# 1AC

## 1AC---Class Action

Advantage One is Class Action.

#### Compulsory arbitration immunizes corporations from liability.

Newton ’20 [Dawn; October 26; Business attorney who focuses on franchise law, intellectual property, and data privacy, J.D. from the Hastings College of the Law at the University of California; Donahue Fitzgerald, “Avoid Class Action Suits by Using Arbitration Agreements,” <https://donahue.com/resources/publications/avoid-class-action-suits-using-arbitration-agreements-2/>]

The U.S. Supreme Court paved the way for businesses to avoid class actions by using arbitration agreements.

On April 27, 2011, the U.S. Supreme Court issued a ruling in AT&T Mobility LLC v. Vincent Concepcion that barred a consumer class action and overturned prior California law that prohibited class action waivers in arbitration agreements. In doing so, the Court created a mechanism by which savvy businesses can avoid class actions in a number of different contexts, which may, in turn, dramatically impact the way companies resolve disputes with their customers, employees and contractors.

Underlying Facts and Procedure

In 2002, in response to a promotional advertisement, the plaintiffs in the case, Vincent and Liza Concepcion, entered into a cell phone contract with AT&T (Cingular Wireless at the time) that entitled them to two “free” phones. After the Concepcions signed the company’s mobile service agreement, they discovered that although AT&T’s advertisement promised that they would receive the phones for free, they were charged $30.22 as tax on the phones’ retail value.

In March 2006, the Concepcions sued AT&T in the United States District Court in the Southern District of California, alleging that it engaged in false advertising and fraud. The District Court later consolidated the case into a class action lawsuit.

AT&T challenged the lawsuit, arguing that its standard mobile service contract, signed by the Concepcions and all other customers, precluded a class action because of two key provisions in the agreement: a mandatory arbitration clause and a class action waiver that required consumers to bring claims only in their individual capacity. AT&T asked the court to send the matter to arbitration in each plaintiff’s individual capacity as was required by the contracts.

The plaintiffs opposed AT&T’s request, arguing that a well-established decision by the California Supreme Court–Discover Bank v. Superior Court, 36 Cal.4th 148 (2005)–had already determined that class action waivers in consumer contracts were unreasonable, or “unconscionable,” and therefore unenforceable. The District Court agreed with the Concepcions and denied AT&T’s request to compel arbitration, and the Ninth Circuit Court of Appeals affirmed that decision. Critical to the decisions of both the District Court and the Ninth Circuit Court of Appeals was the determination that the Federal Arbitration Act (“FAA”), the purpose of which is to ensure the enforcement of arbitration agreements according to their terms, did not override, or “preempt,” the California Supreme Court’s decision in Discover Bank.

The Supreme Court Ruling

In a remarkably broad 5-4 ruling, divided along ideological lines, the Supreme Court reversed the Ninth Circuit, holding that the Discover Bank rule is inconsistent with the FAA and is therefore preempted.

Writing for the majority, Justice Antonin Scalia stated that the purpose of the FAA was twofold: to ensure the enforcement of private agreements to arbitrate and to make dispute resolution more efficient. By essentially allowing the parties to re-write arbitration agreements after the fact, Scalia explained, the rule in Discover Bank “stands as an obstacle to the accomplishment and execution” of Congress’s purpose and objective in enacting the FAA.

Specifically, the Court found that allowing class arbitration under the Discover Bank rule is inconsistent with the FAA for three reasons: 1) when compared with two-party arbitration, class arbitration makes the process slow, expensive and procedurally difficult; 2) class arbitration requires a formality that was not envisioned by Congress when the FAA was passed; and 3) class arbitration greatly increases the risk to defendants because a single mistake by the arbitrator, multiplied by thousands of class members, could result in a “devastating loss.” As a result, according to the Court, the rule in Discover Bank is likely to dissuade companies from using arbitration as a means of resolving disputes, thereby undermining the FAA.

Implications

The implications of the ruling in AT&T Mobility are significant. Indeed, the ruling comes only a year after the Supreme Court’s decision in Stolt-Neilsen S.A. v. Animalfeeds International, Corp., 130 S. Ct. 1758 (2010), in which the Court held that if parties to an arbitration agreement did not intend to allow class claims, arbitrators have no power to impose class-wide arbitrations under agreements that are merely “silent” on the issue. Both decisions further the Court’s clear directive that the FAA, and its liberal policy favoring arbitration, should be given full effect.

Given the ubiquity of arbitration clauses in everything from cell phone plans to franchise agreements to employment contracts, the potential ramifications of the Court’s decision in AT&T Mobility are sweeping.

Impact on Franchisors & Distributors

Many franchisors with arbitration provisions in their franchise agreements have felt insulated from the threat of class-action suits since last year’s Stolt-Nielsen decision. However, some lingering doubt remained, since some California courts had allowed franchisee attempts to either avoid arbitration or force a class certification since Discover Bank was decided, citing that case as authority. The AT&T Mobility decision puts an end to this argument.

Franchisors or distributors that draft consumer contracts for use by others in their distribution chain should also pay close attention to this decision. By incorporating arbitration clauses into those consumer agreements, they can eliminate the threat of class actions, provided that the terms for the individual arbitration are not unfair. The majority opinion noted that AT&T’s arbitration policy prevented AT&T from seeking its own attorney’s fees if it prevailed in arbitration, but provided for payment of twice the plaintiff’s attorney’s fees if AT&T lost. In addition AT&T’s policy agreed to pay an additional $7,500 to any plaintiff who obtained a better arbitration award than AT&T’s last settlement offer. While the opinion does not require a consumer arbitration clause to match these terms, the FAA would not require the enforcement of an obviously unfair arbitration provision.

Impact on Employers

The Supreme Court’s ruling is also certain to have far-reaching consequences for employers that use arbitration agreements with their employees. In particular, it could dramatically change wage and hour litigation, which often relies on class actions as a vehicle to redress relatively small individual losses. Without being able to consolidate these claims into a large class action, it will likely become increasingly difficult for individual employees to secure legal representation. Indeed, if an employer’s arbitration agreement is drafted correctly, because of the ruling in AT&T Mobility, it can potentially eliminate the risks to employers of facing a wage and hour class action (or any other employment-based class actions).

#### It's a shield against antitrust lawsuits AND increasing.

Rubinoff ’20 [Matt; May 1; J.D. Candidate at Pennsylvania State University, Managing Editor; Arbitration Law Review, “Too Big to Arbitrate? Class Action Waivers, Adhesive Arbitration, and Their Effects on Antitrust Litigation,” vol. 12]

I. Introduction

Access to arbitration provides a simple, expeditious, and “feasible form of fairness in adjudication.”1 In avoiding the typically tedious and strenuous litigation procedures, arbitration remains “a vital part of the litigation alternatives in the [United States] legal system.”2 However, what happens when one side, particularly the poorer and weaker party, unwillingly or unknowingly submits to an alternative legal path that hinders their chances of success, fairness, or justice?

Class action procedures allow courts to manage lawsuits that would otherwise be impossible or improbable if each class member were required to join as an individually named plaintiff.3 The process “enables vindication of claims that otherwise could never be litigated, no matter how meritorious.”4 Antitrust lawsuits, for example, typically provide access to class actions for consumers or employees, and represent the primary way to compensate victims who suffer a loss, while subsequently providing a strong deterrent to future illicit behavior.5 In the age of global marketing, communication, and access to information, “it is not uncommon for many individuals to be harmed in essentially identical ways by mass-produced products or standard corporate practices.”6 Class actions attempt to level the playing field for those groups otherwise economically disadvantaged or lacking access to the justice system.7

Over the years, however, large corporations have begun to routinely opt to use arbitration agreements to prevent class action lawsuits through class action “waivers” imbedded in their contractual agreements.8 These waivers too often remain hidden or unseen to the final user of the product or employee of the company until after they have already agreed to the contract. 9 Accordingly, parties seeking a cause of action are not only forced to arbitrate any dispute, but also are prevented from certifying the lawsuit in a class action proceeding.10

Footnote 8:

8 See Megan Leonhardt, Lawmakers want to give Americans back their right to sue companies, CNBC (Sept. 10, 2019, 5:32 PM), https://www.cnbc.com/2019/09/10/lawmakers-want-to-give-americans-backtheir-right-to-sue-companies.html (“A recent academic study found that 81 of the biggest 100 companies in America have put legal clauses in the fine print of their customer agreements that bar consumers from suing them in federal court…”).

End of footnote 8.

This article discusses class action waivers in mandatory arbitration agreements and examines their possible impact on access to antitrust violation claims. Part II introduces the recent history of class action waivers and the United State Supreme Court’s continued enforcement of such waivers. Part III addresses the impact of case law on antitrust laws and procedures. Finally, Part IV concludes with recent developments, and what the near future might look like for mandatory arbitration agreements and class action waivers.

II. The Supreme Court and Class Action Waivers

Arbitration settles legal disputes in a functional, flexible procedure, avoiding the unneeded expenses and delays of a common judiciary. 11 On a macro level, arbitration represents a justifiable avenue to not only avoid backing up the court system with claims, but also expediates the process for legitimate parties seeking relief.12 From this perspective, the goals of arbitration inherently seek to support litigants who do not have the time or money to compete with a powerful opposition. However, as the Supreme Court, as well as the legislature, increasingly relies on the enforcement of arbitration agreements, the bigger, stronger parties strategically use customized adhesive arbitration contracts and class-action waivers as a protective shield to further benefit themselves and effectively limit any future liabilities.13

#### Overwhelming empirical and statistical evidence verifies private, class action suits are necessary to deter cartelization.

Lande ’16 [Robert; Spring 2016; Venable Professor of Law at the University of Baltimore School of Law, Director of the American Antitrust Institute; Antitrust, “Class Warfare: Why Antitrust Class Actions Are Essential for Compensation and Deterrence,” vol. 30]

Our recent empirical studies demonstrate five reasons why antitrust class action cases are essential: (1) class actions are virtually the only way for most victims of antitrust violations to receive compensation; (2) most successful class actions involve collusion that was anticompetitive; (3) class victims’ compensation has been modest, generally less than their damages; (4) class actions deter significant amounts of collusion and other anticompetitive behavior; and (5) anticompetitive collusion is underdeterred, a problem that would be exacerbated without class actions.

Recent court decisions undermine class action cases, thus preventing much effective and important antitrust enforcement.1

Class Actions Are Virtually the Only Way for Most Victims of Federal Antitrust Violations to Receive Compensation

The antitrust statutes provide that violations result in automatic treble damages for the victims.2 The legislative history 3 and case law indicate that compensation of victims is a goal, perhaps the dominant goal, of antitrust law’s damages remedy.4 Class actions play an essential role in ensuring that the treble damages remedy serves its intended function of “protecting consumers from overcharges resulting from price fixing.”5 As the Supreme Court noted, “[C]lass actions . . . may enhance the efficacy of private [antitrust] actions by permitting citizens to combine their limited resources to achieve a more powerful litigation posture.”6 Accordingly, “courts have repeatedly found antitrust claims to be particularly well suited for class actions . . . .”7

Without class actions, cartels and other antitrust violators that inflict widespread economic harm would have little to fear from the treble damages remedy. This is because, as a practical matter, class action cases are virtually the only way for most victims of anticompetitive behavior to receive compensation.8 A 2013 study that Professor Joshua Davis and I conducted documents the benefits of private enforcement by analyzing 60 of the largest recent successful private U.S. antitrust cases (defined as suits resolved since 1990 that recovered at least $50 million in cash for the victims9 ). These actions returned a total of $33.8–$35.8 billion in cash to victims of anticompetitive behavior.10 These figures do not include products, discounts, coupons, or the value of injunctive relief or precedent—only cash.11 Consequently, these totals significantly understate the actual benefits of this litigation to the victims involved. And, of course, this study covered only 60 suits (albeit 60 of the largest private recoveries) out of the many hundreds of private cases filed in the United States during this period.

Of these 60 large private cases, 49 were class action suits.12 These cases recovered a total of $19.4–$21.0 billion—the majority of the amount analyzed in our study.13 Since these were among the largest private actions ever filed, specific conclusions based upon these results may not generalize perfectly to all class action cases. They do suggest, however, that without class action cases, effective and significant victim compensation would be reduced dramatically.

Most Successful Class Actions Involve Collusion that Was Anticompetitive

Almost every private antitrust case that results in a remedy does so through a settlement,14 so the underlying merits of the plaintiffs’ claims usually have not been definitively assessed by a court or jury. Critics sometimes use this fact to support assertions that class actions usually are meritless, that plaintiffs often receive huge sums from cases not involving anticompetitive conduct, and that private antitrust actions often amount to legalized blackmail or extortion.15

Antitrust class actions arise in widely varied market and factual settings, and views about the merits of specific cases and the litigation risks involved vary as well. This makes it extremely difficult to draw objective conclusions about the merits of settlements.

