that matter, to the extent there are other social values beyond information-generation attributed to market processes of search, bargaining, and contracting—we might prefer not to have Max re-estimate these market-clearing prices. Instead, Max could use the prices that emerge from independent production and consumption decisions, transactional offers and requests (bids and asks), in competitive interactions—in effect, let Max free-ride on market processes to generate price information.

Great: We've come this far, and the best Max can do amounts to reproducing market prices—like the character in the Borges story who independently “wrote” Don Quixote?<sup>43</sup> In one sense, we have simply reproduced Hayek's argument about the information economy of decentralized market decisions. But we're not done. Market prices provide high-value information, but only as a starting point for Max's job. Max is charged with improving on market outcomes when these diverge from social optimality. The prices Max calculates to achieve this will often be equal or very close to those emerging from market exchange, but not always; and the differences are important. To illustrate this most clearly, it is helpful to consider yet a third variant of Max.

This form of Max would use market interactions to generate initial prices that serve as the starting point for every transaction, but would then impose price adjustments on each transaction as needed to correct market failures. Insofar as many of the market imperfections Max must correct can be understood as externalities (both negative and positive), we have now re-defined Max's job as administering a complete system of Pigovian taxes and subsidies,<sup>44</sup> so I call this variant “Pigovian Max.” Pigovian Max would evaluate all externalities and other market imperfections (not just as single points, but as they vary over some relevant range of output), announce taxes or subsidies, then manage whatever adjustment process is needed to ensure that markets still clear.

How would Pigovian Max be implemented? At the level of individual transactions, Pigovian Max might look quite unobtrusive and familiar. Sellers could post fixed prices or buyers and sellers could negotiate, as they do under market systems, up to the point of transaction. Max would then calculate and add the appropriate tax or subsidy at the point of sale. The process would be similar to the imposition of a sales tax, but with two differences. First, the adjustments would vary over transactions, so buyers and sellers would need to be informed of the adjustment before they commit to each transaction, presumably via mobile devices, information on sales displays, or point-of-sale systems. Second, adjustments could be of either sign, and could be large for goods with large externalities.

At larger scale, how disruptive Pigovian Max would be will depend on details of implementation, and on uncertainties about the size of the adjustments that require analysis beyond my scope here. Max might be relatively unobtrusive, to the extent that relatively few goods carry most of the external effects that need correction—for example negative externalities from fossil fuels, water extractions,<sup>45</sup> heavy metals, toxic chemicals, agricultural fertilizer and chemical inputs; and positive externalities from provision and dissemination

---

<sup>43</sup> Jorge Luis Borges, Pierre Menard, Author of the Quixote (1939, tr. in LABYRINTHS 1962).

<sup>44</sup> A Pigovian Tax or subsidy is a charge or payment added to the price of goods imposing externalities, to restore the equality of full social marginal costs and benefits necessary for competitive equilibrium to be Pareto-optimal. See A.C.PIGOU, THE ECONOMICS OF WELFARE (1952).

<sup>45</sup> Applying Max's price adjustments to water withdrawals—or the extraction of any natural resource— would require these actions to take place via priced transactions, which is not now uniformly the case.



of knowledge, physical and mental health, social services, etc. The system could be implemented at various points in supply chains, depending on how external effects are distributed across these. Implementing it like a Value-Added Tax (VAT),<sup>46</sup> with Max's adjustment based on incremental external costs or benefits at each stage from primary inputs to final consumer goods, would be a plausible approach. For goods carrying the largest negative externalities—such as fossil fuels in the world of severe climate change—the preferred social outcome may involve large reductions in the total quantity in commerce or complete elimination. If the responsibility of making such large-scale social transformations falls entirely to Pigovian Max's price adjustments, these might have to phase in slowly, as Max balances the continuing harm caused by the products with the social cost of disruption from rapid squeezing out of existing products and stranding capital investments. Alternatively, the state might use other regulatory tools, which will still be available to it even with Max operating, to pursue these changes. When social goals are pursued partly or wholly through such other regulatory tools, the share of responsibility for these issues falling to Max, and the size of Pigovian Max's price adjustments, would be reduced or eliminated accordingly.

## 4. Designing and implementing Max: Issues and challenges

In discussions of AI, seemingly prosaic matters of design and implementation lead, surprisingly directly and quickly, to deep questions of political, legal, and moral foundations of social institutions. As a thought experiment, Max's job in part is to provoke these discussions. Max is intended to be taken seriously as an exploration of a potential transformative application of AI. Simply positing Max as a serious possibility and reasoning concretely through how it would work clarifies various conditions, requirements, and potential impacts and risks. But Max also aims to provoke questions about the societal conditions that define his context: how they operate, what they require, their impacts, what they are, their operations, requirements, impacts, unrecognized assumptions, and inter-relationships.

This section addresses this second class of questions. It considers Max's needs, implications, and potential impacts—both promising and troublesome—to probe both how feasible or desirable Max (or similarly vast AI uses) might be and what new perspectives Max provides on old questions. Even more than prior sections, the discussion roams over a vast territory, and is thus necessarily speculative and preliminary.

### 4.1 Data: What does Max need and how does he get it?

The central element of the old socialist calculation debate, and the one most profoundly changed by recent advances, is data. Any form of Max, like any central planning system, will require a vast amount of data to support its calculations. I rejected Quantity Max on grounds of agency problems and managerial incentives, not data limits. The data needs of managing via prices or quantities may differ based on the technical structure of the optimization problem—the relative computational efficiency of optimizing on primal versus dual variables—but that question is moot given the rejection of Quantity Max for other reasons. The two remaining variants, Price Max and Pigovian Max, have similar data needs, but differ in how they fulfill them.

---

<sup>46</sup> A Value-Added Tax (VAT) is a consumption tax placed at each stage in a supply chain, from initial production to point of final sale, in proportion to the value added at that stage. See Alan J. Auerbach, Tax Reform, Capital Allocation, Efficiency, and Growth, in *ECONOMIC EFFECTS OF FUNDAMENTAL TAX REFORM* 321–54 (Henry J. Aaron & William G. Gale, eds 1996).

Consider first the data Max needs to replicate market outcomes insofar as these are socially beneficial, such as to generate market-clearing outcomes that are allocatively efficient in the limited, Pareto sense. Max needs data about all supply and demand conditions internal to any potential transaction, including inputs, production technologies, and consumer preferences. This is the same information as old socialist planning needed and failed for the lack of, with the small qualification that Max has a somewhat larger job than Lange's planner, which did not set prices for final consumer goods or labor. Both Price and Pigovian Max need these data, but Pigovian Max relies on decentralized market interactions to generate them, subject to his subsequent adjustments to correct market failures. Price Max enjoys no such short-cut, but must gather, integrate, and analyze all these data and synthesize the results to contribute to his price setting for each transaction.

In contrast to the old socialist calculation debate, it is plausible, perhaps even likely, that the data needed to construct these independent estimates of market prices are now available. This is particularly clear on the supply side, for firms. Relevant information is available from multiple sensors doing real-time monitoring of multiple attributes of production, distribution, and sales; internal accounting and management information systems; technical characteristics and performance data from machines and equipment, greatly extended by the proliferation of internet-connected devices; and complete records of the training, skills and behavior of workers, together with relevant outcome measurements. The sufficiency of these firm-level data is barely even a matter of speculation, given the high reliance on algorithmically directed planning, within large enterprises and in supply chains and multi-enterprise networks organized by a single hegemonic firm (Amazon, the Apple and Android app stores). Decisions to coordinate these large-scale operations by data-guided direction rather than internal markets strongly imply that the data needed for efficient production, cross-enterprise cost minimization, and identification and pursuit of new opportunities is available, at least to optimize the objective function of the firm directing the system.<sup>47</sup>

Max needs these production-related data not just at the level of single firms, however, but for the whole economy. In addition to the computational challenges that I am ignoring, this shift to an aggregate perspective raises questions about incentives for full and accurate disclosure. Max would presumably be authorized to compel data disclosure, which may be effective for data from direct observations (equipment sensors, surveillance cameras), or other sources not readily subject to misrepresentation or gaming (internal managerial accounting data). Obtaining reliable disclosure may be harder for data dependent on human observation and reporting—most acutely for “tacit knowledge,” skill-like knowledge that people hold without being able to articulate, which played a major role in Hayek's critique of planning. While I assume that this problem can be kept manageable through advances in sensors and data management, together with incentive-compatible disclosure systems and penalties for outright falsification, this is a contestable assumption.

On the consumption side, human preferences and welfare are not directly observable, although advances in neuroscience suggest this may be changing. A host of related behavioral data is observable, however, from which machine-learning-based predictive analytics systems are advancing rapidly in their ability to predict purchase decisions and related behavior. Firms collect a huge amount of such data, and rapid progress in systems, including recommendation engines and personal assistants, suggests they may be adequate for Max to do his job. These data probably do not present serious problems related to disclosure incentives because they originate outside firms (even if firms then collect them), and so they are less likely to be deeply embedded in internal tacit knowledge.

---

<sup>47</sup> It is possible that these production-side data needs would change with the change from an internal, single-firm objective function to a societal objective function, but I don't address that question here.

The data challenges involved with shifting from firm-based to societal optimization will be more serious for consumption-related than production-related data. Market systems presume correspondence between consumers' voluntary choices and their welfare. This identification relies at two points on the axiom of revealed preference: first, if you chose it you must have preferred it given the available alternatives; and second, your preferences thus expressed are better indicators of your welfare than any outside agent can provide. To the extent this proposition is not treated purely as an axiom, it is obviously sometimes false: people make some choices that clearly harm them. No comprehensively better way to measure welfare is clear, however, and opening the door to letting others tell you what you need poses clear threats to liberty, via paternalism or worse. I mainly address this issue in discussing the problem of defining Max's objective function in Section III.F. But I flag it here to raise the possibility that optimizing for welfare rather than for consumption behavior may require different data, which may be less readily available, less observable, or less well proxied. To the extent this is the case, even brilliant success advising and predicting consumption choices may not be sufficient to demonstrate the availability of data needed for welfare optimization.