Nevertheless, there are good reasons to believe that the vast majority of class action cases in the Davis/Lande study involved legitimate claims. Forty-one of the 49 class actions involved allegations of collusion,16 and the same conduct supporting the settlements gave rise to criminal penalties in 20 cases; to civil relief by the FTC or DOJ in 8 cases; to civil relief by a state or other governmental unit in 9 cases; to a trial that the defendants lost and that was not overturned on appeal in 7 cases; to a class being certified in 22 cases; and to plaintiffs surviving or prevailing at summary judgment in 12 cases.17 Overall, 44 of the 49 class action suits (90 percent) exhibited at least one of these forms of legal validation as to their merits. (The 5 actions that did not have at least one of these indicia settled too early for a substantive evaluation of their merits).18

These results are broadly consistent with a finding that Professor John Connor derived from an analysis of 130 private recoveries worldwide in international cartel cases for which he could obtain the necessary data.19 He found that of the 50 largest worldwide settlements, measured by their monetary recoveries in constant dollars, 49 had been filed against international cartels.20 Of these, 51 percent were follow-ups to successful DOJ prosecutions, and another 8 percent were filed after fines by the EC or other non-U.S. antitrust authorities.21 Using a different data set, Connor and I found that 36 of 71 (also 51 percent) successful U.S. class action recoveries followed successful DOJ criminal cases.22

This data does not prove that these or any other specific class action cases involved anticompetitive conduct. But critics who assert that most antitrust class actions are little more than legalized blackmail rely only on anecdotes, hypotheticals, and opinions (often of defendants in the cases), without support from studies, and with no reliable empirical evidence that the actions lack merit or that settlement amounts are excessive compared to the anticompetitive harm.23 To be fair, one should compare the above indicia of validity to the absence of any systematic evidence underpinning the critics’ charges.

#### Cartels bulldoze economic growth and cascade through umbrella effects on innovation.

Crowe ’16 [Jonathan and Barbora Jedlickova; 2016; Professor of Law at Bond University; Lecturer and Fellow at the Center for Public, International, and Comparative Law at the University of Queensland; Federal Law Review, “What’s Wrong with Cartels?” vol. 44]

Cartels, then, are widely viewed as immoral, as well as economically inefficient. The academic literature, however, has tended to focus primarily on economic factors. 15 There has been significant recent discussion of the rationales for criminalising cartel conduct, 16 but relatively few attempts to systematically integrate economic and moral considerations. The leading exception is Stuart Green's framework for analysing whitecollar crime, 1 7 which has been applied to cartels by a number of authors. 18 The present article suggests an alternative approach to this issue. We offer an integrated account of the wrongness of cartels that emphasises the relationship of cartel behaviour to the moral duty to promote the common good. Cartels are wrong because they undermine the role of open and competitive markets as a salient response to an important social coordination problem in a way that causes serious harm to community welfare. This combination of factors supplies a robust justification for both civil and criminal sanctions in appropriate cases.

The article begins by briefly exploring the traditional economic justification for prohibiting cartels. This economic case is relatively uncontroversial, but it provides necessary background for the wider argument. We then introduce the idea of common good duties. We argue that people have a duty to contribute to salient solutions to coordination problems that seriously threaten the common good of their communities. The next section explores the role of the competitive market in solving social coordination problems: we argue that support for market mechanisms and the competition values they embody can form part of a person's common good duties. This provides a normative basis for prohibiting cartels. We conclude by connecting this account of the wrongness of cartels with the respective roles of civil and criminal remedies. The resulting theory offers a principled foundation for the current framework of cartel regulation in Australia.

II. The Economic Harms of Cartels

Cartels are forms of collusive conduct that involve participants deciding to stop competing in certain ways -for instance, by setting the price, setting and decreasing output or dividing territories or customers. There is no uniform definition of 'cartel', but many formulations focus on the economic impacts of cartel behaviour. 19 Legal prohibitions on cartels typically target a range of anticompetitive conduct, including price-fixing, output restrictions, market-sharing arrangements and bid-rigging. 20 In this article, we will focus primarily on naked horizontal collusion, which is purposely used to restrict competition and has no obvious social benefit. We will use the term 'cartels' to refer to the above-mentioned varieties of anticompetitive horizontal collusion among independent economic entities. 21

The current approach to regulating cartels-as implemented, for instance, in Australia, the US and the EU- is broadly based on a framework of welfare economics informed by a neoclassical focus on economic efficiency and overall social welfare. 22 The neoclassical outlook, associated most notably with the Chicago School, imports a conception of economic agents as rational actors seeking to maximise their utility, along with a classical liberal emphasis on the beneficial consequences of a free and competitive market.23 This yields a strongly consequentialist approach to the regulation of cartel behaviour. The focus of regulatory efforts, on this view, falls on combating the harms caused by cartels in the form of decreased outputs. The purpose of regulation is to increase the costs of cartelisation and remove the incentives people might otherwise have to collude rather than compete.

The classical liberal tradition treats a free and competitive market as having high social value because it leads to higher productivity, lower prices and increased innovation. 24 A free and competitive market, on this simplified model, naturally reflects demand and supply. Suppliers (companies and other entities) compete for customers' attention essentially because they want to sell as many products as possible, for as high a profit as possible. Suppliers are therefore motivated to innovate and differentiate their products and services, while decreasing their cost of production. They have incentives to be more efficient. Price signals enable this process by aggregating the information available to actors in the market and expressed in individual transactions. 25 The price system is highly dynamic- it adjusts constantly as players in the market take account of new information and use it to guide their choices. It will probably never lead to perfect coordination of preferences under actual market conditions, but it plays this role more effectively than other available methods. 26 This information-sharing function of a competitive market is disturbed if competitors take measures to restrict competition.

There is one primary reason that someone would stop competing and start colluding: profit generation. For instance, if price competition becomes intensive through the reduction of the price of substitutable products, the market may reach a point where the total cost is equal to the price charged by the producer. Such a producer must assess its options for starting to profit again. It can, for example, lower its costs or perhaps improve its product - or it can become involved in a cartel, such as a price cartel. Cartel participants become involved willingly because of the prospect of generating better (and commonly monopolistic) profit. This increases the welfare of the cartel members, but decreases total and consumer welfare. The final price consumers pay is higher and there is less incentive to innovate or become more efficient (for example, by cutting costs) on the side of the cartel participants.

Cartels, then, harm social welfare by disrupting the competitive functioning of the market. They damage 'confidence in the market's ability to act as the chosen mechanism for transfer and distribution'. 2 7 There are also more specific forms of harm associated with cartels. Most importantly, cartels lead to higher prices and lower output of the products or services concerned. 28 Price cartel harm therefore has two major components. 29 The first is the transfer of wealth from customers to the price fixers generated by artificially high prices. This also decreases the total social welfare because the overcharge customers pay as part of the cartel price could be spent on additional goods or services. The second component of harm arising from price cartels is the reduction of the quantity of products sold. A higher price means consumers buy less of the product and thus generate a so-called 'deadweight loss', which is a sum of the lost utility from potential purchases made by consumers and the potential additional sales if the price had been a competitive one. 30

Other consequences of cartels include decreased innovation and efficiency. The members of a price-fixing cartel lose the motivation to produce their products for less than others and thus become less efficient. They can also lose the motivation to innovate and improve their products to attract more customers or, alternatively, increase prices once their products are better than others available. These effects may extend beyond the members of the cartel to other participants in the market. For example, cartels have been associated with 'umbrella effects' that arise where competing firms are assured of buyers, provided they set their price below that of the dominant company in the market. 31 If the dominant entity is part of a cartel, then other sellers will also be able to profit from its anticompetitive prices.

III. The Problem of the Baseline

Cartels, then, are economically harmful. They increase prices and decrease innovation, efficiency and output in the market, thereby making society as a whole worse off than it would otherwise have been. However, this consideration alone does not explain why cartels are morally objectionable. The proposition that cartels make the market less efficient is a descriptive claim; it cannot, by itself, generate a moral conclusion. The notion of harm, one might say, is normatively loaded: once we accept a general duty not to cause harm, we can then derive the normative conclusion that it is wrong to cause harm of the specific kind involved in creating cartels. However, this appeal to the wrongfulness of harm raises two further issues.

#### Microeconomic theory confirms---enforcing against collusion is the centerpiece of growth and innovation---AND makes growth sustainable.

Martinez ’21 [Diego and Pramuan Bunkanwanicha; carbon dated to August 17; M.A. in Finance from Universidad Complutense de Madrid, E.Ph.D. Candidate at ESCP Business School; Professor of Finance at ESCP Europe, Ph.D. in Economics from the University of Paris; ESCP Research Institute of Management, “Good Faith Competition as a Natural Mechanism for Sustainable Economic Growth,” <https://academ.escpeurope.eu/pub/IP%202021-31-EN.pdf>]

I. Introduction

Microeconomic theory defines the market as perfect competition when firms provide goods at a price that equals their marginal cost. Some common characteristics of a perfectly competitive market include homogenous products, all buyers and sellers as price takers, there is complete information, and no entry and exit barriers. Under the assumption of prices equal marginal costs, firms would have no or little incentive to innovate.

It is reasonable to expect that most industries are characterized by some degree of heterogeneity and product differentiation. In this situation, the competition encourages profit-maximizing firms to innovate to achieve abnormal returns.

Rooted in management literature known as the resource-based view of the firm, Barney (1991) argues that sustainable competitive advantage derives from the resources and capabilities a firm controls that are valuable, rare, imperfectly imitable, and not substitutable. It is arguable that the firm's sustainable competitive advantage should be connected with the environment where the firm operates. Good faith competition incentivizes firms to build sustainable competitive advantages through R&D investments, product differentiation, advertising, and capital-and cost-efficiencies. Firms need to invest in tangible and intangible resources to create competitive advantages and generate abnormal returns (returns on equity higher than the cost of equity). Firms also need to continue investing in maintaining those advantages over time to create long-term value.

Kline and Rosenberg (2010) define the process of innovation as a series of changes that affect not only hardware but also production, markets, and organizations. In fair competition markets, a firm's search for creating competitive advantages provides a continuous investment process and stimulates innovation, providing economic growth, employment, and welfare enhancement (Baumol and Strom 2007, OECD 2007, Daniels 1996).

Sustainable economic growth has important implications for society. In the long run, economic growth is mainly explained by technological progress. Sustained economic growth has an amplified effect on per capita income, and it is an effective mechanism to reduce poverty rates (Barro and Sala-i-Martin 2004, Sala-i-Martin 2006, Dollar et al. 2013). United Nations' 2030 Agenda for Sustainable Development1 includes eradicating poverty as an indispensable requirement for sustainable development. In fair markets, firms competing for competitive advantages take a crucial role, bringing the power of innovation that generates economic growth, resulting in an improved standard of living for the wider society. However, some firms may have incentives to collude to obtain extra-profits, harming consumers and, at the same time, negatively affecting the power of innovation. Regulators have to ensure the fair functioning of markets.

II. Advantages of good faith competition

The positive effect on society of firms' rivalry is based on three central ideas. The first one is that firms pursue a profit maximization strategy and expect to achieve abnormal returns. The second one is that industries have some degree of heterogeneity and product differentiation. Lastly, firms compete in fair markets. In this scenario, firms pursuing abnormal returns will make investments in order to develop competitive advantages. Investment in R&D is one of the most important activities driving competitive advantage, and firms in competitive industries enter into innovation races to differentiate their products. Innovation affects long-term economic growth through technological progress. The European Central Bank supports innovation as an essential driver of economic progress that benefits consumers, businesses, and the economy as a whole.

Fair market competition is one of the pillars for obtaining positive effects from rivalry. National and supranational organizations acknowledge the benefits of good faith competition. The Autorité de la concurrence, the competition regulator in France, argues that competition forces companies to be innovative and to stimulate growth and jobs. The European Union states that having firms competing fairly in the market benefits society. Consumers receive higher quality products at better prices, and competition incentivizes firms to innovate to differentiate their products and make firms more competitive in global markets.

In fair markets, the search for competitive advantages stimulates innovation and strengthens long-term economic growth. The Presidency Report to the Council of the EU (September 20th, 2019) on developing long-term strategies of sustainable growth identifies Research and Innovation (R&I) as a critical driver in response to the main challenges of the European economic growth model. Economic growth does not need to be explosive but recurrent over the long term. An example of the positive effects of long-term economic growth on income per capita is the U.S. economy. The US GPD per capita grew at a yearly rate of 1.8% between 1870 and 2000, resulting in an increase of 10 times, from $3,340 to $33,330 measured in 1996 dollars. However, reducing the yearly growth rate to 0.8%, the per capita rent in 2000 would have been $9,450, only 2.8 times the value of 1870, and the U.S. would be ranked in 45th position instead of 2nd out of 150 countries (Barro and Sala i Martin 2004).

Arguably, designing good faith competition markets is a natural mechanism to promote sustainable economic growth. Fair competition stimulates innovation, which is the main contributor to sustainable economic well-being.

#### Slow growth collapses the liberal order AND causes global hotspot escalation---extinction.

Oppenheimer ’21 [Michael; 2021; Clinical Professor, Center for Global Affairs, New York University. Senior Consulting Fellow, Scenario Planning at the International Institute for Strategic Studies. Former Executive Vice President, The Futures Group. Member, Council on Foreign Relations. Member, The Foreign Policy Roundtable at the Carnegie Council on Ethics and International Affairs. Member, The American Council on Germany; The Future of Global Affairs: Managing Discontinuity, Disruption and Destruction, “The Turbulent Future of International Relations,” Ch. 2]

Four structural forces will shape the future of International Relations: globalization (but without liberal rules, institutions, and leadership)1; multipolarity (the end of American hegemony and wider distribution of power among states and non-states2); the strengthening of distinctive, national and subnational identities, as persistent cultural differences are accentuated by the disruptive effects of Western style globalization (what Samuel Huntington called the “non-westernization of IR”3); and secular economic stagnation, a product of longer term global decline in birth rates combined with aging populations.4 These structural forces do not determine everything. Environmental events, global health challenges, internal political developments, policy mistakes, technology breakthroughs or failures, will intersect with structure to define our future. But these four structural forces will impact the way states behave, in the capacity of great powers to manage their differences, and to act collectively to settle, rather than exploit, the inevitable shocks of the next decade.

Some of these structural forces could be managed to promote prosperity and avoid war. Multipolarity (inherently more prone to conflict than other configurations of power, given coordination problems)5 plus globalization can work in a world of prosperity, convergent values, and effective conflict management. The Congress of Vienna system achieved relative peace in Europe over a hundred-year period through informal cooperation among multiple states sharing a fear of populist revolution. It ended decisively in 1914. Contemporary neoliberal institutionalists, such as John Ikenberry, accept multipolarity as our likely future, but are confident that globalization with liberal characteristics can be sustained without American hegemony, arguing that liberal values and practices have been fully accepted by states, global institutions, and private actors as imperative for growth and political legitimacy.6 Divergent values plus multipolarity can work, though at significantly lower levels of economic growth-in an autarchic world of isolated units, a world envisioned by the advocates of decoupling, including the current American president.7 Divergent values plus globalization can be managed by hegemonic power, exemplified by the decade of the 1990s, when the Washington Consensus, imposed by American leverage exerted through the IMF and other U.S. dominated institutions, overrode national differences, but with real costs to those states undergoing “structural adjustment programs,”8 and ultimately at the cost of global growth, as states—especially in Asia—increased their savings to self insure against future financial crises.9

But all four forces operating simultaneously will produce a future of increasing internal polarization and cross border conflict, diminished economic growth and poverty alleviation, weakened global institutions and norms of behavior, and reduced collective capacity to confront emerging challenges of global warming, accelerating technology change, nuclear weapons innovation and proliferation. As in any effective scenario, this future is clearly visible to any keen observer. We have only to abolish wishful thinking and believe our own eyes.10

Secular Stagnation

This unbrave new world has been emerging for some time, as US power has declined relative to other states, especially China, global liberalism has failed to deliver on its promises, and totalitarian capitalism has proven effective in leveraging globalization for economic growth and political legitimacy while exploiting technology and the state’s coercive powers to maintain internal political control. But this new era was jumpstarted by the world financial crisis of 2007, which revealed the bankruptcy of unregulated market capitalism, weakened faith in US leadership, exacerbated economic deprivation and inequality around the world, ignited growing populism, and undermined international liberal institutions. The skewed distribution of wealth experienced in most developed countries, politically tolerated in periods of growth, became intolerable as growth rates declined. A combination of aging populations, accelerating technology, and global populism/nationalism promises to make this growth decline very difficult to reverse. What Larry Summers and other international political economists have come to call “secular stagnation” increases the likelihood that illiberal globalization, multipolarity, and rising nationalism will define our future. Summers11 has argued that the world is entering a long period of diminishing economic growth. He suggests that secular stagnation “may be the defining macroeconomic challenge of our times.” Julius Probst, in his recent assessment of Summers’ ideas, explains:

…rich countries are ageing as birth rates decline and people live longer. This has pushed down real interest rates because investors think these trends will mean they will make lower returns from investing in future, making them more willing to accept a lower return on government debt as a result.

Other factors that make investors similarly pessimistic include rising global inequality and the slowdown in productivity growth…

This decline in real interest rates matters because economists believe that to overcome an economic downturn, a central bank must drive down the real interest rate to a certain level to encourage more spending and investment… Because real interest rates are so low, Summers and his supporters believe that the rate required to reach full employment is so far into negative territory that it is effectively impossible.

…in the long run, more immigration might be a vital part of curing secular stagnation. Summers also heavily prescribes increased government spending, arguing that it might actually be more prudent than cutting back – especially if the money is spent on infrastructure, education and research and development.

Of course, governments in Europe and the US are instead trying to shut their doors to migrants. And austerity policies have taken their toll on infrastructure and public research. This looks set to ensure that the next recession will be particularly nasty when it comes… Unless governments change course radically, we could be in for a sobering period ahead.12

The rise of nationalism/populism is both cause and effect of this economic outlook. Lower growth will make every aspect of the liberal order more difficult to resuscitate post-Trump. Domestic politics will become more polarized and dysfunctional, as competition for diminishing resources intensifies. International collaboration, ad hoc or through institutions, will become politically toxic. Protectionism, in its multiple forms, will make economic recovery from “secular stagnation” a heavy lift, and the liberal hegemonic leadership and strong institutions that limited the damage of previous downturns, will be unavailable. A clear demonstration of this negative feedback loop is the economic damage being inflicted on the world by Trump’s trade war with China, which— despite the so-called phase one agreement—has predictably escalated from negotiating tactic to imbedded reality, with no end in sight. In a world already suffering from inadequate investment, the uncertainties generated by this confrontation will further curb the investments essential for future growth. Another demonstration of the intersection of structural forces is how populist-motivated controls on immigration (always a weakness in the hyper-globalization narrative) deprives developed countries of Summers’ recommended policy response to secular stagnation, which in a more open world would be a win-win for rich and poor countries alike, increasing wage rates and remittance revenues for the developing countries, replenishing the labor supply for rich countries experiencing low birth rates.

Illiberal Globalization

Economic weakness and rising nationalism (along with multipolarity) will not end globalization, but will profoundly alter its character and greatly reduce its economic and political benefits. Liberal global institutions, under American hegemony, have served multiple purposes, enabling states to improve the quality of international relations and more fully satisfy the needs of their citizens, and provide companies with the legal and institutional stability necessary to manage the inherent risks of global investment. But under present and future conditions these institutions will become the battlegrounds—and the victims—of geopolitical competition. The Trump Administration’s frontal attack on multilateralism is but the final nail in the coffin of the Bretton Woods system in trade and finance, which has been in slow but accelerating decline since the end of the Cold War. Future American leadership may embrace renewed collaboration in global trade and finance, macroeconomic management, environmental sustainability and the like, but repairing the damage requires the heroic assumption that America’s own identity has not been fundamentally altered by the Trump era (four years or eight matters here), and by the internal and global forces that enabled his rise. The fact will remain that a sizeable portion of the American electorate, and a monolithically proTrump Republican Party, is committed to an illiberal future. And even if the effects are transitory, the causes of weakening global collaboration are structural, not subject to the efforts of some hypothetical future US liberal leadership. It is clear that the US has lost respect among its rivals, and trust among its allies. While its economic and military capacity is still greatly superior to all others, its political dysfunction has diminished its ability to convert this wealth into effective power.13 It will furthermore operate in a future system of diffusing material power, diverging economic and political governance approaches, and rising nationalism. Trump has promoted these forces, but did not invent them, and future US Administrations will struggle to cope with them.

What will illiberal globalization look like? Consider recent events. The instruments of globalization have been weaponized by strong states in pursuit of their geopolitical objectives. This has turned the liberal argument on behalf of globalization on its head. Instead of interdependence as an unstoppable force pushing states toward collaboration and convergence around market-friendly domestic policies, states are exploiting interdependence to inflict harm on their adversaries, and even on their allies. The increasing interaction across national boundaries that globalization entails, now produces not harmonization and cooperation, but friction and escalating trade and investment disputes.14 The Trump Administration is in the lead here, but it is not alone. Trade and investment friction with China is the most obvious and damaging example, precipitated by China’s long failure to conform to the World Trade Organization (WTO) principles, now escalated by President Trump into a trade and currency war disturbingly reminiscent of the 1930s that Bretton Woods was designed to prevent. Financial sanctions against Iran, in violation of US obligations in the Joint Comprehensive Plan Of Action (JCPOA), is another example of the rule of law succumbing to geopolitical competition. Though more mercantilist in intent than geopolitical, US tariffs on steel and aluminum, and their threatened use in automotives, aimed at the EU, Canada, and Japan,15 are equally destructive of the liberal system and of future economic growth, imposed as they are by the author of that system, and will spread to others. And indeed, Japan has used export controls in its escalating conflict with South Korea16 (as did China in imposing controls on rare earth,17 and as the US has done as part of its trade war with China). Inward foreign direct investment restrictions are spreading. The vitality of the WTO is being sapped by its inability to complete the Doha Round, by the proliferation of bilateral and regional agreements, and now by the Trump Administration’s hold on appointments to WTO judicial panels. It should not surprise anyone if, during a second term, Trump formally withdrew the US from the WTO. At a minimum it will become a “dead letter regime.”18

As such measures gain traction, it will become clear to states—and to companies—that a global trading system more responsive to raw power than to law entails escalating risk and diminishing benefits. This will be the end of economic globalization, and its many benefits, as we know it. It represents nothing less than the subordination of economic globalization, a system which many thought obeyed its own logic, to an international politics of zero-sum power competition among multiple actors with divergent interests and values. The costs will be significant: Bloomberg Economics estimates that the cost in lost US GDP in 2019- dollar terms from the trade war with China has reached $134 billion to date and will rise to a total of $316 billion by the end of 2020.19

Economically, the just-in-time, maximally efficient world of global supply chains, driving down costs, incentivizing innovation, spreading investment, integrating new countries and populations into the global system, is being Balkanized. Bilateral and regional deals are proliferating, while global, nondiscriminatory trade agreements are at an end. Economies of scale will shrink, incentivizing less investment, increasing costs and prices, compromising growth, marginalizing countries whose growth and poverty reduction depended on participation in global supply chains. A world already suffering from excess savings (in the corporate sector, among mostly Asian countries) will respond to heightened risk and uncertainty with further retrenchment. The problem is perfectly captured by Tim Boyle, CEO of Columbia Sportswear, whose supply chain runs through China, reacting to yet another ratcheting up of US tariffs on Chinese imports, most recently on consumer goods:

We move stuff around to take advantage of inexpensive labor. That’s why we’re in Bangladesh. That’s why we’re looking at Africa. We’re putting investment capital to work, to get a return for our shareholders. So, when we make a wager on investment, this is not Vegas. We have to have a reasonable expectation we can get a return. That’s predicated on the rule of law: where can we expect the laws to be enforced, and for the foreseeable future, the rules will be in place? That’s what America used to be.20

The international political effects will be equally damaging. The four structural forces act on each other to produce the more dangerous, less prosperous world projected here. Illiberal globalization represents geopolitical conflict by (at first) physically non-kinetic means. It arises from intensifying competition among powerful states with divergent interests and identities, but in its effects drives down growth and fuels increased nationalism/populism, which further contributes to conflict. Twenty-first-century protectionism represents bottom-up forces arising from economic disruption. But it is also a top-down phenomenon, representing a strategic effort by political leadership to reduce the constraints of interdependence on freedom of geopolitical action, in effect a precursor and enabler of war. This is the disturbing hypothesis of Daniel Drezner, argued in an important May 2019 piece in Reason, titled “Will Today’s Global Trade Wars Lead to World War Three,”21 which examines the preWorld War I period of heightened trade conflict, its contribution to the disaster that followed, and its parallels to the present:

Before the First World War started, powers great and small took a variety of steps to thwart the globalization of the 19th century. Each of these steps made it easier for the key combatants to conceive of a general war.

We are beginning to see a similar approach to the globalization of the 21st century. One by one, the economic constraints on military aggression are eroding. And too many have forgotten—or never knew—how this played out a century ago.

…In many ways, 19th century globalization was a victim of its own success. Reduced tariffs and transport costs flooded Europe with inexpensive grains from Russia and the United States. The incomes of landowners in these countries suffered a serious hit, and the Long Depression that ran from 1873 until 1896 generated pressure on European governments to protect against cheap imports.

…The primary lesson to draw from the years before 1914 is not that economic interdependence was a weak constraint on military conflict. It is that, even in a globalized economy, governments can take protectionist actions to reduce their interdependence in anticipation of future wars.