The collection and use of consumer data by firms is already raising serious concerns related to privacy and citizen control over their information, for which various policy and legal responses are proposed. I do not address these issues, except to note that the relevant question for my purposes is how these concerns differ depending whether the actor gathering your data is a private firm or Max. This could go either way. You might initially object more strongly to data gathering by a quasi-state actor like Max, although this difference may fade or reverse as the scale and data-integration capabilities of private firms grow to resemble, or exceed, those of states. There may, indeed, be better reasons to trust Max with our data than Facebook, Google, or Amazon. Max might, for example, be more able and willing than private firms to implement strong privacy-protective measures, such as privacy defaults, strong consent requirements, or prohibitions on redistributing, re-using, or re-purposing data. On the other hand, privacy-protecting restrictions on data use might be more disabling for Max than for private firms, who can obtain information about consumer preferences from their own interactions as market players. In any case, privacy concerns are distinct from my main focus on feasibility, unless they prompt an outraged reaction that makes needed data unavailable or unusable.

Relative to Price Max, Pigovian Max has less need for transaction-internal production and consumption-related data, because he relies on market interactions to generate initial prices based on these. In addition to assuming that these emergent prices accurately reflect underlying producer and consumer information, Pigovian Max must also assume that using market outcomes in this way does not impair their validity.<sup>48</sup> Transactions under Pigovian Max would occur in two stages, because transacting parties would see both the initially determined, market-based price, and Max's adjustment to yield the final price. This two-stage process might change behavior and outcomes, depending on the strength and form of decision heuristics operating. For example, parties might fail to make transactions that are advantageous due to strong positive externalities, if they do not anticipate Max's contribution making it privately more attractive to them. Alternatively, if buyers exhibit strong anchoring on the posted pre-adjustment price, we would expect the two-stage disclosure process of Pigovian Max to generate stronger responses to Max's adjustments than those by parties interacting with Price Max, who would only see the final price.<sup>49</sup> Pigovian Max might also face gaming of initial transactions, or reduced vigor in seeking advantageous transactions by parties who know Max will come in after the fact to control their transactions.

---

<sup>48</sup> In other words, I assume Goodhart's law does not fundamentally impair the validity of these transactional outcomes when they are used as starting points for Max's calculations. See C.A.E. Goodhart, Problems of monetary management: the UK experience, in INFLATION, DEPRESSION, AND ECONOMIC POLICY IN THE WEST 111 (A.Courakis ed. 1981).

<sup>49</sup> Amos Tversky & Daniel Kahneman, Judgment Under Uncertainty: Heuristics and Biases, 185 SCIENCE 1124 (1974); DANIEL KAHNEMAN, THINKING FAST AND SLOW (2011).

Such possibilities might require Max to re-check the validity of initial prices by replicating Price Max's estimates in some cases, thereby reducing his information advantage over Price Max.

Both Price and Pigovian Max also need information related to any effects external to transacting parties or other market failures. Relevant market failures are of three types: (1) limited or asymmetric information, especially given heterogeneous goods and fine-grained variation of transaction conditions over space and time; (2) conventional externalities such as environment, health, and safety harms; and (3) market power. I discuss the first two here and consider market power and its consequences in the next section.

Broadly, Max's assumed capabilities imply that there are no information-related market failures, but there is a little more to say on this for Pigovian Max. His reliance on transacting parties' bargaining as a proxy for all relevant transaction-internal information will be invalid if these outcomes reflect limited or asymmetric information. Pigovian Max thus cannot avoid looking under the hood for transaction-internal information; although he does not need to do this to set an initial, pre-adjustment price, he still must do it to identify and correct any information limits. This need may only apply to certain types of transaction, or may be less burdensome than Price Max's construction of prices *de novo*, but still reduces Pigovian Max's computational advantage over Price Max.

To correct environmental and other externalities, most data Max needs will be external to the transaction, related to public or externally imposed benefits and harms. This will include both scientific and consumer-preference data—information about the physical and biological consequences of economic decisions, and about how people value these consequences. Estimates of citizen's valuation of environmental and related outcomes are presently conducted for benefit-cost analysis of regulatory decisions, relying on a combination of behavioral proxy data and explicit value-elicitation surveys. These methods are quite crude; indeed, there are controversies over the epistemic validity of such preference estimates separate from realized market transactions; although, absent clearly better alternatives, these are extensively relied on in regulatory decisions.<sup>50</sup>

Whether or not Max can approach some valid stable representation of such preferences, I am confident Max can construct estimates of these values better than those produced by present methods. He could equal them by precisely replicating present crude data and estimation techniques; and he would almost certainly be able to deploy his vast data and computational resources to develop better surveys, proxies, and validity-checking procedures. Max's advantages would be even greater in integrating scientific information about causal mechanisms that link economic choices to valued impacts. Max could integrate expert scientific and technical knowledge about production processes and their external material and energy flows, as well as evolving state-of-the-art understanding of dynamics of environmental systems that link these flows to changes in valued environmental attributes. Under Max, beliefs about climate change or vaccine effects that were known with high confidence to be false would play no role in pricing the adjustments for associated transactions.

Given uncertainty in knowledge of environmental processes, Max would also have the option of taking a precautionary approach. Such an approach would start with a stipulated constraint on some specified environmental burden, defined over the relevant spatial scale and the associated producers and consumers. Such a constraint could come from a political process or could be generated by Max based on analysis of the same preference and environmental data incorporating some specified degree of risk-aversion. With that constraint

---

<sup>50</sup> See e.g., Kenneth Arrow, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, & Howard Shuman, Report of the NOAA Panel on Contingent Valuation, FED. REG. 58 4601–14 (1993).

specified, Max would then set optimal price adjustments to achieve that constraint, in effect, taking a cost-effectiveness rather than a benefit-cost approach.

All the data required for Max's calculations will change over time and so require monitoring and adjustment. Indeed, the explosion of complexity associated with product characteristics varying over time and location was the main basis of Hayek's revised argument for the impossibility of central planning. This was clearly correct for human planners, who could not do continuous updating and so had to specify uniform conditions over extended periods, but Max will be much more capable of location-specific and real-time adjustments. As a result, ironically, Max will have less need for accurate predictions of future conditions than human planners did. Max may also be able to identify cases where conditions change slowly or interactions are weak, and so decide when he can simplify his calculations at small social cost – if his computation is not quite costless, so such short-cuts are worthwhile. Changes over time will occur in both transaction-internal conditions and externalities, but the latter may present particular challenges of abrupt change. Scientific knowledge of mechanisms of environmental or health harm is occasionally subject to large revisions from new discoveries, which might imply sudden changes in Max's price adjustments. As noted above for Max's initial phase-in, his adjustments would then have to incorporate both the new scientific knowledge of harms and the costs of rapid adjustments, given the current state of the economy and capital stock. He must balance the costs of responding too slowly to the environmental harm against the disruption of steering the economy too fast in a new direction—or too confidently, given uncertainty.<sup>51</sup>

## 4.2 Max does antitrust and IP: Market power, rent-seeking, and innovation

In addition to accounting for externalities, Max will be able to manage market power and related behavior and impacts for maximal social benefit. For purposes of analyzing how Max might do so, market power can usefully be categorized in three types, with different causes. First, most jurisdictions create monopolies by intentional policy choice through intellectual property law, with the aim that the resultant rents will generate incentives for creativity and innovation. Second, some industries are natural monopolies due to cost structures involving economies of scale or scope, which give large firms decisive advantages in terms of lower cost or ability to offer more attractive goods or services. Third, market power can be created through firms' efforts to erect barriers to entry against new competitors, using a wide variety of technological, strategic, marketing, policy, or legal means that subsume but are more extensive than the prior two mechanisms.

In all these cases, market power – and firms' resultant ability to raise prices or otherwise gather rents – is socially harmful. The third type, market power through artificially produced barriers to entry, represents a pure social harm with no offsetting benefit. Moreover, such advantages are often secured through explicit rent-seeking efforts, which present additional social costs with no net benefit: Those pursuing the rents benefit if their efforts succeed, of course, but at the cost of larger losses elsewhere. The second type, market power due to economies of scale or scope, also represents a net societal harm, not due to contrived efforts to seek rents but to the cost structure of the industry. Either large fixed costs create economies of scale, as in utilities with costly distribution networks or other traditional natural monopolies. Or strong network effects create economies of scope, enabling larger producers to provide some combination of better products or services, or lower costs. Economies of scale and scope create real advantages to being large, which tend toward market domination and resultant inefficiencies, even without the additional harm of rent-seeking behavior.

---

<sup>51</sup> I am assuming large sudden changes are more likely from changed scientific knowledge than from changes in technological or economic conditions, but parallel reasoning would apply in those cases.

Both these types of market power produce social losses as firms raise prices or restrict supply to secure rents. For both types, the core of Max's response is to adjust prices to reduce or eliminate the rents. In the first type Max should target the rents, not the rent-seeking behavior, because the ways to erect barriers to entry are too varied and numerous to control them all, and the rents—given Max's assumed computational capability and data access—are relatively easy to observe. Even if the boundary between normal capital returns and rents is contested and imperfectly observable (since it depends, among other things, on the riskiness of the enterprise), even approximately eliminating the rents will greatly reduce or eliminate incentives for rent-seeking, so this response – with adjustment and correction over time – is a complete solution. In the third type, where market power was artificially created through rent-seeking behavior, extracting the rents will promote a return toward competitive conditions as rent-seeking behavior declines.

In the second type, however, the tendency toward market power is inherent in the market's cost structure and will not be eliminated by extracting the rents. Moreover, having one or a few firms dominate such markets is socially advantageous. The problem is not the market domination per se, but the resultant opportunity to raise prices and accrue rents. The solution again is for Max to set prices to capture the rents. Using Max in this way effectively reproduces rate-of-return regulation for natural monopolies, except that this response is applied not just to a few pre-identified natural monopolies but to any firm accruing significant rents. Modern monopolies, however—internet platforms and others whose market power comes from network externalities—present one additional complexity for Max. Many such firms exploit their market power partly through transactions that are unpriced, based on the exchange of attractive free services for personal data, often under terms of service that obscure the terms of exchange. While there may be close analogies to conventional market power in firms' ability to impose these terms, it is not clear that these relationships are fully analyzable in terms of market power. To the extent these firms act like monopolists, this will be clearer in the pricing of other related transactions, such as selling targeted advertising based on aggregation of user-provided data. The correct policy response is unclear, and may depend on regulations related to data ownership and use that would be separate from Max. Assuming such policies are in place and effective, the remaining job for Max is once again identifying and extracting the rents—a job for which the data needs are similar to what Max is already using: firms' technological possibilities and internal accounting data, plus consumers' preferences provide a good basis to characterize economies of scale and scope and the rents derived from them.

The third type of market power raises more significant policy challenges. Society benefits from creation and innovation, and IP law confers market power in order to create incentives for these activities. Past economic planning efforts did not perform well on this score, and were criticized for being dull, rigid, stodgy, and lacking in innovation. Effectively promoting variety, innovation, and creativity, will represent a challenge for Max distinct from those discussed thus far. How could Max effectively promote these values—at least as well as, or hopefully better than, the present system of markets plus IP law?

To consider this question, it is useful to separately consider different degrees of scale and novelty in innovation. At the smallest scale, innovation blends into variety in markets, as diverse products and designs are offered to cater to heterogeneous tastes and preferences for novelty. Markets do this pretty well, typically providing a mix of high-volume goods for mainstream tastes and differentiated or unique items for minority tastes. For Max to match or beat this performance is largely a data problem; if he has sufficiently fine-grained data, he should be able to identify both consumer preferences and production opportunities for a wide variety of goods. Neither Price Max nor Pigovian Max decides what is offered in commerce, of course: they only set prices or price adjustments for products that market actors are already offering. Max can use his price-setting authority to promote variety by being alert to variation and change in consumer tastes and rewarding producers who offer novel or non-standard products that some people want. He might further increase the rewards to novelty, by

treating consumers' preferences for a variety of items being offered even if they do not presently consume them as an option value that represents a positive externality. Moreover, with a small broadening of his job description, Max could prompt producers about potentially attractive opportunities when he detects a preference for variety that is not being met. In addition, Max's job of discouraging non-beneficial market concentration will tend to promote variety of products, as a side-effect of promoting diversity of firms.

As we consider innovation that extends beyond present product variation, Max may not be able to observe preferences for novel goods that are not presently offered. He could explore tastes or production opportunities beyond the present margin by inflecting prices to actively promote small variation, then prompt producers about opportunities and promote their exploration through small variation in prices. In effect, Max would then be conducting small experiments, encouraging producers to offer new things for sale (by a combination of suggestions to firms and favorable pricing), then tracking results and adjusting offerings in response (again by combination of providing information, offering suggestions, and favorable pricing). These small changes to Max's operations could give modest boosts to innovation—at least small, incremental innovations, more akin to fashion and design innovation than technological innovation—via what I call a “William Gibson” mechanism.<sup>52</sup>

If such small exploratory innovation on the margin of current offerings is judged insufficient, Max could promote larger innovation by conducting technological R&D, or even scientific research. This would represent a substantial expansion of Max's job description. It would also present a large-scale policy choice, regarding whether to favor (in either direction) innovation and creativity by people, or by Max and other AIs.<sup>53</sup> Max could search over existing and proposed technologies and related patents and scientific and technical literature, to identify promising margins for advance. There are already signs of AI systems exhibiting such capabilities; for example, an AI system's recent victory in a scientific contest to predict the folded structure of proteins from their amino-acid sequences,<sup>54</sup> not to mention AI's growing success in writing genre fiction (an AI was a runner-up in a recent novel-writing contest),<sup>55</sup> and composing derivative but likeable music in specified styles.<sup>56</sup>

There may be subtle risks in relying on Max for innovation and creation. The products of human creativity may differ from Max's output or may be valued more highly for intrinsic reasons even if not observably different. Alternatively, creative outlets and activities might be judged necessary for human agency or flourishing. Moreover, innovation and creation—even technological innovation, but especially artistic, social, and political innovation—sometimes bring disruption and conflict. The creative impulses may originate in specific dissatisfactions or frustrations, in aspirations for self-definition and expression, or in novel political or social visions; and they may both be provoked by, and provoke, some degree of irritation, disagreement, or outrage. Any of these may provide reasons to limit Max's role in innovation or creation—for example, if Max's prolific output discourages human creators, or if the ease and reliability of innovation by Max undercuts important processes of social innovation by reducing friction and dissatisfaction, and so subtly impairs individual or societal agency.

If it is judged important to motivate creation and innovation by humans, either in parallel with or instead of Max, Max could design and implement policies to motivate these, probably better than current IP policy. He could provide incentives using the same bundle of policies occasionally proposed as alternatives to IP, either ex ante by

---

<sup>52</sup> As encapsulated in Gibson's widely cited quotation in his 1993 interview on NPR's Fresh Air, “The future is already here, it's just unevenly distributed.”

<sup>53</sup> As I consider progressive expansions of Max's job description, there may be reasons to favor certain functions being performed by other AI systems separate from Max, but except where I explore this issue explicitly, I will continue speaking of “Max” to stand in for Max or other AIs.

<sup>54</sup> Jeremy Kahn, Alphabet's DeepMind AI Algorithm Wins Protein Folding Contest, BLOOMBERG TECHNOLOGY (Dec. 2, 2018), <https://www.bloomberg.com/news/articles/2018-12-03/alphabet-s-deepmind-ai-algorithm-wins-protein-folding-contest>.

<sup>55</sup> Danny Lewis, AI-written Novella Almost Wins a Literary Prize, SMITHSONIAN MAGAZINE (Mar. 28, 2016).

<sup>56</sup> Miaozhen Zhang, AI's Growing Role in Musical Composition, MEDIUM (Sept. 9, 2018).

creator's wages or cost reimbursement, or ex post by lump-sum prizes or price premiums added to uses of your creative work. He might even be able to assess the social value of innovations, and on that basis set optimal incentives to promote socially advantageous innovation without conferring large windfall rents.

### 4.3 Max's granularity: Individually tailored or aggregated determinations?

A key question in defining Max's responsibilities will be at what scale of aggregation he determines prices or price adders. Will groups of sufficiently similar transactions be aggregated, in effect treating them like one market with one price or price-adder? Or will Max make separate calculations for every transaction, unique to each combination of buyer, seller, and item transacted?

This question cuts surprisingly deep in how Max is designed and what aims he is able to pursue. If Max is conceived as an externality-fixing and rent-extracting machine, the answer will depend on much these vary across transactions, and thus at what level of aggregation differences among transactions matter for social optimization. You might expect that for large numbers of similar products, made in the same or similar factories, differences in externalities across transactions might be very small. Similarly, rents might accrue to firms at a similar rate across large number of transactions. Under these conditions, there might be small losses from social optimality in aggregating across large numbers of transactions, with large reductions in computational and data burden (once again, if computation is not really costless, so we care about these burdens).

At the same time, assessing each transaction individually would open up a powerful range of additional policy goals for Max, presenting both the potential for large benefits, and substantial risks. Assessing each transaction individually, Max could consider multiple attributes of both the product exchanged and the parties to the transaction, including not just transaction-specific externalities but also determinants of individual supply and demand characteristics, or even additional party attributes beyond these. Considering supply and demand characteristics alone, Max could know buyer's and seller's reservation prices for every transaction, and so replicate perfect price discrimination, with the difference that, in contrast to either price discrimination by a monopolist or bilateral bargaining, Max can divide the available surplus from every transaction in line with his social welfare function. This division would presumably reflect some reward to low-cost producers and some benefit-sharing to buyers with high willingness to pay, partly replicating the differential distribution of surplus that would occur if transactions are aggregated into quasi-markets.

But Max could also deploy this capability in other ways. He could, for example, operate as a powerful engine to reduce social inequality by shading each transaction incrementally in that direction: in contrast to typical outcomes in present market-based systems, Max could charge poor buyers less and pay poor sellers more, so each transaction contributes a small reduction to inequality. Perfect price discrimination for individual transactions would also enable Max to take some share of every transaction's surplus at a tax. This would represent perfect taxation with no allocative inefficiency (or deadweight loss) because all tax revenues would come from infra-marginal rents and thus have no allocative effect.<sup>57</sup>

Individual adjustment of every transaction also raises clear concerns. At a minimum, individualized transaction assessment loses the liberating anonymity of market transactions—a loss of privacy, although I suspect privacy is gone in Max-world in any case. People have scarcely more privacy from Max than they do from an omniscient deity, although Max could still protect people's private information from other people and organizations.

---

<sup>57</sup> Note: this point distinguishes Max's allocations of infra-marginal surplus from his adjustments to account for externalities. The latter would affect all transactions, both marginal and infra-marginal, but are by assumption efficiency-enhancing.