In retrospect, the 30 years of tariff hikes, trade wars, and currency conflicts that preceded 1914 were harbingers of the devastation to come. European governments did not necessarily want to ignite a war among the great powers. By reducing their interdependence, however, they made that option conceivable.

…the backlash to globalization that preceded the Great War seems to be reprised in the current moment. Indeed, there are ways in which the current moment is scarier than the pre-1914 era. Back then, the world’s hegemon, the United Kingdom, acted as a brake on economic closure. In 2019, the United States is the protectionist with its foot on the accelerator. The constraints of Sino-American interdependence—what economist Larry Summers once called “the financial balance of terror”—no longer look so binding. And there are far too many hot spots—the Korean peninsula, the South China Sea, Taiwan—where the kindling seems awfully dry.

Multipolarity

We can define multipolarity as a wide distribution of power among multiple independent states. Exact equivalence of material power is not implied. What is required is the possession by several states of the capacity to coerce others to act in ways they would otherwise not, through kinetic or other means (economic sanctions, political manipulation, denial of access to essential resources, etc.). Such a distribution of power presents inherently graver challenges to peace and stability than do unipolar or bipolar power configurations,22 though of course none are safe or permanent. In brief, the greater the number of consequential actors, the greater the challenge of coordinating actions to avoid, manage, or de-escalate conflicts. Multipolarity also entails a greater potential for sudden changes in the balance of power, as one state may defect to another coalition or opt out, and as a result, the greater the degree of uncertainty experienced by all states, and the greater the plausibility of downside assumptions about the intentions and capabilities of one’s adversaries. This psychology, always present in international politics but particularly powerful in multipolarity, heightens the potential for escalation of minor conflicts, and of states launching preventive or preemptive wars. In multipolarity, states are always on edge, entertaining worst-case scenarios about actual and potential enemies, and acting on these fears—expanding their armies, introducing new weapon systems, altering doctrine to relax constraints on the use of force—in ways that reinforce the worst fears of others.

The risks inherent in multipolarity are heightened by the attendant weakening of global institutions. Even in a state-centric system, such institutions can facilitate communication and transparency, helping states to manage conflicts by reducing the potential for misperception and escalation toward war. But, as Waheguru Pal Singh Sidhu argues in his chapter on the United Nations, the influence of multilateral institutions as agent and actor is clearly in decline, a result of bottom-up populist/nationalist pressures experienced in many countries, as well as the coordination problems that increase in a system of multiple great powers. As conflict resolution institutions atrophy, great powers will find themselves in “security dilemmas”23 in which verification of a rival’s intentions is unavailable, and worst-case assumptions fill the gap created by uncertainty. And the supply of conflicts will expand as a result of growing nationalism and populism, which are premised on hostility, paranoia, and isolation, with governments seeking political legitimacy through external conflict, producing a siege mentality that deliberately cuts off communication with other states.

Finally, the transition from unipolarity (roughly 1989–2007) to multipolarity is unregulated and hazardous, as the existing superpower fears and resists challenges to its primacy from a rising power or powers, while the rising power entertains new ambitions as entitlements now within its reach. Such a “power transition” and its dangers were identified by Thucydides in explaining the Peloponnesian Wars,24 by Organski (the “rear-end collision”)25 during the Cold War, and recently repopularized and brought up to date by Graham Allison in predicting conflict between the US and China.26

A useful, and consequential illustration of the inherent challenge of conflict management during a power transition toward multipolarity, is the weakening of the arms control regime negotiated by the US and the Soviet Union during the Cold War. Despite the existential, global conflict between two nuclear armed superpowers embracing diametrically opposed world views and operating in economic isolation from each other, the two managed to avoid worst-case outcomes. They accomplished this in part by institutionalizing verifiable limits on testing and deployment of both strategic and intermediate-range nuclear missiles. Yet as diplomatically and technically challenging as these achievements were, the introduction of a third great power, China, into this twocountry calculus has proven to be a deal breaker. Unconstrained by these bilateral agreements, China has been free to build up its capability, and has taken full advantage in ramping up production and deployment of intermediate-range ground-launched cruise missiles, thus challenging the US ability to credibly guarantee the security of its allies in Asia, and greatly increasing the costs of maintaining its Asian regional hegemony. As a result, the Intermediate Nuclear Force treaty is effectively dead, and the New Start Treaty, covering strategic missiles, is due to expire next year, with no indication of any US–Russian consensus to extend it. The US has with logic indicated its interest in making these agreements trilateral; but China, with its growing power and ambition, has also logically rejected these overtures. Thus, all three great powers are entering a period of nuclear weapons competition unconstrained by the major Cold War arms control regimes. In a period of rapid advances in technology and worsening great power relations, the nuclear competition will be a defining characteristic of the next decade and beyond. This dynamic will also complicate nuclear nonproliferation efforts, as both the demand for nuclear weapons (a consequence of rising regional and global insecurity), and supply of nuclear materials and technology (a result of the weakening of the nonproliferation regime and deteriorating great power relations) will increase.

Will deterrence prevent war in a world of several nuclear weapons states, (the current nuclear powers plus South Korea, Iran, Saudi Arabia, Japan, Turkey), as it helped to do during the bipolar Cold War? Some neorealist observers view nuclear weapons proliferation as stabilizing, extending the balance of terror, and the imperative of restraint, to new nuclear weapons states with much to fight over (Saudi Arabia and Iran, for example).27 Others,28 examining issues of command and control of nuclear weapons deployment and use by newly acquiring states, asymmetries in doctrines, force structures, and capabilities between rivals, the perils of variable rates in transition to weapons deployment, problems of communication between states with deep mutual grievances, the heightened risk of transfer of such weapons to non-state actors, have grave doubts about the safety of a multipolar, nuclear-armed world.29 We can at least conclude that prudence dictates heightened efforts to slow the pace of proliferation, while realism requires that we face a proliferated future with eyes wide open.

The current distribution of power is not perfectly multipolar. The US still commands the world’s largest economy, and its military power is unrivaled by any state or combination of states. Its population is still growing, despite a recent decline in birth rates. It enjoys extraordinary geographic advantages over its rivals, who are distant and live in far worse neighborhoods. Its economy is less dependent on foreign markets or resources. Its political system has proven—up to now—to be resilient and adaptable. Its global alliance system greatly extends its capacity to defend itself and shape the world to its liking and is still intact, despite growing doubts about America’s reliability as a security guarantor. Based on these mostly material and historical criteria, continued American primacy would seem to be a good bet, if it chooses to use its power in this way.30

So why multipolarity? The clearest and most frequently cited evidence for a widening distribution of global power away from American unipolarity is the narrowing gap in GDP between the US and China. The IMF’s World Economic Outlook forecasts a $0.9 trillion increase in US GDP for 2019–2020, and a $1.3 trillion increase for China in the same period.31 Many who support the American primacy case argue that GDP is an imperfect measure of power, that Chinese GDP data is inflated, that its growth rates are in decline while Chinese debt is rapidly increasing, and that China does poorly on other factors that contribute to power—its low per capita GDP, its political succession challenges, its environmental crisis, its absence of any external alliance system. Yet GDP is a good place to start, as the single most useful measure and long-term predictor of power. It is from the overall economy that states extract and apply material power to leverage desired behavior from other states. It is true that robust future Chinese growth is not guaranteed, nor is its capacity to convert its wealth to power, which is a function of how well its political system works over time. But this is equally the case for the US, and considering recent political developments is not a given for either country.

As an alternative to measuring inputs—economic size, political legitimacy, technological innovation, population growth—in assessing relative power and the nature of global power distribution, we should consider outputs: what are states doing with their power? The input measures are useful, possibly predictive, but are usually deployed in the course of making a foreign policy argument, sometimes on behalf of a reassertion of American primacy, sometimes on behalf of retrenchment. As such, their objectivity (despite their generous deployment of “data”) is open to question. What is undeniable, to any clear-eyed observer, is a real decline in American influence in the world, and a rise in the influence of other powers, which predates the Trump administration but has accelerated into America’s free fall over the last four years. This has produced a de facto multipolarity, whether explainable in the various measures of power—actual and latent—or not. This decline results in part from policy mistakes: a reckless squandering of material power and legitimacy in Iraq, an overabundance of caution in Syria, and now pure impulsivity. But more fundamentally, it is a product of relative decline in American capacity—political and economic—to which American leadership is adjusting haphazardly, but in the direction of retrenchment/restraint. It is highly revealing that the last two American presidents, polar opposites in intellect, temperament and values, agreed on one fundamental point: the US is overextended, and needs to retrench. The fact that neither Obama nor Trump (up to this point in his presidency) believed they had the power at their disposal to do anything else, tells us far more about the future of American power and policy—and about the emerging shape of international relations—than the power measures and comparisons made by foreign policy advocates.

Observation of recent trends in US versus Russian relative influence prompts another question: do we understand the emerging characteristics of power? Rigorously measuring and comparing the wrong parameters will get us nowhere at best and mislead us into misguided policies at worst. How often have we heard, with puzzlement, that Putin punches far above his weight? Could it be that we misunderstand what constitutes “weight” in the contemporary and emerging world? Putin may be on a high wire, and bound to come crashing down; but the fact is that Russian influence, leveraging sophisticated communications/social media/influence operations, a strong military, an agile (Putin-dominated) decision process, and taking advantage of the egregious mistakes by the West, has been advancing for over a decade, shows no sign of slowing down, and has created additional opportunities for itself in the Middle East, Europe, Asia, Latin America, the Arctic. It has done this with an economy roughly the size of Italy’s. There are few signs of a domestic political challenge to Putin. His external opponents are in disarray, and Russia’s main adversary is politically disabled from confronting the problem. He has established Russia as the Middle East power broker. He has reached into the internal politics of his Western adversaries and influenced their leadership choices. He has invaded and absorbed the territory of neighboring states. His actions have produced deep divisions within NATO. Again, simple observation suggests multipolarity in fact, and a full explanation for this power shift awaiting future historians able to look with more objectivity at twenty-first-century elements of power.

When that history is written, surely it will emphasize the extraordinary polarization in American politics. Was multipolarity a case of others finding leverage in new sources of power, or the US underutilizing its own? The material measures suggest sufficient capacity for sustained American primacy, but with this latent capacity unavailable (as perceived, I believe correctly, by political leadership) by virtue of weakening institutions: two major parties in separate universes; a winnertake-all political mentality; deep polarization between the parties’ popular bases of support; divided government, with the Presidency and the Congress often in separate and antagonistic hands; diminishing trust in the permanent government, and in the knowledge it brings to important decisions, and deepening distrust between the intelligence community and policymakers; and, in Trump’s case, a chaotic policy process that lacks any strategic reference points, mis-communicates the Administration’s intentions, and has proven incapable of sustained, coherent diplomacy on behalf of any explicit and consistent set of policy goals.

Rising Nationalism/Populism/Authoritarianism

The evidence for these trends is clear. Freedom House, the go-to authority on the state of global democracy, just published its annual assessment for 2020, and recorded the fourteenth consecutive year of global democratic decline and advancing authoritarianism. This dramatic deterioration includes both a weakening in democratic practice within states still deemed on balance democratic, and a shift from weak democracies to authoritarianism in others. Commitment to democratic norms and practices—freedom of speech and of the press, independent judiciaries, protection of minority rights—is in decline. The decline is evident across the global system and encompasses all major powers, from India and China, to Europe, to the US. Right-wing populist parties have assumed power, or constitute a politically significant minority, in a lengthening list of democratic states, including both new (Hungary, Poland) and established (India, the US, the UK) democracies. Nationalism, frequently dismissed by liberal globalization advocates as a weak force when confronted by market democracies’ presumed inherent superiority, has experienced a resurgence in Russia, China, the Middle East, and at home. Given the breadth and depth of right-wing populism, the raw power that promotes it—mainly Russian and American—and the disarray of its liberal opponents, this factor will weigh heavily on the future.

The major factors contributing to right-wing populism and its global spread is the subject of much discussion.32 The most straightforward explanation is rising inequality and diminished intergenerational mobility, particularly in developed countries whose labor-intensive manufacturing has been hit hardest by the globalization of capital combined with the immobility of labor. Jobs, wages, economic security, a reasonable hope that one’s offspring has a shot at a better life than one’s own, the erosion of social capital within economically marginalized communities, government failure to provide a decent safety net and job retraining for those battered by globalization: all have contributed to a sense of desperation and raw anger in the hollowed-out communities of formerly prosperous industrial areas. The declining life expectancy numbers33 tell a story of immiseration: drug addition, suicide, poor health care, and gun violence. The political expression of such conditions of life should not be surprising. Simple, extremist “solutions” become irresistible. Sectarian, racial, regional divides are strengthened, and exclusive identities are sharpened. Political entrepreneurs offering to blow up the system blamed for such conditions become credible. Those who are perceived as having benefited from the corrupt system—long-standing institutions of government, foreign countries and populations, immigrants, minorities getting a “free ride,” elites—become targets of recrimination and violence. The simple solutions of course, don’t work, deepening the underlying crisis, but in the process politics is poisoned. If this sounds like the US, it should, but it also describes major European countries (the UK, France, Italy, Germany, Poland, Hungary, the Czech Republic), and could be an indication of things to come for non-Western democracies like India.