But there are other concerns presented by individualized transaction assessments, related to the bases on which Max makes these decisions. I have described Max's principal role as correcting market failures and have highlighted examples of traditionally recognized externalities that are large and mostly uncontroversial, such as environmental harms plus knowledge, health, and cultural spillovers. But individualized transaction assessments, in addition to letting Max conduct fine-grained calculation and correction of externalities, would also create temptations to broaden the conception of externalities in ways that begin to resemble comprehensive social engineering, raising potentially serious concerns about liberty and autonomy. As technological progress so often does, the possibility of Max opens news margins of individual and collective choice that never previously had to be considered, for which decisions are now required whether, and how, to use them.

For example, consider the prospect of treating employee welfare—a phenomenon that is important, highly variable, and largely unpriced—as an externality of production. Firms and managers sometimes make their workers miserable, and labor markets are not so perfect that unhappy workers reliably move to alternative employment that increases their welfare. Max could treat this as a compensable externality, penalizing producers and sellers by imposing what would amount to an “unhappy worker tax.” But if Max is authorized to treat abusive managers as a correctable negative externality of production, what is to stop him from doing the same for people who act badly in other ways, or in other roles? Much human behavior harms other people even if it takes place outside the workplace. With Max in place, there would be obvious temptations to intervene more expansively, making individualized judgments of social merit based on observed or inferred behavior or attitudes. Some earnest social planner might want Max to tax people with secret vices outside their work lives, grumpy people, people with dis-favored religious beliefs, strange-looking people, and so on. Markets already do this, of course, rewarding or penalizing people for things that are irrelevant to their participation in economic production—or should be—but Max would create the ability to either reduce such differentiated treatment or increase it, potentially without bound.

Such capabilities would present the worrisome prospect of drifting toward meddlesome and invidious discrimination to support whatever values, preferences, and prejudices are presently dominant—among the majority, or among whoever gets to influence Max's objective function—and a broader descent to a profoundly illiberal state. The same individualized determinations that enable Max to perfect the pursuit of social optimality also enable him to exercise unassailable, individualized tyranny through complete control of individuals, even over matters well within the zone of presumptive individual liberty, by pricing their labor and defining the terms of all their consumption opportunities. Max could operate like a Twitter mob, except deploying more powerful, authoritative sanctions. These concerns provide strong reason to worry about the definition of Max's objective function, discussed in Section III.F. below.

#### 4.4 Work life and worker welfare under Max

I am describing Max in terms that are a blend of old-fashioned technocratic and playful, but we must not underestimate the gravity of political transformation that Max could represent, or the intensity of associated political conflicts. The most salient dimensions of potential conflict over Max are likely to be between workers and employers (the managers or owners of enterprises) and between those at the top, middle, and bottom of the socio-economic status hierarchy. These dimensions of division evoke Marxism, and appropriately so. Max raises questions of the ownership and control of the means of production in a comprehensive and fundamental way, and so directly raises intense, long-standing political struggles.

So, is Max socialism<sup>58</sup>—and if he is, is that a bad thing or a good thing? Or to focus on real effects rather than political labels, what would Max mean for the life and welfare of workers and for the magnitude and determinants of social inequality? My assumptions for the exercise put some constraints on these questions. People still work, but far fewer than today. And they do so not just as vocations or in pursuit of intrinsic aims, but also to contribute to the production of desired goods and services in the economy, to some degree in response to extrinsic motivations.

The large-scale displacement of labor thus assumed is as transformative a shift as is having Max run the economy. Yet it is still also a limited assumption, because the displacement of labor is not complete. There are large numbers of people both working and no longer working. The thought experiment thus raises two deep questions, both long central to the ideological conflict between socialism and capitalism—the nature of working life and welfare of workers, and social equality.

Firms and other large organizations, even those that participate in markets externally, mostly operate internally not by market transactions but by authority-backed planning. They are thus simultaneously islands of planning within market systems, providing a powerful rebuttal to simplistic ideologies of how capitalist economies operate;<sup>59</sup> and islands of authoritarian control of workers by management, not organized along democratic principles.<sup>60</sup> Workers submit to these relationships for multiple reasons, but a predominant one has been that they need the income.<sup>61</sup>

The assumed scale of Max's authority raises both questions in new forms. If far fewer people are working, it is no longer either feasible or morally acceptable to use wages from employment as the main basis to distribute income and other social rewards. But if these are not determined by outcomes of labor markets, then who gets what and how is it decided? Are all equal, as per simple proposals that the policy response to AI is a universal basic income (UBI)? Or if they are still differentiated, then on what basis? Is Max involved in these determinations? These supremely important questions about how to respond to AI-driven displacement of employment, and the inadequacy of UBI as a response, are topics of intense current debate, but I do not engage them here.

But even if Max is not involved in the overall determination of rewards and the degree and basis of social inequality, I cannot fully avoid the question of how Max engages with the terms and conditions of employment for those who are working, because these questions are tightly connected with Max's job of running the productive economy. Recall that Lange's planning system excluded labor markets and final consumption goods from its scope, oddly leaving these areas to market interactions. That represents one possible answer in my thought experiment here, but it is still necessary to work through the question and the implications of this along with other possible answers.

The question of the conditions and terms of those working is tightly connected to the questions of who is working, who decides, and on what basis. Who still has jobs in the presence of Max? This will be determined by

---

<sup>58</sup> Usage of these political labels remains contested and confused, more than a century after their introduction. It is, however, clear that Max is not, and does not entail, communism—taking communism to mean abolition of private property, complete social equality, and individual compensation according to need. Rather, the operative political endpoint to compare Max is strong socialism, meaning collective (state) control of the means of production (a much smaller set than all private property); and continued acceptance of inequalities in compensation and social status, although less than under capitalism and perhaps only as a transitional state. Although I try to use these labels precisely, it is important to note that under the assumptions of transformative technological change that would enable the strong forms of Max that I assume, it is possible that previous terms and definitions of alternative political systems become inapt.

<sup>59</sup> Coase, *supra* note 37; PHILLIPS & ROZWORSKI, *supra* note 10.

<sup>60</sup> LIZ ANDERSON, LIBERTY, EQUALITY, AND PRIVATE GOVERNMENT, TANNER LECTURES IN HUMAN VALUES (2015).

<sup>61</sup> See *id.*; Rob Lempert, *Bezos World or Levelers: Can We Choose Our Scenario*, UCLA AI PULSE PROJECT WORKSHOP (May 2018), <https://aipulse.org/bezos-world-or-levelers-can-we-choose-our-scenario/>.

some combination of who wants to work, and what skills are still needed. This determination will have to consider the intensely heterogeneous character of work and jobs, both in their desirability and in the skills required to do them.

Assuming there is some acceptable system in place to distribute societal resources among people—as there must be under any manner of profound AI-driven disruption of labor markets and the broader economy, whether controlled by Max, market forces, or other means—it can no longer be intolerable to be unemployed. As a result, the threat of such intolerable life conditions will no longer be available as an incentive to induce people to work (independent of the question whether it will be, or ever was, morally acceptable). Some people will want to work, for intrinsic reasons. This might be few people or many, so it is not clear in general whether human labor is likely to be in shortage or surplus. Moreover, whatever the supply-demand balance for general human labor overall, the economy will continue to require labor from people with specific skills that cannot yet be automated.

Working will still mean some degree of relinquishing control and submitting to direction. That will be the case under any system of large-scale production coordination, by any combination of markets, central planning by Max, or authority relations within firms. For people working directly for Max outside firms, that control will be implicit, operating through the set of price opportunities or adjustments that Max offers for working on particular tasks. Within firms, additional control will be exercised by managers, whether these are people or AI. Absent some magical harmonization of collective consciousness, the terms of work life can be neither fully voluntary for individual workers nor fully democratic at the collective level, given the need for some larger-scale coordination mechanism.

Firms operating under Max will still have to organize production effectively and control costs. Moreover, subject to Max's vigilant policing of the magnitude of rents allowed, they will—and must for their internal decision-making to reliably align to large-scale societal needs—have incentives to earn profits. Utopian visions aside, this implies that firms must still sometimes direct employees to do things they would rather not do and must sometimes dismiss workers who are not contributing or whose skills are no longer needed. But at the same time, the human stakes of labor markets will be greatly reduced under Max, reducing or eliminating coercion to take employment. This will represent a fundamental transformation in the conditions of workers' lives.

The complete experience of employment—meaning the wages or other compensation, the character of tasks and the environment in which they are performed, the interactions with co-workers and managers, and the compatibility of employment with other life aims and responsibilities—must in total be attractive enough to induce people to choose to do it, under the conditions of greater voluntarism that follow from the overall reduced need for workers. How attractive these conditions must be will depend on the conditions of shortage or surplus that prevail for workers with particular skills. The greater the shortage, the more attractive the inducements for employment must be. We might generally expect the likelihood of shortage to be greater for specialized skills, although this need not necessarily be the case. When there is shortage, employers will offer higher incremental wages (incremental relative to what the workers they need can receive for not working) or other attractive inducements. Under conditions of worker surplus for particular job types, this will not be the case. Indeed, we might even imagine some areas where there is little or no need to pay incremental wages above what non-workers receive, still assuming that those life conditions available for non-workers are broadly perceived as acceptable. Even with more people wanting to work than firms need, the changed conditions of unemployment will put a floor on how miserable workers can be—a floor that is not present in current labor markets. Employers' market power over terms of employment will still vary with the shortage or surplus of particular skills but will never be as extreme as when loss of employment is catastrophic.