We have emphasized throughout this chapter the interaction of four structural forces in shaping the future, and this interaction is evident here as well. Is it merely coincidence that the period of democratic decline documented by Freedom House, coincides precisely with the global financial and economic crisis? Lower growth, increasing joblessness, wage stagnation, superimposed on longer-term widening of inequality and declining mobility, constitute a forbidding stress test for democratic systems, and many continue to fail. And if we are correct about secular stagnation, the stress will continue, and authoritarianism’s fourteen-year run will not be over for some time. The antidemocratic trend will gain additional impetus from the illiberal direction of globalization, with its growth suppressing protectionism, weaponization of global economic exchange, and weakening global economic institutions. Multipolarity also contributes, in several ways. The former hegemon and author of globalization’s liberal structure has lost its appetite, and arguably its capacity, for leadership, and indeed has become part of the problem, succumbing to and promoting the global right-wing populist surge. It is suffering an unprecedented decline in life expectancy, and recently a decline in the birth rate, signaling a degree of rot commonly associated with a collapsing Soviet Union. While American politics may once again cohere around its liberal values and interests, the time when American leadership had the self-confidence to shape the global system in its liberal image is gone. It may build coalitions of the like-minded to launch liberal projects, but there will be too much power outside these coalitions to permit liberal globalization of the sort imagined at the end of the Cold War. In multipolarity, the values around which global politics revolve will reflect the diversity of major powers, their interests, and the norms they embrace. Convergence of norms, practices, policies is out of the question. Global collective action, even in the face of global crises, will be a long shot. To expect anything else is fantasy

Unbrave New World and Future Challenges

At the outset of this chapter we described these structural forces as interacting to produce more conflict and diminished prosperity. We also predicted a world with shrinking collective capacity to address new challenges as they arise. What specifically will such a world look like? We address below three principal challenges to global problem solving over the next decade.

Interstate Conflict

In the world experienced by most readers of this volume, conflict is observed within weak states, sometimes promoted by regional competitors, by terrorist groups, or by great powers, acting through surrogates or by indirect means. Sometimes, as in Syria, this conflict spills over to contiguous states and contributes to regional instability, and challenges other regions to respond effectively, a challenge that Europe has not met. Much of this will continue, but the global significance of such local conflicts will be greatly magnified by increasing great power conflict, which will feed—rather than manage or resolve—local instabilities and will in turn be exacerbated by them. Great powers will jockey for advantage, support their local partners, escalate preemptively. Conflicts initially confined to failing states or unstable regions will be redefined by great powers as global in scope and significance.

This tendency of states to view local conflicts in the context of a zero-sum, global struggle for power is familiar to students of the Cold War, but now with the additional challenges to collective action, expanded uncertainty and worst-case thinking associated with the power transition to multipolarity. We can easily observe increased conflict in US–China relations, as we will in US–Russia relations as future US administrations try to make up for ground lost during the Trump presidency, especially in the Middle East. We can observe it among powerful states with mutual historical grievances, now with a weakening presence of the hegemonic security guarantor and having to consider the renationalization of their defense: Japan-South Korea, Germany-France. We can observe it among historical rivals operating in rapidly changing security landscapes: India-China. We can observe it within the Middle East, as internal rivalries are appropriated by regional powers in a contest for regional dominance. We can observe it clearly in Syria, where the regime’s violent suppression of Arab Spring resistance led to all-out civil war, attracted outside support to proxy forces by aspiring regional hegemons Saudi Arabia and Iran, enabled the rise of ISIS, and eventually to great power intervention, principally by Russia. In a world of effective great power collaboration or American primacy, the Syrian civil war might have been settled through power sharing or partition, or if not, contained within Syria. The collapse of Yugoslavia, occurring during a period of US “unipolarity” and managed effectively, demonstrates the possibilities. Instead, with the US retrenching, Middle East rivals unconstrained by great powers, and great power competition rising, the Syria civil war was fed by outside powers, then metastasized into the region, and—in the form of refugee flows—into Europe, fundamentally altering European politics. Libya may be at the early stages of this scenario.

This is not the end of the Syria story. Russia has established itself as a major player in Syria and the Middle East’s power broker, the indispensable country with leverage throughout the region. China is poised to reap the financial and power benefits of Syrian reconstruction. The US has just demonstrated, in its act of war against the Iranian regime, its willingness, without consultation, to put its allies’ security in further jeopardy, accentuating the risks of security ties with Washington and generating added opportunities for Russia and China. The purpose here is not to critique US policy, but to point out the dramatically shifting power balance in a critical region, toward multipolarity. The dangers of such a shift will become apparent as some future US president attempts to reassert US influence in the region and finds a crowded playing field.

Can a multipolar distribution of power among several states whose interests, values, and political practices are divergent, all experiencing bottom-up nationalist pressures, all seeking advantages in the oversupply of regional instability, be made to work? I think not. Will this more dangerous world descend into direct military confrontation between great powers, and could such confrontation lead to use of nuclear weapons? Here the question becomes, what will this more dangerous world actually look like; what instruments of coercion will be available to states as technology change accelerates; how will states employ these instruments; how will deterrence work (if at all) among several states with large but unequal levels of destructive capacity, weak command, and control, disparate— or opaque—strategies and simmering rivalries; can conflict management work in a world of weak institutions? The collapse of the Cold War era nuclear arms control regime, the threat to the Non-Proliferation Treaty represented by the demise of the JCPOA, and multiple indications of an accelerating nuclear arms race among the three principle powers, augurs badly. Given the structural forces at play, and without predicting the worst, we are indeed entering perilous times.

Global Poverty and Inequality

Despite the challenges of volatility and disruptive change inherent in globalization, the world under American liberal leadership has managed a dramatic reduction of extreme poverty. According to World Bank estimates, in 2015, 10 percent of the world’s population lived on less than $1.90 a day, down from nearly 36 percent in 1990.34 In fact, as of September 2018, half the world is now middle class or wealthier.35 The uneven success of the UN Millennium Development Goals (MDGs) exemplifies this achievement, and demonstrates what is possible when open markets are managed through strong global institutions, effective leadership and interstate collaboration. What this liberal hegemonic system did not achieve, however, was a fair distribution of the gains from globalization within states, and among those states that for various reasons were not full participants in this system.

This record of partial achievement leaves us with a full agenda for the next fifteen years, but without the hegemonic leadership, strong institutions, ascendant liberalism or robust global growth that enabled previous gains. There are powerful reasons to question the sustainability of these poverty reduction gains, leading to doubts about the realization of the Sustainable Development Goals, which have replaced the MDGs as global development targets.36 (See Jens Rudbeck’s chapter and Sidhu’s UN chapter for SDGs). Skeptics have pointed to slowing global growth, specifically in China, whose demand for imported commodities was a major factor in developing country growth and job creation; growing protectionism in developed country markets, fueled by bottom-up forces of nationalism, and from top-down by a weakened global trading regime and increased geopolitical rivalry; the effects of accelerating climate change on agriculture, migration and communal conflict in poor countries; and the growth burst among poor countries from the rapid transition to more efficient use of resources, a transition that is now slowing down.37

Perhaps the greatest concern in this scenario is a general deterioration in the developing country foreign investment climate. Foreign direct investment (FDI) has been a major contributor to growth, job creation, and poverty alleviation among poor countries. It has incentivized growthfriendly policies, reduced corruption, introduced technology and effective management practices, and linked poor countries to foreign markets through global supply chains.38 It has stimulated growth of indigenous manufacturing and service companies to supply new foreign investments.

It has been the major cause of economic convergence between rich and poor countries. From 2000 to 2009, developing economies’ growth rates were more than four percentage points higher than those of rich countries, pushing their share of global output from just over a third to nearly half.39 However, FDI flows into poor countries are imperiled by the structural forces discussed here. Political instability arising from slower growth and environmental stress will increase investors’ perception of higher risk, reinforcing their developed country bias. Protectionism among developed countries will threaten the global market access upon which manufacturing investment in developing countries is premised, causing firms to pare back their global supply chains. As companies retrench from direct investment in poor countries, the appeal to those countries of Chinese debt financed infrastructure projects, under the Belt-Road Initiative with little or no conditionality, but at the risk of “debt traps,” will increase.

Global Warming

The question posed at the beginning of this section is whether the international system, evolving toward multipolarity and rising nationalism, will find the collective political capital to confront challenges as they arise. Global warming is the mother of all challenges, and the weakness in the system’s capacity to respond is clear. With the two major political/economic powers and greenhouse gas emitters locked in deepening geopolitical conflict (and with one of them locked in climate change denial, possibly through 2024), the chances of significantly slowing global warming or even ameliorating its effects are very slim. We are reduced to the default option, nation-specific adaptation to climate change, which will impose rising human, political and economic costs on all, and will widen the gap between rich countries with adaptive capacity (of varying degrees), and the poor, who will suffer deteriorating economic, political, and social conditions. (For a contrary, optimistic view see Michael Shank’s chapter, which credits new actors—like cities—as playing a more constructive role in climate mitigation.) This would bring to a close liberal globalization’s greatest achievement; the raising of 1.1 billion people out of extreme poverty since 1990,40 with all its associated gains in quality of life (in the WHO Africa region, for example, life expectancy rose by 10.3 years between 2000 and 2016, driven mainly by improvements in child survival and expanded access to antiretrovirals for treatment of HIV).41

Several forces are at work here. The problem itself is graver—in magnitude and in rate of worsening—than predicted by climate scientists. The UN Intergovernmental Panel on Climate Change (IPCC), the major source of information on global warming, has consistently underpredicted the rate of climate deterioration. This holds true even for its “worst-case scenarios,” meaning that what was meant as a wake-up call has in fact reinforced complacency.42 (see Michael Shank’s chapter for further discussion of climate change). The IPCC, in its 2019 report, has tried to undo the damage by emphasizing the acceleration in the rate of warming and its effects, the only partially understood dynamic of climate change, and—given wide uncertainty—the possibility of unpleasant surprises yet to come. This strengthens the scientific case for urgency—to both severely limit greenhouse gas emissions, and to increase investment in ameliorating the effects.

Unfortunately, the crisis comes at a moment when the climate for collective action is ice cold. Geopolitical competition incentivizes states to out produce each other, regardless of the environmental effects. Multipolarity complicates collective action. Economic stagnation mandates job creation, making regulation politically toxic. Bottom-up nationalism/populism causes states to pursue “relative gains,” meaning that if the nation is seen as gaining in a no-holds-barred economic competition with others, the negative environmental effects can be tolerated. A post-Trump presidency would help, with the US rejoining the Paris Agreement, and lending its weight to tighter regulation, increased R and D, and stronger economic incentives to reduce carbon emissions. Keep in mind, however, that President Obama was fully behind such efforts, but in a deeply polarized America was unable to implement measures needed to fulfill the Paris obligations through legislation, and his executive orders to do this were swiftly overturned by Trump.

Conclusion

It may be tempting to hope that post-Trump, the US can regain its global leadership and exert its considerable power in a liberal direction, but with enough self-awareness of its relative decline to share responsibility with others. This was, I believe, the broad direction of the Obama strategy, evidenced by the JCPOA and the Trans-Pacific Partnership: liberal, collective solutions to global problems, as US dominance receded.

This would constitute an optimistic scenario, and it confronts two major problems: can US internal politics support it (can, for example, the country legislate controls on carbon, essential for the global credibility and durability of such commitments); and is the world ready to reengage with American leadership, given the damage to its reputation and the structural forces discussed in this chapter?

My educated guess is no, on both counts. The rot within is extensive, the concrete evidence clear in the economic inequality/immobility numbers, the life expectancy numbers, the deep political polarization, between the two major parties, between regions, between cities and rural areas. We are in fact a long way from fitness for global leadership, and the recognition of this by others will accelerate the decline of American influence. The rest of the world is well on its way toward adjusting to post-American hegemony, some by renationalizing their defense, or by cutting deals with adversaries, by building new alliances or by seizing new opportunities for influence in the vacuum left by American retrenchment. The evidence for this will accumulate. Observe the current and emerging Middle East, where all these post-hegemonic strategies are visible.