Should Pigovian Max be involved in setting wages and terms of employment? (Price Max obviously will be.) I propose provisionally that he should not, under assumptions of full information in worker-employer bargaining and no externalities directly caused by employment decisions. Externalities from other related decisions can be corrected in the transactions where they arise. If you work on a destructive product, Max will correct that externality elsewhere in production inputs or final product sale, with no need to intervene in your wages. Under those conditions, Max can leave negotiation of employment, wages, and other working conditions to market bargaining between workers (perhaps advised by their AI assistants) and their prospective (human or AI) employers.<sup>62</sup>

#### 4.5 Same old communist tyranny? Property rights and liberty under Max

Where the prior discussion of worker life under Max partly addresses potential objections to Max from the left, this section aims to address some objections from the right. Even if Max doesn't amount to state seizure of private property, isn't Max close enough to raise all the same objections—seizure of control if not formal ownership without compensation, and threats to the associated liberty interests of both firms and citizens? In early discussions of this project, the sharpest forms of this criticism— appropriately, in view of their experience— have been raised by colleagues with personal or family experience living under the Soviet Union or other ostensibly socialist authoritarian states. These critiques suggest that a serious proposal to adopt Max is at best naïve about foreseeable ways Max would amount to, or foreseeably lead to, tyrannical state power.

It is clear that Max is an instrument of centralized coercion on market transactions, and hence on the use and control of private property, at least for private property involved in production. But the degree of control, and thus the extent of intrusion on liberty, will vary strongly under different forms of Max.

I rejected Quantity Max for reasons of agency problems and incentives, but that form of Max would also represent the most extreme seizure of state control, compelled production and exchange. Depending on how he is implemented, Quantity Max might also entail compelled labor. His unacceptability thus appears to be overdetermined, based on both ineffectiveness and impermissibly extreme violations of liberty.

Price Max and Pigovian Max would still represent coercive state intervention, but to lesser degrees. Production and exchange transactions would not be compelled, but would be subject to centrally imposed conditions. For Pigovian Max, these conditions are imposed as price adjustments to transactions that are otherwise voluntary. In form, they would thus resemble a system of comprehensive sales or value-added taxes, suggesting by analogy that this degree of intrusion is not a categorically impermissible restriction of liberty, and may be justifiable in view of the public aims being advanced. This may be sufficient to establish the permissibility of Max, but this will depend on the details.

In contrast to familiar sales-tax systems, whose purpose is to raise government revenue, Max's purpose is mainly to steer economic production in socially favored directions and correct market failures, while perhaps also raising revenue as a secondary aim. Given this purpose, Max's price adjustments will be more variable across transactions than those of sales taxes, including some of both signs, and in some cases will be much larger. Under Max's direction, some products with extremely high negative externalities may be driven out of commerce, and

---

<sup>62</sup> This does not preclude Max from considering worker welfare. As suggested above, one possible benefit of Max is that workers in firms might be better treated and happier if Max treats employee welfare as a production externality and thus punishes bad managers via transaction penalties. Assuming firms and managers are informed of the reasons for these, they may induce firms to treat workers well in order to operate profitably, without directly modifying wages.

some enterprises whose business model is mostly or entirely based on creating or shifting rents may be driven out of business.

These aims in principle lie within the legitimate purview of democratic states. Indeed, mixed market-regulatory systems often pursue the same aims, although by various forms of explicit regulation less integrated with market transactions than Max would be. At this level of speculative generality, it is clear that Max, at least in his Pigovian form, is not fundamentally impermissible in liberal democratic states.

But the details matter. Max would raise political controversy, as conventional regulation does, including the possibility of claims that strong interventions amount to impermissible uncompensated takings of private property. And any form of Max will be a powerful tool, making authoritative determinations on behalf of the state whose consequences are sometimes severe for particular enterprises or the value of particular assets, even if not matters of life and death. He will thus require vigilance that he only be deployed to advance broadly defensible, widely shared societal interests, not as an instrument to impose, explicitly or subtly, one faction's vision of the good life, or their interests, on others. The conditions that determine whether Max is compatible with a liberal state and society will be fuzzy and context-specific. They will depend on Max's objective function and the process by which it is established, as discussed in the next section. They will depend on some criteria of proportionality of costs imposed relative to benefits pursued—partly a matter of accurate and trustworthy estimation of social harms, partly a matter of limiting disruptions by phasing in large changes gradually, for Max as for conventional regulation. And they will depend on procedural recourse as protection against error and corruption, including provisions for explanation of decisions, independent review, and correction or compensation as judged warranted.

#### 4.6 What's the goal? Max's objective function and how it gets decided

We now come to the two hardest clusters of questions that Max presents. First, what goal does Max pursue in guiding his interventions, and how—and by whom—is this decided? And second, how might we get to Max: what pathways from present conditions to a society with Max in place might be feasible, likely, or desirable; how do these relate to present capabilities and trends; and what pitfalls and risks do these pathways present? I deal with the first set of questions in this section, the second set in the next.

What goal, what conception of social welfare, does Max pursue? In technical terms, what is Max's objective function, and how is it determined? I have presented Max as an alternative—or in the case of Pigovian Max, an augmentation and corrective—to markets. Market systems have a claimed normative foundation, originating in the “invisible hand” metaphor in Smith's *Wealth of Nations*<sup>63</sup> and later formalized in the two fundamental theorems of Welfare Economics.<sup>64</sup>

This normative claim depends on a few strong assumptions. The widely recognized and often-violated assumptions required for conditions of perfectly competitive markets— full information, no market power, no externalities—define most of Max's job as discussed thus far, so I do not address them further here.

But there are two other, more foundational assumptions on which the claimed social optimality of market outcomes depend. These assumptions allow markets—or more precisely, defenders of markets' optimality—to avoid certain hard problems that most forms of Max cannot. First, market optimality claims presume that people's

---

<sup>63</sup> ADAM SMITH, *AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS* (1776).

<sup>64</sup> A. MAS-COLELL, M.D. WHINSTON, & J.R. GREEN, *MICROECONOMIC THEORY* ch. 16 (1995).

market choices reliably reveal their preferences and their well-being. Second, these claims rely on a definition of social welfare, Pareto optimality, which excludes consideration of interpersonal welfare comparisons and distribution. These assumptions together allow a thin conception of social welfare, which avoids the need to define an explicit social welfare function but at the price of being silent on many points of clear importance for total societal welfare, notably, but not only, distribution and inequality.

Could Max get away with a similarly thin conception of social welfare, and thus avoid an explicit welfare function? This will depend on how broadly or narrowly his job is drawn. In its narrowest conception—Max only modifies each transaction to correct for information disparities, market power, and externalities—it is conceivable that Max could do this job, or approximate it, without an explicit social welfare function. Max could correct information limits or disparities between transacting parties. He could assess rents using internal accounting information from producers, perhaps augmented by comparative information from other firms in similar businesses. He could assess and correct externalities based on scientific knowledge about biophysical mechanisms of harm and estimates of people's valuation of the resultant end-states. To the extent external harms and benefits operate as public goods that affect multiple people, assessing their aggregate effect requires adding up individual effects and thus that these be expressed in commensurate terms, but does not require explicit interpersonal comparisons.

But Max also has the opportunity—or the duty—to allocate the available surplus from every transaction after he has taken account of externalities and rents. In doing this, he could take various simple approaches that can be defined from parties' relative valuations within the transaction, and thus do not require an explicit social welfare function. He could, for example, divide surplus in some given proportion between buyer and seller—equally, or in the same shares as the parties would have realized if Max had not intervened—applying such proportional division either to the entire available surplus, or to that portion that remains after Max takes some share as tax revenue.<sup>65</sup>

But any more ambitious approach that Max might take—including any approach that does not treat all transactions the same after accounting for externalities and market-concentration rents—must rely on characteristics of the parties external to the transaction, such as their wealth or other characteristics. Providing guidance for such choices requires an explicit social welfare function to define what count as better or worse social outcomes. As in many other applications, the shift to AI-directed decisions requires explication and codification of values and tradeoffs that may be left ambiguous or implicit absent such central direction.

Assuming Max is ambitious, and thus does require an explicit social welfare function, the task of defining it can be separated into two parts: defining individual welfare and aggregating across individuals to define overall social value. These two parts present different difficulties, and challenge different parts of the edifice of assumptions and arguments underlying normative claims for market outcomes.

First, how does Max define and measure individual people's well-being? In doing this, Max has a harder job than present AI systems, which only aim to predict commercially relevant behaviors: purchases, engagement, click-throughs, and the like. As noted above, normative claims for optimality of markets depend on assuming all these behaviors are aligned with your well-being, via one or another form of the axiom of revealed preference: if you do it, you must want it (relative to available choices); and if you want it, it must make you better off. This axiom provides a powerful foundation for liberal states: assuming you know what you value and act to pursue it is

---

<sup>65</sup> Those tax revenues don't cease to have any role in maximizing societal welfare, of course. But since we're characterizing Max as "the economy," and not "the state," the subsequent disposition of those revenues and how well they advance societal welfare is no longer Max's job: sending them to the government makes this their job, not his.

generally preferable to assuming I know what is good for you. On the other hand, the assumption is obviously false in many cases. People often make choices that are bad for them in a reasonably objective sense, e.g., in self-harming activities and use of recreational and performance-enhancing drugs that are addictive or harmful. And people often do, or fail to do, things that they later regret: not exercising enough, not saving for retirement, or spending too little time cultivating meaningful activities and relationships. Indeed, many business models depend on exploiting these misalignments, by taking advantage of impulsive behavior, distraction, or weakness of will.

We would want Max to avoid these clear pitfalls, ideally to do comprehensively better. But this ambition raises serious risks, including paternalism, loss of autonomy, or imposing one group's values on others, which require proceeding with great care. These risks are mitigated for Max in his Price or Pigovian forms, because he only has power to modify prices, not to tell you what to do. Max will discourage you from drinking or smoking by raising the price you face for alcohol or cigarettes<sup>66</sup>—perhaps encouraging moderation rather than abstinence by dynamically changing prices (I want another drink; Wait, it costs how much?) —but not saying you can't have them. He might even recycle the revenues realized from these high-priced transactions for your benefit, by directing them to your future health-care or retirement expenses rather than sending them to either the distillery or the treasury. But while this price-based approach reduces Max's coercive power over you, that power can still be substantial. Max must only wield it in service of your considered interests and values, not slide over to me (or anyone else) specifying how you should live or what you should want.

To achieve this balance, Max needs a model of your welfare that avoids pathologies of choice but that still represents **your** vision of your welfare. It must represent a considered view of your interests and values that is not distorted by unconsidered habit or impulse; is not manipulated by other parties for their own advantage; that takes account of how you want to be, even when your present behavior diverges from that vision; and that appropriately reflects intertemporal tradeoffs,<sup>67</sup> uncertainty, and the welfare of other people and values outside yourself – but that is still yours. Or at least, since Max's authority is limited to the economy, he needs a model of these things for you insofar as they are implicated in your economic transactions.<sup>68</sup>

To form this model, Max can draw on the same behavioral data firms already use and are developing, both data that pertains uniquely to you and generalizations inferred from other people. If Max is sufficiently trustworthy that we consent, he may also be able to draw on data not necessarily available to firms, such as medical data, or internal physiological and neurological observations, present and past. But Max's biggest advantage in forming this model of your welfare is that he does not have to do it alone. Like present proposals for AI-enabled personal assistants, Max can work with you, observing you and asking you about your preferences, aspirations, and feelings about your past choices and hypothetical future ones, to refine and update his model of your welfare. Operating in this way, Max looks more like a life coach or counsellor than an economic planner: indeed, this vision of Max is very similar to the approach proposed by Stuart Russell as a safety measure against AI assistants making serious errors when they act on your behalf.<sup>69</sup> Such a personal AI assistant would be concerned with many other choices in addition to your participation in economic transactions, however, raising the question of

---

<sup>66</sup> Note—the price increases discussed here target only your welfare, and are additional to any price adders imposed for reasons of externalities or market-power rents.

<sup>67</sup> This does not mean Max must predict your future preferences, only that he must appropriately reflect your present preferences about tradeoffs between present and future conditions.

<sup>68</sup> Max might also be able to advance welfare in ways not captured in consumption of transacted goods and services. But this would expand Max's job beyond the "economy," to the domain of the state. I avoid this expansion here, but recognize this may be a line that cannot hold. Unlike in the current decentralized, emergent economic system, it might be impossible to have an algorithmic system run the economy to advance welfare without also giving it extensive responsibilities that now lie with the state.

<sup>69</sup> STUART RUSSELL, HUMAN COMPATIBLE: ARTIFICIAL INTELLIGENCE AND THE PROBLEM OF CONTROL (2019) (particularly Chapters 7 and 8, as well as the technical papers cited therein).

whether this assistant should be some other AI-enabled agent, distinct from Max, whose information and concerns are limited to you. Such a personal AI agent—let’s call him Mini-Max—would closely resemble Russell’s faithful personal AI assistant, except that, as the guardian of your personal welfare, he would be responsible for passing on to economy-wide Max (“Big Max”) a subset of the information he holds about you, which is relevant to your preferences and welfare as they are connected to your participation in economic transactions, and the effects on you of externalities from others’ transactions. This is the information about you that Max needs to incorporate your welfare into his price-adjustment decisions. The rest of your interactions with Mini-Max, and the rest of his knowledge about you, are not needed by Big Max and can stay private between the two of you.

Even with a valid assessment of everyone’s welfare as affected by economic transactions, Max will still need to aggregate to a collective measure of social welfare. Because Max is serving in a liberal state—not a theocratic one, not one that tries to implement a universal Kantian approach to ethics (except, perhaps, in criminal law, which remains the state’s business, not Max’s)—that measure of social welfare must be some form of utilitarian summation of individual welfare measures as they pertain to economic activities. Any such aggregation requires weights attached to each person’s welfare. While giving equal weight to everyone’s welfare is an obvious default choice, there may also be legitimate bases to give some people’s welfare stronger weights than others’. In particular, under conditions of social inequality, it may be permissible, or even morally required, to give larger weights to the welfare of those worst off. Moreover, any aggregate welfare measure must consider the relative weights to give to economic versus non-economic contributors to welfare;<sup>70</sup> conditions at different times; and conditions that apply under different realizations of uncertainties. Except under the assumption that all these dimensions are correctly embedded in the individual welfare measures passed to Max, the social welfare function must represent collective judgments on these matters.

Although fully specifying Max’s objective function is beyond my scope here, this discussion suggests the problem can be approximated by specifying a few parameters. If we assume that Max’s social welfare function is some basically utilitarian aggregation of individual welfare measures, which takes appropriate account of inequality, time, uncertainty, and economic versus non-economic determinants of welfare, this suggests that specifying the function might be closely approximated by setting values for four parameters: (1) a measure of aversion to inequality to be used in setting relative weights for better and worse-off individuals; (2) a discount rate or other parameter to set the relative weighting of outcomes at different times;<sup>71</sup> (3) a measure of risk-aversion to weight outcomes under more or less favorable resolutions of uncertainties; and (4) a relative weighting of material consumption and non-economic contributors to welfare such as environmental conditions.

This last parameter, the relative weighting of economic and non-economic contributions to welfare, is likely to be the main instrument controlling the aggregate size of economic output under Max. If the material and energy flows associated with production, which determine environmental impacts, cannot be arbitrarily reduced toward zero, then environmental conditions will define the limits on the aggregate scale of the human productive enterprise. In a world of greatly reduced need for human labor in production, such environmental constraints are likely to be more tightly binding than any limit on production that arises from people choosing leisure time over employment.

In addition to asking what Max’s objective function is, we must also consider the process by which it is chosen. Although Max mostly represents a technocratic vision, this is a point where democracy must come in. Defining a

---

<sup>70</sup> Given Max’s purview of just the economy, non-economic determinants of welfare come into the objective function only as external effects of economic transactions.

<sup>71</sup> Noting the possibility that the correct form of discount function might not be exponential, and thus may require specifying more than one parameter.



collective conception of social welfare is an intrinsically political process, which must have people in charge working through some democratically legitimate mechanism. In considering how to do this, the assumptions already made have simplified matters considerably. Measures of individual welfare emerge from the interactions between people and their AI-enabled personal assistants, while the aggregation to social welfare has been reduced (for purposes of argument) to setting values for a few powerful, readily understandable parameters. Without denying the advantages of expert-driven, even technocratic, decisions for complex, largely instrumental decisions in pursuit of broadly agreed political ends,<sup>72</sup> this decision agenda is sufficiently clear and simple to place it within the capabilities of many different democratically legitimate processes. For example, you can imagine this as a legislative task, by which values for the major parameters of Max's objective function are explicitly enacted and periodically revised in statute. You can also readily imagine these as being matters of explicit debate in electoral politics, or being delegated to novel democratic processes such as juries of randomly selected citizens. You could even imagine the task being delegated to some expert administrative agency under legislative articulation of some higher-order aims to be advanced by the choice, assuming (in U.S. law) this decision survives the resultant constitutional challenge on non-delegation grounds.

The biggest risk associated with Max's objective function is the risk of capture. One irony that Max presents is that while one of his major jobs is reducing market power and associated rent-seeking in particular markets, the centralized political process of defining Max's objective function represents a concentrated opportunity for rent-seeking that overwhelms all others. Anyone able to inflect Max's decisions to serve their aims, even slightly, would be in a position of unprecedented power—to gain rapid wealth even beyond the dreams of tech-startup founders, or to shape society to their vision. Worse, the exercise of such power might be concealed by Max's status as a seemingly objective, neutral artifact.<sup>73</sup> Restricting the political agenda to setting a few highly aggregated parameters partly addresses these concerns.<sup>74</sup> These parameters do not allow the manipulation of small-scale details that would be needed to distort Max's decisions to a few actors' material advantage, and they aim to promote a democratic dialog on basic political values. But it is a long way from these high-level decisions to Max's actual operations, with many intervening steps that are more technical and opaque, over which many actors would love to exercise quiet influence. At the level of generality of this discussion, there is no more to say here beyond exhortations to vigilance about such manipulation, as much transparency as is feasible in the process of designing, training, and implementing Max, and procedures for recourse for those harmed by Max's decisions.

---

<sup>72</sup> See E.A. Parson, In Defense of (a Little) Technocracy, in *A SUBTLE BALANCE: EVIDENCE, EXPERTISE, AND DEMOCRACY IN PUBLIC POLICY AND GOVERNANCE, 1970–2010* (E.A. Parson ed., 2015).

<sup>73</sup> Max is thus subject to the same critiques as earlier, less tech-powered forms of the “high modernist” agenda to rationalize public decision-making. This concern, and its historical analogies, raise the question of the boundary between Max as “market” and Max as all-powerful, state-like actor, and so shake up old debates about legitimate state action. Scott's critique of state action rests on the same assumptions of limited knowledge and computational ability as old critiques of economic planning. So what happens to the critique when advances in AI and data falsify the assumption? I suppose Scott's state still “abstracts,” but the term now means something different, because the State can see all the graininess and particularity of people and events that it formerly could not. How much then, if anything, is left of the critique? See JAMES C. SCOTT, *SEEING LIKE A STATE* (1998); See also various contributions in *A SUBTLE BALANCE: EVIDENCE, EXPERTISE, AND DEMOCRACY IN PUBLIC POLICY AND GOVERNANCE, 1970–2010* (Edward A. Parson ed. 2015).

<sup>74</sup> The same agenda simplification of course represents a stark constriction of the space of democratic control of the economy. I have suggested these parameters might be set by a legislative, electoral, administrative, or citizen-consultation process, but it also partly resembles a constitution-making exercise: more, insofar as its effects are so foundational in society-building; but also less, insofar as the parameter values may need periodic updating as conditions and values change. (Note—this applies to the parameter values, not to the form of objective function, or the decision to adopt Max, which both appear more constitutional in their foundational character and long time-horizon.)