#### Expectation of strong growth stops nuclear war and nationalist bloc formation.

Kampf ’20 [David; June 16; PhD Fellow at the Center for Strategic Studies at The Fletcher School, MA in International Affairs from Columbia University; World Politics Review, “How COVID-19 Could Increase the Risk of War,” https://www.worldpoliticsreview.com/articles/28843/how-covid-19-could-increase-the-risk-of-war]

Other theories posit that economic bonds between countries have limited wars in recent decades. Dale Copeland, a professor of international relations at the University of Virginia, has argued that countries work to preserve ties when there are high expectations for future trade, but war becomes increasingly possible when trade is predicted to fall. If globalization brought peace, the recent wave of far-right nationalism and populism around the world may increase the chances of war, as tariffs and other trade barriers go up—mostly from the United States under President Donald Trump, who has launched trade wars with allies and adversaries alike.

The coronavirus pandemic immediately elicited further calls to reduce dependence on other countries, with Trump using the opportunity to pressure U.S. companies to reconfigure their supply chains away from China. For its part, China made sure that it had the homemade supplies it needed to fight the virus before exporting extras, while countries like France and Germany barred the export of face masks, even to friendly nations. And widening economic inequalities, a consequence of the pandemic, are not likely to enhance support for free trade.

This assault on open trade and globalization is just one aspect of a decaying liberal international order, which, its proponents argue, has largely helped to preserve peace between nations since World War II. But that old order is almost gone, and in all likelihood isn’t coming back. The U.N. Security Council appears increasingly fragmented and dysfunctional. Even before Trump, the world’s most powerful country ratified fewer treaties per year under the Obama administration than at any time since 1945.

Trump’s presidency only harms multilateral cooperation further. He has backed out of the Paris Agreement on climate change, reneged on the Iran nuclear deal, picked fights with allies, questioned the value of NATO and defunded the World Health Organization in the middle of a global health crisis. Hyper-nationalism, rather than international collaboration, was the default response to the coronavirus outbreak in the U.S. and many other countries around the world.

It’s hard to see the U.S. reluctance to lead as anything other than a sign of its inevitable, if slow, decline. The country’s institutionalized inequalities and systemic racism have been laid bare in recent months, and it no longer looks like a beacon for others to follow. The global balance of power is changing. China is both keen to assert a greater leadership role within traditionally Western-led institutions and to challenge the existing regional order in Asia. Between a rising China, revanchist Russia and new global actors, including non-state groups, we may be heading toward an increasingly multipolar or nonpolar world, which could prove destabilizing in its own right.

Finally, the pacifying effect of nuclear weapons could be waning. While vast nuclear arsenals once compelled the United States and the Soviet Union to reach arms control agreements, old treaties are expiring and new talks are breaking down. Mistrust is growing, and the chance of an unwanted U.S.-Russia nuclear confrontation is arguably as high as it has been since the Cuban missile crisis.

The theory of nuclear peace may no longer hold if more countries are tempted to obtain their own nuclear deterrent. Trump’s decision to abandon the Iran nuclear deal, for one thing, has only increased the chance that Tehran will acquire nuclear weapons. It’s almost easy to forget that, just a few short months ago, the United States and Iran were one miscalculation or dumb mistake away from waging all-out war. And despite Trump’s efforts to negotiate nuclear disarmament with Kim Jong Un’s regime in Pyongyang, it is wishful thinking to believe North Korea will give up its nuclear weapons. At this point, negotiators can only realistically try to ensure that North Korea’s nuclear menace doesn’t get even more potent.

In other words, by turning inward, the United States is choosing to leave other countries to fend for themselves. The end result may be a less stable world with more nuclear actors.

If leaders are smart, they will take seriously the warning signs exposed by this global emergency and work to reverse the drift toward war.

If only one of these theories for peace were worsening, concerns would be easier to dismiss. But together, they are unsettling. While the world is not yet on the brink of World War III and no two countries are destined for war, the odds of avoiding future conflicts don’t look good.

The pandemic is already degrading democracies, harming economies and curtailing international cooperation, and it also seems to be fostering internal instability within states. Rachel Brown, Heather Hurlburt and Alexandra Stark argue that the coronavirus could in fact sow more civil conflict. If this proves accurate, the increase in civil wars is likely to lead to more external meddling, and these next proxy wars could soon precipitate all-out international conflicts if outsiders aren’t careful. With the usual deterrents to conflict declining around the world, major wars could soon return.

#### Serial collusion devastates innovation in the chemical sector.

Kovacic et al. ’21 [William, Robert Marshall, and Michael Meurer; 2021; Global Competition Professor of Law and Policy at George Washington University Law School; Distinguished Professor of Economics at Pennsylvania State University; Professor of Law at Boston University; Boston University School of Law Research Paper Series, “Patents and Price Fixing by Serial Colluders,” No. 21]

In a recent article on price fixing, we coined the term “serial colluder” to designate multi-product firms that have participated in many cartels, involving a range of participants, and initiated at different dates.15 Several chemical firms meet this definition because of their participation in at least thirty different chemical cartels spanning at least three decades.16 Our earlier article also addressed the business model of serial colluders and the failure of anti-cartel law to deter such behavior. In some cases, weak monitoring and high-powered incentive payments to product division managers may have fostered multiple cartels without encouragement from, or even contrary to the instructions of, upper management. This “rogue manager” explanation of serial collusion is often invoked by corporate directors seeking a story that deflects blame away from them. A more troubling explanation for serial collusion is that price fixing is an integral part of the business model of certain firms, and high-level managers advocate for and assist with collusion throughout the firm. We believe serial colluders in certain industries have run “portfolios of cartels.” In support of this “business model” explanation, in previous work we presented various kinds of indirect evidence that serial colluders in the chemical industry have indeed run a portfolio of cartels.17 Unaddressed in that previous work is an examination of how serial colluders may use patents and patent licensing schemes to initiate or maintain a cartel.

In Section I of this paper, we find that serial colluders increased patenting during the duration of their cartels, which is consistent with the theory that these firms use new patents to support cartelization. The magnitude of this increase is above and beyond incremental increases in patenting over time. We also find that “core” serial colluders (but not other major serial colluding chemical firms) increased patenting on products that they did not produce but that were being cartelized by their fellow colluders, which is consistent with the view that serial colluders engage in reciprocal practices across distinct markets.18 On the whole, our analysis of patenting practices for serial colluders in the chemical space suggests ongoing use of patents to initiate or maintain cartels, a practice that may apply to other industries with serial colluders as well.

Finding that the empirical data support our hypothesis of serial colluders using patents to create and maintain cartels, we next probe in Sections II and III reasons for why this conduct might evade agency enforcement and effectively help to coordinate cartels. Unlike the older cartels that openly used patents to directly restrain output, modern serial colluders running a portfolio of cartels potentially use patents in ways that are indirect and less likely to be noticed by private plaintiffs and government enforcers. We then explore how cartel participants in the modern era (excepting pay-for-delay cases like Actavis) appear to use patents to deter entry into cartelized markets, facilitate intrafirm communications and actions in support of collusive conduct, and communicate with other serial colluders about their portfolio of cartels under the guise of discussing their portfolio of patent licenses.

For the remainder of the Article, we discuss how the existing antitrust jurisprudence regarding patents and price fixing requires major upgrades to account for the dramatic modern improvements in our understanding of the economics of collusion. In older cases, judges recognized that firms could use patent licenses directly to restrict output, raise prices, or boost competitors’ marginal costs, 19 but they may not have appreciated the many indirect ways that patents can increase cartel stability and profitability. As discussed in greater detail below, patents provide an avenue for ongoing communication among rivals about output and pricing. Patent pools and cross-licensing arrangements are especially useful for organizing cartels across product types. Furthermore, licensing regimes may permit a firm to organize supportive resources within the firm without raising legal compliance concerns.

Anticipating these benefits to cartel formation and maintenance, this Article goes on to suggest that serial colluders may engage in strategic patenting. That is, they procure patents to advance cartel goals rather than to promote innovation. We present data on global patent procurement by price fixers in the chemical industry that is consistent with this view. Importantly, firms managing a portfolio of cartels can use patents in a reciprocal way to stabilize cartels across markets where not all firms participate as producers in each market. Within the network of chemical cartels, for example, we see evidence that certain firms use patents to promote cartels in markets for products they do not produce. Firms may use the threat of a patent lawsuit to punish deviators and discourage outsiders from attempting to enter a cartelized market. They may also use patent licenses to audit licensee sales and monitor compliance with cartel rules. One firm might perform such a service for other firms in the collusive network with the expectation that the non-participant would get similar help managing their own portfolio of cartels from other serial colluders in the future.

Further, in this Article, we probe deeply into the ways serial colluders can coordinate their patent practices to enhance cartel profits and stabilize their cartels. Our previous work on serial collusion documented that modern anti-collusion enforcement has not adequately deterred massive, prolonged multi-market price-fixing schemes.20 We also explained how various forms of reciprocity among serial colluders increased their cartel profits and made cartels more resilient.21 We expand on this topic with respect to the use of patents for cartelization, which we touched on only briefly in previous work.

This Article also describes gaps in existing antitrust enforcement and scholarly analysis of patenting practices. Recognition of serial collusion helps us to identify further flaws in the conventional treatment of patent licenses that allegedly facilitate price fixing. As one example, case law favors vertical patent licenses by applying rule of reason analysis to restrictions that could earn per se condemnation if organized as horizontal licenses.22 Such deference stems partly from worries that anti-collusion enforcement could weaken returns to patents and discourage research and innovation, as well as concerns that there may be legitimate reasons for suppliers, manufacturers, retailers to coordinate some activities. Yet, past practice of serial colluders show that firms can and do evade per se condemnation by simply organizing a middle man to stand as an upstream patent pool organizer. Thus, we reject such deference for vertically organized patent licenses in the context of serial colluders that are managing a portfolio of cartels, because what appears to be a vertical relationship is often part of the network of connections among serial colluders. Similarly, the leading scholarly commentary on patents and price fixing suggests that socially desirable licenses can be sorted from socially harmful licenses by determining whether significant rents flow to the licensor.23 This test may be effective in the context of an isolated cartel affecting a single market.24 As we explain in Section IV, this test has little or no value in the context of serial collusion where the firms are managing a portfolio of cartels.

Finally, in this Article, we provide additional policy recommendations tailored to the abuse of patents by serial colluders. Our earlier work lays out various reforms to anti-collusion policy that could mitigate the harms of serial collusion. In Section V, we go further and explain how certain patent-related behaviors by firms that do not participate directly in cartelizing a particular market can be used to infer collusion in that market (when the outsider is part of a network of serial colluders). We also discuss penalties and liability that antitrust and patent agencies should impose on firms that use their patents to facilitate collusion by others. Specifically, we argue for generous application of the patent misuse defense to render unenforceable patents used to facilitate price fixing.25 Entry would be easier and patent-based cartel punishments would be eliminated if cartel patents are left unenforceable. Finally, we identify possible adjustments in the institutional arrangements by which the federal antitrust enforcement agencies address the use of patents and patent licensing to facilitate collusion.

This Article is organized as follows. Section I presents empirical evidence that serial collusion is a serious problem, that serial colluders in the chemical industry use the patent system intensively in ways that suggest strategic patenting, and that their patenting behavior is consistent with their use of patents to enhance multi-market price fixing. Section II considers the evolution of antitrust doctrine and policy related to patent assertion and licensing as collusive devices. Notwithstanding existing strictures, this section reviews how patent practices can facilitate cartelization. Section III turns to the role that patents can play in supporting serial collusion. Section IV discusses the modernization of doctrines related to patents and price fixing in response to the threat of serial collusion. Section V offers policy recommendations and additional concluding comments.

I. Serial Collusion and Patents: Case Study in the Global Chemical Industry

Serial collusion in the chemical industry dates back to the 1880s and has reappeared in most decades since then.26 German chemical firms have been prominent price-fixers and often cartel ring-leaders, but they have been joined by chemical firms from the United States, England, France, Belgium, the Netherlands, Canada, Switzerland, South Korea, and Japan.27 Dozens of different chemical products have been affected by price fixing at some point.28 Historically, some of these collusive agreements were regional; others were global. Some were short-lived; others spanned decades. This history, and the specific role of patents to instituting and maintaining cartels in the global chemicals market, is described below.