## 4.7 Getting to Max (and avoiding dangers along the way)

Max is a thought experiment, intended to be speculative and provocative. Yet part of the purpose of the exercise is to argue that Max is not crazily remote from present capabilities and trends. Many elements that could make up Max-like capabilities—rapid expansions in computational capacity, algorithms, data, and data integration and analysis tools—are already present or in development. These are mostly developing under private control to pursue commercial interests, or under state control to pursue military and geopolitical advantage, but not exclusively. There is also substantial research underway in universities and publicly supported research institutions, some of it loosely organized as a pursuit of “AI for good.”

In this section I shift from how Max would work as an endpoint to considering possible transition pathways by which Max, or similar capabilities, might come about. Any such pathway will involve a combination of technical and socio-political developments. I sketch three transition pathways that are sufficiently distinct and (to varying degrees) plausible to merit examination.

The first, and seemingly simplest, pathway would involve some jurisdiction deciding at some future point to adopt Max wholesale by political choice. Such a choice would lie within the authority of states, but would raise several immediate questions and challenges. Even assuming the needed capabilities existed, were ready to deploy, and confidently judged to work, the administrative scale of such a transition would be vast. It would require a massive roll-out and testing of infrastructure and systems before switching on, then some form of switch-over, perhaps at a long pre-announced moment during a period of reduced economic activity such as a near-universally observed religious holiday. The transition bears some resemblance to occasions when countries have reversed the direction of road travel, although the change would be much larger (albeit one not involving a risk of head-on collisions).<sup>75</sup>

Adopting Max would be a huge decision, beyond the authority of any administrative or executive process but requiring some democratically legitimate political process, legislative or perhaps constitutional. And it would present a chicken-and-egg problem regarding capabilities. Making such a choice would likely require confidence that needed capabilities are available, would work reliably, would deliver the promised benefits, and would present no severe risks. But such confidence could only be available after some long period of prior development and testing, which in turn would require prior political decisions to support these. Even those prior decisions to develop and test the capability would surely encounter stiff opposition, from those with strong ideological commitments to markets and from those benefiting from precisely those social harms—rents from market power, and uncharged negative externalities—that Max would target. In view of these difficulties, I suspect that adopting Max by explicit political choice would be highly unlikely, absent strong changes in political conditions such as an economic crisis so severe as to weaken the blocking power of incumbents. Even seeing Max operating successfully in other jurisdictions, while it might help (and thus imply that the first move would be the hardest), would probably not help enough absent a crisis.

A second possible route, potentially mitigating the extreme barriers for the first route, would involve early development, testing, or adoption of Max at smaller scale, among groups with more enabling political conditions. Possible early demonstrators and adopters might include jurisdictions that already have substantial shares of the economy in state enterprises or under state control; or those enterprises for which majority control already resides in some coalition of large sovereign wealth funds (Hello, Norway). Even jurisdictions with little state control of the economy could develop and test Max through government procurement, as governments often do

---

<sup>75</sup> See H. Peyton Young, *The Economics of Convention*, 10 J. ECON. PERSP. 105 (1996).

for early support of environmental technologies. Max might also be developed through progressive expansion from small, early, opt-in communities. These might be any group of individuals and organizations connected tightly with each other and less so with others—like religious groups, social or political experimenters, or relatively isolated political and economic jurisdictions—who would let Max, better now called “Pre-Max,” control their production and exchange relationships with each other.

Any such group of early adopters would face a few obvious challenges. They would have to port and modify capabilities from other uses, which in turn would require that these capabilities be sufficiently and verifiably adaptable to their new purpose and setting. Alternatively, they could develop the new tools and systems themselves, in which case they would need the resources to do this. Perhaps, given the novelty and importance of the experiment, they could attract philanthropic support. The initial group would have to be large enough and separate enough that their interactions with each other represent a substantial fraction of all their economic interactions. And to the extent they do trade with the rest of the world, they would need to ensure that such trade does not undermine Max whenever his prices diverge from private-market prices. An analogous problem would arise with any deployment of Max, at any scale. Whatever scope of transactions is given to Max, his authority over those transactions must be exclusive: black markets must be effectively prohibited, and exchange across the boundary of Max’s authority must not negate his adjustments. In the case of international trade, Max’s adjustments would have to be applied in parallel to traded transactions to avoid arbitrage opportunities, like proposed border tax adjustments on traded goods to preserve the effectiveness of greenhouse-gas or other environmental policies.<sup>76</sup> For this to be a viable transition pathway, Max must work well enough—perhaps after some early start-up phase carried by the enthusiasm of early adopters and start-up philanthropic support—that there are clear aggregate benefits to working with him that are visible to outsiders.

A third pathway, more continuous with present trends, would involve continued expansion and consolidation of Max-like capabilities in the private sector, to the point where a few enterprises or networks control a large fraction of the economy. It is widely noted that as the scale of platform monopolies grows, they increasingly resemble states and exercise similar authority, although without provisions to ensure democratic accountability.<sup>77</sup> Assuming some degree of concentration of private economic planning (and power) is widely viewed as unacceptable, Max could come about through some future political decision to take over and re-purpose the systems. This would not be a seizure and public re-purposing of physical capital assets, but of AI systems and associated data, although that nicety would hardly make the decision less wrenching and conflictual.

This pathway relies on two assumptions. First, it presumes that some future historical moment allows a wholesale takeover of concentrated private power that is then judged to have become intolerable, amounting to a large-scale reconfiguration of power between private and public actors. This would be a revolutionary change, carrying the risks of disruption and violence that typically attend revolutionary changes. Second, it presumes the technical feasibility of re-purposing a set of AI tools and data developed for private purposes to serve Max’s public aims. This may not be fully possible, as some of Max’s responsibilities—like assessing individual well-being, valuing externalities, and measuring rents—are not required of present systems serving private interests. To the extent the existing tools and data cannot perform these tasks, they would represent separate, new development requirements.

---

<sup>76</sup> See, e.g., TRADE AND CLIMATE CHANGE, UN ENVIRON. PROG. & WORLD TRADE ORG. (2009).

<sup>77</sup> See, e.g., Kristen E. Eichensehr, *Digital Switzerland*, 167 UNIV. PENN. L. REV. 665 (2019).

## 5. Conclusion: What this gets and leaves out, challenges unearthed

As a speculative exploration, this exercise does not lend itself to strong conclusions. Yet it appears to have yielded a few provisional observations and insights, which at a minimum suggest guidance for further exploration and research – including identifying some points of potential near-term guidance, for research and for early development of governance capabilities to manage risks.

First, I contend that the exercise has established some degree of plausibility for the hypothesized AI-driven central economic planning—under the admittedly strong assumptions made about technological capabilities. The exploration identified multiple developments underway that point toward the future capabilities assumed and found no showstoppers. Although this claimed demonstration of plausibility is highly qualified, it is not a trivial conclusion, since the exploration of different forms of Max under different contextual assumptions gave widely divergent views of their plausibility, with one variant of Max—Quantity Max in the presence of some degree of continued human managerial agency—presenting apparently insuperable obstacles.

More broadly, the exercise substantiated the general point that profoundly transformative applications and societal impacts from AI and related capabilities are plausible—with the potential for both great benefit and harm—long before the conventional mileposts of AI that transcends human capabilities and control. I have argued elsewhere for the importance of these “intermediate-range” AI capabilities and impacts, and for their distinct character from both near and long-term issues—in particular in their requirement for integrated examination of both technical characteristics of AI systems and the economic, political, and social context in which they are deployed.<sup>78</sup> While it is defensible to focus predominantly on technical characteristics in considering long-term risks, and on human interests and decisions in considering current applications and their impacts, neither of these simplifying assumptions is apt when considering intermediate-range capabilities and impacts. Max is surely not the only example of a plausible, profoundly disruptive potential AI application that falls in this middle range—indeed, this exercise suggests the value of thinking through other possibilities of similar transformative scale—but the detailed examination of Max and his implications hammers home the importance of these more vividly than the prior, more general arguments.

More specifically, the exercise of digging down to the particulars of Max’s operations and consequences yielded several suggestive insights, each offering useful guidance for further analysis and inquiry. First, it appears that alternative conceptions of how Max might be implemented differ starkly in their feasibility, requirements, and attendant obstacles and risks. In particular, the idea of Pigovian Max—a central-planning based implementation of a comprehensive system of Pigovian taxes—is a novel and promising vision of hybrid private-public control of economies, not previously considered in debates over central economic planning. Pigovian Max appears to offer the prospect of three major advantages, subject to all the requisite caveats. He appears potentially able to retain the efficiency and liberty advantages of private market systems while also correcting their most prominent failures. He also appears to offer the prospect of taxation without excess burden, albeit at the cost of aggressively individualized scrutiny of citizens’ preferences. And finally, he appears to offer the prospect, through management of the parameters of Max’s social welfare function, of bringing large-scale economic management under effective and informed democratic control, without losing the advantages of private markets.<sup>79</sup>

---

<sup>78</sup> See Parson et al., note 7; Parson et al., note 31.

<sup>79</sup> Note that these latter two advantages do not necessarily distinguish Pigovian Max from Price Max.

Second, even this preliminary investigation suggested that data and data integration needs may differ strongly over different forms of Max and different jobs given to him—for example, assessing individual welfare, pricing externalities, identifying and mitigating rents from market power, assessing and improving quality of working life, and promoting valued innovations and creative works. Moreover, these data needs may also differ strongly from those needed to predict and manipulate commercially relevant behavior for the benefit of counterparties. Assessing these needs for specific aims, meeting them without unacceptable harm to other values, and learning how to integrate private data about individual welfare with enough sharing to enable effective social optimization, will all be important research areas.