A. Historical and Modern Cartelization of the Global Chemical Industry

Patents played a significant role in chemical cartels during the first half of the twentieth century. 29 Margaret Levenstein observes that “[d]uring most of the 30 years preceding World War I, bromine producers in the United States and Europe colluded, pooling output, dividing up markets, and raising prices.”30 In the period leading up to World War II, German chemical firms engaged in a variety of practices that Heinrich Kronstein has called “monopolizing by patents.” 31 One technique employed by the “combine” of chemical companies was to direct the research arm of each participant to procure as many patents as possible, to use them for strategic ends.32 From his study of patents and cartelization in 1920s Germany, Kronstein reported that “[m]ore and more the chemical industry began to apply for patents on practically everything. The research laboratories of the few remaining chemical works, connected among themselves by cartel and working agreements, systematically studied entire fields and closed them by a large number of patents.”33 In fields such as plastics and pharmaceuticals, “[e]ach publication in any chemical review or each patent application of any applicant in any country was given to the staff of the research laboratory to find anything that could be patented, no matter if the patent was a patent of evasion or supplement or protection against other inventors.”34 This phenomenon Kronstein described resembles the pattern of recent patenting behavior in the chemical sector we document below—where patenting activity by cartel participants increases dramatically during the period of illegal collaboration for the purpose of consolidating market share for existing firms and keeping out entrants.35

A second method documented by Kronstein and other researchers involves the extensive use of patent licensing agreements among major U.S. and foreign chemical producers and their subsidiaries to establish effective networks for global cartelization.36 Kronstein reports that in the decades leading up to World War II, “[t]he participation of an American enterprise in a world cartel chiefly through the device of patent exchange became very common.”37 In 1946, George Stocking and Myron Watkins reported “that a division of market territories for products coming within the scope of [cartel] patents and secret processes in a given field usually entail[ed] a complete division of territories for all related products.”38

A third method of cartelization involved the use of multiple licensing arrangements to cartelize entire domestic markets. In the late 1930s, the DOJ successfully challenged Ethyl Gasoline Company for creating an elaborate system of licensing arrangements for the production and use of tetra-ethyl lead to stabilize prices for motor fuel. 39 In another prominent American example of the technique applied outside the chemical sector, in the 1940s, the DOJ prosecuted United States Gypsum for using minimum price terms in patent licenses to cartelize the gypsum wallboard industry.40 For about a decade, Gypsum had granted licenses with largely identical price restrictions to nearly all of the industry’s numerous firms.41 In upholding the government’s challenge to Gypsum’s licensing terms, the Supreme Court observed, “the industry is completely regimented, the production of competitive unpatented products suppressed, a class of distributors squeezed out, and prices on unpatented products stabilized.”42

The rash of chemical industry cartelization has continued to modern times. In the three decades since 1980, the European Commission (EC) prosecuted chemical producers for collusion in 32 separate markets. 43 Notable American antitrust cases brought against chemical producers during this period ended cartels in the markets for lysine, citric acid, and vitamin C. 44 Since 2010, the Korean Fair Trade Commission (KFTC) fined participants in a chemical additives cartel. 45 Today, the EC is investigating an ethylene cartel, 46 and a massive investigation of serial collusion by generic drug companies is ongoing in the United States.47 Whereas the scope of these investigations has not focused on what role patents may have played in helping to facilitate these cartels, we suspect that patents did play a role.48 We explore this conjecture by examining the patenting behavior of colluding firms before, during, and after agency enforcement to explore whether these firms may have pursued patents for strategic ends.

#### Extinction.

Danielpour ’14 [Steven; April 2014; Director of Specifications at HOK, Professor at the Pratt Institute; PaintSquare, “Sustainable Coatings: Shifting the Paradigm,” https://www.paintsquare.com/archive/?fuseaction=view&articleid=5271]

New technologies and processes will help deliver the innovations needed to respond to mankind’s greatest challenges, says HOK’s firmwide director of specifications.

Whether you’re an architect or facility owner interested in ensuring healthy buildings and communities, a contractor navigating the many shades of “green” coatings or a supplier responding to market demand for these coatings, sustainability matters.

But what makes a coating “sustainable” in the built environment, and why should we care?

Wikipedia defines sustainability as “a characteristic of a process or state that can be maintained at a certain level indefinitely.” Production, distribution and application of sustainable coatings must meet current needs without compromising our ecosystems’ ability to sustain future populations.

Continuing to build the way we have built, using the materials we have used for centuries, is no longer viable in light of diminishing energy, water and other resources. Key megatrends, including population growth, climate change and a proliferation of information, make sustainable coatings all the more critical.

So it’s exciting to see the industry responding with advanced technologies that are healthier for building occupants and the environment, while achieving high performance and durability. We see it in such innovations as the newest generation of PVDF (polyvinylidene fluoride) coatings, polysiloxane coating systems, advanced anti-microbials and more.

And just ahead we can expect to see phase-changing coatings that will respond chemically to cooler or warmer conditions, for instance, to improve energy efficiency in the building envelope. We may see roofing materials that reflect and absorb heat as appropriate, using phase-changing materials and nanotechnology.

Such cutting-edge technologies, along with processes that reduce waste, reuse byproducts and allow reformulation into new products, promise game-changing improvements for coatings.

Let’s take a closer look at the drivers that will make sustainable coatings increasingly important, and the processes and technologies on the horizon.

Responding to Dwindling Resources

The logic is simple: If we continue to consume natural resources faster than they can be replenished, and if we produce wastes for future generations to deal with, we’ll have a harder and harder time maintaining life on Earth as we know it.

Scientific research on species extinction makes it clear that human survival depends on maintaining our ecological cycle, as well as those of other species and their habitats. Yet we’re barreling like a runaway train toward depleting some key resources.

Petroleum: Petrochemicals, a necessary feedstock for high-performance coatings, derive from fossil fuels that took millions of years to create; they are not readily replenished. Sustainable resource management requires that we conserve irreplaceable resources through closed-loop manufacturing, reusing manufacturing byproducts and recycling waste into new products.

Water Resources: Only 3 percent of the Earth’s water is potable, and most of this supply is locked in the polar ice cap. Just 0.003 percent of the world’s water is readily available for human consumption, and 16 percent of that is used to manufacture building materials and construct buildings. Worse yet, due to pollution, 40 percent of streams, 45 percent of lakes and 50 percent of estuaries in the United States were deemed not clean enough to support fishing and swimming in a 2000 Environmental Protection Agency study. The Index of Watershed Indicators reports that only 15 percent of our watershed has relatively good water quality.

Forests: Rain forests play an important role in maintaining Earth’s air quality, absorbing carbon dioxide emissions and VOCs (volatile organic compounds), while replenishing the air with oxygen. Statistics show that the annual rate of global deforestation is equal to an area the size of the state of Georgia. This is critical, because it has been estimated that when more than 70 percent of an ecosystem is lost, the remainder may be unable to sustain the environment needed for survival.

Waste: The United States generates enough garbage daily to fill 63,000 garbage trucks, which, lined up, would stretch 400 miles from Los Angeles to San Francisco. The building industry accounts for 20 percent of this waste stream.

Energy: The U.S. Department of Energy estimates that improvements in U.S. building energy efficiency using existing technology could save $20 billion. Forty percent of the world’s energy is used to construct and operate buildings.

The numbers are grim, but designers and suppliers have real options for countering these trends. We can employ what I like to call the Seven Principles of Sustainable Design:

Use Low-Impact Materials: Select non-toxic, sustainably produced or recycled materials that require little energy to process.

Promote Energy efficiency: Use less energy to manufacture more efficient products.

Select for Quality and Durability: Use durable, longer-lasting and better-functioning products to minimize replacement frequency.

Design for Reuse and Recycling: Design products, processes and systems for performance in a commercial “afterlife.”

Employ Bio-Mimicry: Use scientific data to redesign industrial systems along biological lines, enabling the constant reuse of materials in continuous closed cycles.

Substitute for High-Use Service: Shift modes of consumption from single ownership to public/shared ownership (e.g., private automobile to car-sharing service). Promote minimal resource use per unit of consumption.

Choose Renewable Sources: Use materials extracted from nearby (local or bioregional), sustainably managed renewable sources that can be composted (or fed to livestock) when usefulness has been exhausted.

Responding to a Changing Society

Beyond the challenges we face in conserving scarce resources, a few key megatrends underscore the importance of sustainable coatings.

Population Growth: World population doubled from 2.5 billion in 1950 to 5 billion in 1990; it is projected to reach 9.8 billion in 2050. The population is also shifting from rural areas to major metropolitan areas, with people migrating for better employment, commerce and quality of life. New construction will be required to support growth and urbanization. We’ll need to replace, upgrade, repurpose and conserve existing structures and infrastructures.

Climate Change: Once mislabeled “global warming,” the significant, lasting change from relatively mild, predictable weather patterns to more unpredictable patterns increasingly will affect industrialized farming and dense urban populations. We’ll see more pressure to produce materials, products and assemblies that can withstand extreme variances in weather. Basic code-compliant solutions that are “good enough” today will no longer be acceptable.

We’re now designing disaster-mitigation plans and hardening essential facilities and infrastructure, as new codes require mitigation of rising water levels and storms we once saw every 100 years.

We can expect to see carbon dioxide emissions regulated, promoting net-zero buildings whose every feature is designed to reduce energy use and associated carbon emissions.

Greater emphasis will be placed on energy efficiency and energy recovery, as well as water-resource management and conservation.

Information Explosion: Information is growing exponentially, and a corollary increase in access to this information through the Internet means that people are more informed than ever about optimum human health and the risks associated with exposure to chemicals. We pore over studies seeking to define the “tipping point” for toxemia in terms of parts per billion of key compounds. We worry about information that links exposure to changes of DNA affecting future generations.

These health concerns are driving changes that have tremendous implications for building materials.

* New Regulations: States increasingly introduce regulations designed to control exposure and assure public health. The International Green Construction Code is now used for baseline sustainability in regular building codes.
* VOC Limits: VOCs are regulated on the West Coast via the South Coast Air Quality Management District, and on the East Coast via the Ozone Transport Commission. Recent changes in California have lowered VOC limits to a maximum of 50 grams per liter in coatings.
* New Organizations: The Living Building Challenge introduced a chemical “Red List” banning hazardous chemicals from use on projects.
* More Transparency: As a result of requirements in LEED v4 for product transparency, manufacturers of products used on LEED projects must detail the chemical content of the products in HPDs (health product declarations) and EPDs (environmental protection declarations).
* New Social Contract: Major petroleum chemical companies are forced to address the population’s desire to shift from oil and coal to natural gas and to renewable energy and biomass materials.

Technology Explosion: The last 20 years of mergers and acquisitions led to large chemical plants manufacturing single resins. The future lies in small batch processing of custom chemicals and new processing technologies. These include nano-technology, micron-level changes to alter product performance; phase-changing materials, capable of storing and releasing large amounts of energy; and regenerative chemicals that respond to environmental changes.

What do these megatrends mean for the chemical industry? They portend a shift in processes, standardization and approach.

Closed-Loop Processes: Manufacturing closed loops are economically advantageous, reduce/ eliminate waste, reuse byproducts and allow reformulation into new products without downcycling. Shaw Carpet is one success story, creating nylon 6 fibers that can be recycled 100 percent into new carpet. Shaw’s activity resulted in record profits, as producing carpet with nylon 6 requires no new petrochemicals.

Tightening of Standards/LCA: Life-cycle costing is the true measure of value instead of traditional first-cost thinking. That is important where better products require less maintenance. Tightening standards will help designers maintain quality through specifications.

In fact, as coatings technologies advance, our reliance on standards increases. Standards organizations whose certifications for sustainable offerings fail to keep up with national programs, or whose certifications don’t perform as intended, will be bypassed. For coatings specific standard groups to survive, they must align with national standards and address high performance and durability.

Stricter Guidelines: In the healthcare and science laboratory industries, stricter guidelines will be required to combat hospital-acquired infections and address the harsh chemicals/disinfectants necessary to stem infections.

Alchemizing Toxic Chemicals: The storage of large quantities of toxic chemicals at various waste sites necessitates that we incorporate toxic chemicals in ways that alchemize them, creating non-toxic, stable, safe products that can be reused and recycled, without toxicity. For example, LEED supports the use of fly ash, the byproduct of coal manufacturing, as cement replacement in concrete production. This activity will decrease chemical reservoirs of fly ash so that they no longer pose a health hazard.

Creating Coatings for the Future

Coatings technology has evolved as manufacturers respond to market needs and awareness.

Getting the Lead Out: For decades lead was added to paints and coatings to improve durability and color retention. Research into the hazards of lead paint, and lead dust, made the industry move from lead to safer alkyd formulations. Recent awareness of the high VOC content has led manufacturers to replace alkyds with lower-VOC acrylic latex systems.

Where initial productions met market skepticism regarding performance and durability, formulation improvements now offer paint coatings with low VOCs and better performance, durability, color retention and color-hiding capability than older technologies.

Improving Corrosion-Resistant Coatings: Corrosion-resistant coatings for architecturally exposed structural steel have been three-coat systems consisting of organic or inorganic zincrich primers, epoxy intermediate and aliphatic polyurethane topcoats as the most durable high-performance coatings. Advances in the last 20 years have led to two-coat polysiloxane coating systems that, for mild to moderate atmospheric exposure, provide excellent corrosion resistance along with color and gloss retention said to surpass that of polyurethanes.

Improving Coatings to Protect Aluminum: Coatings to protect aluminum required chromate pretreatment for surface preparation and bonding of PVDF resin coatings. Awareness of the toxicity of hexavalent chromate prewashes led to development of coatings that do not need chromate prewashes but offer the same service life and durability.