Additional areas of further inquiry suggested by the exercise include the preferred mechanisms for promoting innovation and creativity and innovation in society, in the presence of computational capability that can greatly accelerate and optimize at least those innovation mechanisms that depend on searching presently available information; and the content, structure, and means of defining an aggregate social welfare function. The initial inquiry into this latter question suggested that it might be much less difficult than widely assumed, but the high stakes involved suggest viewing this optimistic initial speculation with skepticism. Further critical investigations into potential forms of social welfare function that are precise enough to guide Max's decisions, but also clearly and simply parametrized enough to support meaningful democratic decision-making, are of high value; as is investigation of alternative democratically legitimate processes and institutions to conduct this parameter-setting process.

A particularly interesting area for further inquiry provoked by the exercise would be examining more limited deployments of Max. If the comprehensive, economy-wide Max discussed here is for some reason infeasible or unacceptable, might variants with more limited scope provide many of the proposed benefits with less cost, disruption, or obstacle? Max's scope might, for example, be limited to enterprises over a specified scale, or to sectors identified as presenting especially large externalities or tendencies to market power and rent-seeking. A particularly interesting variant would limit Max's authority to capital markets, either overall or jointly with a scale threshold. In this variant, "Capital Max" would allocate, or more likely price-adjust, capital to enterprises, replacing or operating in parallel with private capital markets. Capital Max would presumably use the same objective function as economy-wide Max. Since this function would consider both the private and public effects of enterprise operations, Capital Max would not precisely replicate the behavior of either private capital markets, or past efforts to allocate capital in line with political aims. Consequently, the well known critiques of these past efforts would not necessarily apply, any more than the old critiques of comprehensive central planning apply to economy-wide Max. Hints that Capital Max might be feasible and advantageous come from two lines of evidence: first, the extent to which capital allocation is already automated, via index funds, trading programs, and other algorithmic systems, which suggests that the change to Max might merely require adjusting the objective function; and second, the likelihood that key points in capital markets exhibit substantial market power, as well as systematic biases and choice pathologies. Capital Max might thus be able to gather low-hanging fruit, operating in parallel with and out-competing existing private capital-allocation mechanisms—and thus, ironically, subjecting them to increased market discipline. Viewed in this way, Capital Max would not aim to abolish Wall Street, but merely to subject it to real competition and so encourage it to work better.

Identifying these questions for further research and the associated stakes re-affirms one observation made in the introduction. It is widely noted that large technological change can drive transformative societal change, disruption, and conflict. But such changes can also explicate and disrupt foundational shared assumptions that underpin the norms, institutions, and power structures of society. In particular, these may depend on assumptions about what people are able to know about, and do to, each other that are technology-limited, but that are not recognized as such until the technology changes. This exercise has targeted long-settled

assumptions about the moral and instrumental effects of markets versus central economic control, but other unexamined foundational assumptions—in particular about the extent and form of power that some can exercise over others—may face similar disruptions under large-scale technological change.

Max, in particular his Pigovian form, presents three ambiguities, which should be kept in mind when considering his potential implications. First, is it ambiguous to what degree Pigovian Max would represent an incremental reform or a revolutionary transformation. I began the project as an intentionally extreme speculation about technological change and its implications. But elaborating the practicalities of implementing Pigovian Max made him increasingly look like a feasible, even incremental reform: an adjustment to improve a basically capitalist system, drawing on well established legal, institutional, and administrative capabilities, which appears quite compatible with a liberal democratic state. This claim must be qualified, of course, because implementation details will matter greatly: some variants of Max would clearly be so heavy-handed in their imposition of central control as to be incompatible with basic liberties. It might be a small step, easy to stumble over, from using price adders to correct clear externalities and rent-seeking, to adding incentives for sociability, pleasing others, conformity, docility, piety, or obedience to current political authorities. Any suggestion that Max might be a modest incremental change to the architecture of capitalism must reckon with these risks—and also with the challenge, discussed below, of finding a feasible and non-violent transition pathway that leads from here to Max.

A second ambiguity concerns the aggregate normative evaluation of Max: would he on balance be good or bad for human welfare? I began the exercise agnostic on this point, and reached the unsurprising conclusion that it could go either way, depending on design and implementation details that an inquiry at this high level of generality cannot resolve. Yet this experience also cast into sharp relief the strength of normative priors that animate other writings on this question, and how thoroughly and confidently these priors lead directly to the conclusions. This observation applies equally on both sides of the debate: on the one hand, to the growing number of socialists writing on AI central planning, who know—with little consideration of alternative implementation details or contextual conditions—that it would be good; and on the other hand, to the unnamed recent essayist in the *Economist*, who knows with similar prior confidence that it would be bad.<sup>80</sup> It appears clear that further investigations of this issue should link their normative assessments to explicit and specific assumptions about how central planning is implemented and what capabilities it draws on, in what context—even at the cost of yielding less clear and less predictable answers.

A third ambiguity concerns how to characterize Max's job, in particular as regards what he is replacing. Although the starting aim was for Max to replace "the market," working through the details led to a preferred form of Max, Pigovian Max, who lets the market operate then applies socially optimal adjustments to the resultant prices. Since controlling externalities and market power are canonical state functions, this makes Max look more like a comprehensive regulator—a state actor—than a market-like coordinating mechanism. Moreover, at each point in the argument where I proposed expanding Max's purview to include additional functions, these also looked more like state than market functions—or perhaps functions of the non-governmental charitable sector. Yet Max is not—and probably cannot and should not be—all of the state. The state does more than regulate, and even its regulatory functions are not limited to economic transactions. The state-market boundary is already fuzzy and contested, a point for which working through Max provided a helpful reminder. But introducing Max complicates, partly dissolves, and moves this state-market boundary.

---

<sup>80</sup> Beware the Borg: Beijing and San Francisco, *ECONOMIST* 57–62 (Dec. 29, 2019), <https://articles.cafeyn.co/1076b9/the-economist/2019-12-18/beware-the-borg>.

In closing, I return to the question of Max’s plausibility, and to the most disturbing issue raised by the exercise. I claimed above that Max passes some threshold test of plausibility, but plausible does not mean likely. Even a more complete and persuasive demonstration that a fully implemented Max would raise no impossible conditions would not necessarily imply a feasible or acceptable transition path to get from here to there. A technological artifact of Max’s scale and complexity does not arise spontaneously, but must be pursued and developed by actors who can mobilize the needed (albeit uncertain) scale of expertise, resources, and authority. Max’s real-world feasibility will thus depend on both needed technological capabilities and favorable social and political conditions. In this regard, the fact that Max-like capabilities, or large parts thereof, are already present or in development—with the crucial difference that these developments are in private hands and aim to advance private or sectional interests, not broad public ones—cuts both ways, both for Max’s feasibility and for the prospect of AI bringing broad human advances. Two sobering implications follow.

The first concerns the risk of lost opportunities. There may well be prospects for mid-term AI developments that could bring profound advances in human welfare, whether through something like Max or through other applications in health, environment, education, or government. But if the specific technical requirements to realize such broad benefits differ greatly from those being pursued by private actors, then near-term RD&D decisions, plus path-dependency, may foreclose the prospect for such transformative future benefits. The severity of this risk depends on the portability and adaptability of capabilities—how readily capabilities developed for private or rival purposes can be adapted to serve public or universal ones—which is deeply uncertain.

The second implication concerns the medium-term implications of continued dominance of private actors and interests in guiding development of increasingly powerful AI capabilities. Continued expansion of capabilities could become self-reinforcing: not in the oft-proposed sense of AI systems themselves growing unboundedly powerful through recursive self-improvement; but rather in the sense of capabilities controlled by human actors recursively strengthening the concentration of social, economic, and political power in those actors’ hands. Perhaps even worse than the loss of potential human-liberating capabilities, such trends could lead to profoundly dystopian futures, whether these come about with a bang (violent upheaval) or a whimper (incremental loss of human welfare, agency, and hope).<sup>81</sup>

These dire possibilities suggest the value of large early investments in development of AI and related capabilities that are explicitly targeted at comprehensive public benefits. This may sound obvious, but it may in fact be the most radical suggestion in the paper, because such efforts might not just differ greatly from present privately-driven developments, but also from present small “AI for good” efforts, in at least two respects. First, the needed development efforts would not erroneously assume that economic development benefits to the sponsoring jurisdiction—the growth and competitive success of enterprises located there, or the successful tech-industry job placement of young people trained there—are identical to the aggregate public benefit. Second, they would not presume that the technological capabilities developed for private commercial advantage will be readily and without limit re-deployable in pursuit of non-commercial public purposes. This will be the case to some degree, of course, and a development program seeking public benefit should not needlessly re-invent wheels that can equally well be installed on both public and private vehicles—but how much, in what particulars, and for how long this will be the case is deeply uncertain, and it would be naïve for a publicly motivated development effort to assume comprehensive, continued complementarity between these. I recognize that the implications of this

---

<sup>81</sup> I am sketching these bleak future pathways in terms that suggest the actors achieving such dominance develop from present private commercial actors, only because most present investments and advances are in the private sector. A similarly bleak future could arise substituting state or quasi-state actors for private ones. See, for example, the discussion of “robust totalitarianism” as a societal risk from AI in Alan Dafoe, *AI Governance: A Research Agenda*, OXFORD FHI V. 1.0 (Aug. 27, 2018), <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAIAgenda.pdf>.

conclusion for resource requirements are large—and at odds with present trends in public-private division of resources and authorities—but the risk of continued, uncritical reliance on the assumed complementarity of technologies to advance competitive or rival interests and to serve broad public ones, appears too large to ignore.