In addition, the EPA introduced a significant new use rule (SNUR) last September to limit/eliminate perfluorinated compounds (PFCs) in PVDF coatings in response to overwhelming evidence that these chemicals are persistent bioaccumulative toxicants. PFCs were used as surfactants to improve the bond between coatings and metals. Producers of PVDF coatings altered the chemistry to remove perfluorooctanoic acids. Combining PVDF coatings with acrylics, coatings companies created low-VOC, water-based PVDF coatings with the same performance as the solvent-based PVDFs and that can be applied in the field, making initial application and long-term maintenance easy. Other chemical companies altered the chemistry of PVDF coatings further, developing powder coatings that can be applied in the field or in the shop with the same performance as 20-year warrantable fluid-applied systems.

Advancing Anti-Microbials: In high-performance interior coatings for laboratories and hospital facilities, epoxy paints and coatings recently have been replaced with two-component waterborne polyurethane systems based on advancements in polyurea technology. These systems provide high-durability coatings that can contain anti-microbial additives. They have great color retention and durability, while reducing dry time in shop preparations.

Controlling Moisture in Buildings: Rain screen design and energy regulations led to improvement in the building energy envelope through creation of air barrier systems. Controlling the movement of moisture through the building envelope increases the durability and life of the thermal envelope. Fluid-applied air barriers face new challenges as IBC 2012 adopts NFPA 285, mandating assembly fire testing of the exterior envelope. Companies must alter formulations to respond to new requirements for fire test performance.

Overcoming Issues of Fire-Resistant Chemicals: In the last year, fire-resistant coatings came under attack due to studies linking halogenated products to human health issues. Early formulations migrated, leaching chemicals in the environment. Independent research studies showed fire-retardant coatings to be carcinogenic and endocrine disruptors.

Recent formulations provide more durability and intimate bond chemicals in chemical composition of insulation products to prevent leaching and the related hazards. However, the public damage sustained as a result of published reports has led makers of children’s clothing, bedding and toys to remove fire-resistant chemicals from their products. Some design professionals are pushing for building code legislation to remove the requirement that building insulation be fire resistant.

So what developments can we expect in architectural coatings technology?

Phase-Changing Technology: Phase-changing materials will become more mainstream to address changes in environmental conditions. Coatings will change chemistry in response to environmental changes. These coatings will improve the energy efficiency of the building envelope, while minimizing unwanted effects.

Cool Roofing: Some debate surrounds cool roof technology. Reflective roof coatings help reduce energy demand during cooling cycles by reflecting heat from solar radiation. LEED points are available for use of cool roof coatings that help limit the heat island effect in urban environments. Recent research indicates that cool roofs are most effective in reducing the building energy use where the number of cooling days exceeds the number of heating days. Darker roofs may provide better energy performance in colder regions. However, water runoff from black roofs increases the temperature of water in rain and water runoff, and may be harmful to downstream biomes. Cool roof systems not only reduce the heat island effect locally, but also minimize this damage to ecosystems miles away.

Titanium Dioxide Coatings: TiO2 coatings clean surfaces through photocatalytic action, using UV light to activate coatings to bond with carbon dioxide. They produce hydrocarbon runoff and oxygen and clean the environment. Here are some examples.

* As a concrete additive, titanium dioxide maintains white concrete surfaces, minimizing maintenance, by de-bonding with carbon and dirt. It cleans the air by cycling and capturing carbon particles and VOCs.
* In healthcare environments, TiO2 coatings may stimulate antimicrobial action. Operating rooms can “self-clean,” eliminating bacteria between operations by activating enzymatic action through exposure to UV, infrared or other spectral light.
* In another application of TiO2 used as a photocatalyst, the technology is used to coat “self-cleaning glass.” While relatively expensive, such technologies may be valuable in polluted areas like China, where rains pose a durability threat to building materials that self-cleaning chemicals can mitigate.
* Data reveal that indoor air quality is 10 times more toxic than exterior air. Tightening the building envelope has exacerbated this issue. Manufacturers have produced TiO2- based surface treatments that are activated by UV light to actively purify air when applied to interior and exterior surfaces.

Reenvisioning What’s Possible

These are exciting developments, but they’re just the beginning. Here are but a few innovations to look for in coming years.

Regenerative Coatings. Such coatings alter their chemistry to respond to the environment in ways that are regenerative. For example, cool roof coatings will be developed to respond to hot days by reflecting light and to cold days by absorbing heat. This technology will be available through integration of phase-changing materials and nanotechnology.

Better Insulation. Thinner, lighter, more efficient insulations will come down in price, become more mainstream and be adopted by building codes to increase thermal control of the interior environment.

Inherently Fire-Resistant Coatings. These coatings will use nano-technology to produce products that are inherently flame and fire resistant, so fire retardants are not necessary.

More broadly, we can expect the product transparency requirements in LEED v4 and chemical bans from groups like the Living Building Challenge to fundamentally change our approach to materials development and selection. The industry will produce healthier, environmentally sustainable chemicals and products that mitigate problems while maintaining performance and long life.

This effort will depend on greater cooperation among design professionals, chemical companies, manufacturers and fabricators. It will take communication on individual projects, as well as collaboration through cross-industry channels. The challenges and stakes we face are unprecedented in human history, but so are the opportunities.

## 1AC---Private Antitrust

Advantage Two is Private Antitrust.

#### Private antitrust litigation is inevitable.

Foix et al. ’20 [Danyll, Ann O’Brien, Carl Hittinger, and Jeanne-Michele Mariani; October 19; Partner at Baker Hostetler’s Washington D.C. Office; former member of the US Department of Justice in the Antitrust Division; Adjunct Professor teaching antitrust law at the Drexel University Thomas R. Kline School of Law; law clerk; Global Competition Review, “United States: Private Antitrust Litigation,” <https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation>]

Private price-gouging litigation

A recent flurry of price-gouging lawsuits in the United States suggests that this could be only the beginning of a swell of private litigants bringing claims of alleged price-gouging due to covid-19. These lawsuits may foreshadow the nature of future private price-gouging litigation and the parties that might find themselves entangled in such litigation.

The recent lawsuits suggest that price-gouging cases will be premised on various forms of price-related allegations. Some of these cases allege conventional price-gouging, such as a recently filed case by egg purchasers alleging that egg prices nearly tripled between the onset of the pandemic and the end of March and have remained at price levels deemed unlawful (at least 10 per cent higher than usual) under the California Unfair Competition Law (UCL). [26](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-010) Other recent cases suggest claims may be based on market or consumer preference changes due to the pandemic, such as in a case alleging food delivery companies, which have increased in use during the pandemic, wielded their power to force restaurants to charge uniform prices for menu items in violation of federal antitrust laws. [27](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-009) This case also suggests that price-gouging may become a commonplace, additional cause of action in antitrust complaints challenging conduct during the pandemic.

Along with price-gouging and antitrust claims, recent cases show additional causes of action may be expected. Across the country, numerous suits alleging breach of contract and consumer protection violations also have been filed in the wake of covid-19, particularly against gyms, sports and other membership clubs. [28](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-008) In these cases, the plaintiffs usually allege deceptive trade practices when the clubs continued to charge membership or usage fees despite their facilities being closed due to the pandemic. Although gyms and sports clubs have been the primary targets so far, the case theory advanced by the plaintiffs’ lawyers could apply to any business and has already been advanced against resorts and popular theme parks where customers retain yearly passes. [29](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-007)

A broad swath of firms may be accused of price-gouging, as demonstrated in recent cases. As an example, in a case filed in federal court in California, the plaintiffs allege that more than 20 of the largest egg producers, distributors and retailers in the United States [30](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-006) price-gouged in violation of the California UCL. [31](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-005) Relying on the California Attorney General’s statement that its price-gouging law ‘applies to transactions between manufacturers, wholesalers, distributors, and retailers as it does between retailers and consumers’, [32](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-004) the plaintiffs sought damages from any and all of the producers and retailers that increased their transaction prices by 10 per cent or more during the pandemic emergency. This case illustrates that more than customer-facing retailers may be sued and participants at every level in a distribution or supply chain may be subject to price-gouging litigation. [33](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-003)

Similarly, a wide variety of products may be subjected to price-gouging claims, as illustrated by a series of recent cases. A class action case against an online retailer shows that price-gouging claims can include items as diverse as black beans, hair remover, facial cleanser, pain relief tablets and laundry detergent. [34](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-002) Even a single item may lead to litigation. In a recent case against a grocery store chain, the plaintiff focused on toilet paper in alleging that grocery stores sold essential items in excess of their pre-covid-19 prices and in violation of the California UCL. [35](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-001) While price-gouging laws focus on ‘essential’ items, the identification of these items can depend on the nature of a given emergency, and the ongoing covid-19 litigation has defined little limitation for such items during the pandemic.

The private cases filed during covid-19 indicate that class actions may be the popular vehicle for bringing cases with price-gouging and antitrust claims. Nearly all the recent cases brought by consumers have been class actions. Because price-based, consumer allegations can present common issues of fact or law, these allegations may be amenable to class-wide litigation. Also, the plaintiffs bringing price-gouging cases can be incentivised to pursue class claims when they file in jurisdictions that permit the recovery of restitution on behalf of the citizens of the state. [36](https://globalcompetitionreview.com/review/the-antitrust-review-of-the-americas/2021/article/united-states-private-antitrust-litigation#footnote-000) For these reasons, price-gouging cases, especially those involving consumer purchases, will likely continue to be litigated as class actions.

Conclusion

United States price-gouging laws are intended to protect victims of life-altering disasters from having to pay exorbitant prices for essential products. The complex web of state and federal price-gouging laws can be hard to navigate as companies try to price scarce, difficult-to-source, essential products during the pandemic. There is very little legal precedent to look to for guidance, as very few price-gouging cases have been litigated to conclusion. While the covid-19 pandemic has led to the commencement of many federal, state and private cases focused on price-gouging, these cases may take years to play out. The cases recently filed, however, indicate that price-gouging and antitrust allegations may go hand-in-hand for many complaints, any number of products may be targeted for price-gouging, firms at all levels of distribution or supply chains may be vulnerable to claims, and class actions will continue to be the popular vehicle for litigating these cases.

#### Arbitration hamstrings enforcement and structurally incentivizes collusion.

Elhauge ’15 [Einer; 2015; Petrie Professor of Law at Harvard Law School, Founding Director of the Petrie-Flom Center for Health Law Policy, Biotechnology and Bioethics, former Chairman of the Antitrust Advisory Committee to the Obama Campaign; Fordham International Law Journal, “How Italian Colors Guts Private Antitrust Enforcement by Replacing it with Ineffective Forms of Arbitration,” vol. 38]

The same problem infects consents to arbitration clauses that waive the right to effective vindication of antitrust law. If buyers acted together, then they would only consent if those waivers made them better off. But acting individually, each buyer has incentives to consent in exchange for a trivial discount from the inflated market-wide prices that will result when all buyers consent to effectively immunizing antitrust violations against them. It takes only a trivial discount because each buyer knows that their individual decision whether to consent has little effect on whether the market-wide harm from immunizing antitrust violations occurs.

To put it another way, competitive markets are a public good, from which each buyer in a market benefits, whether or not that buyer contributes to the creation of that public good by rejecting conduct or agreements that keep that market competitive. Thus, buyers inevitably have incentives not to contribute; instead they will predictably consent to conduct and arbitration waivers that result in uncompetitive markets.

The future implications are alarming. Given the Italian Colors decision, it is hard to see why all businesses would not at least insert arbitration clauses into their contracts that preclude class arbitration. Given the limited nature of discovery in arbitration, that alone will bring US private enforcement largely into convergence with Europe, and perhaps will leave US private enforcement even less effective than the European Union in the future if the new EU directive leads to stronger national rules on discovery and class actions.

Businesses are likely to go even further given the Supreme Court’s logic that arbitration provisions are permissible whenever they eliminate only the right to prove a claim, rather than the right to pursue it. Under this logic, parties could adopt arbitration provisions eliminating the ability to introduce economic expert testimony altogether, even though that would effectively preclude not only class suits but also suits by corporate plaintiffs that might have large enough stakes to fund an expert. The Court offered two responses to this possibility. First, it said, “it is not a given that such a clause would constitute an impermissible waiver,”12 which alarmingly suggests this possibility might well be in our future. Second, the Court said that this possibility would be different because “such a clause, assuming it makes vindication of the claim impossible, makes it impossible not just as a class action but even as an individual claim.”13 But that rationale conflicts with the Court’s logic that the difference is between being able to pursue a claim and prove it, and disturbingly suggests the Court is resting instead on a hostility to class actions over corporate suits.

Moreover, the Court’s logic would also seem to permit many other possible ways of gutting antitrust enforcement that the Court did not address. Parties could adopt provisions that preclude discovery even more than it is already limited in arbitration, say by barring any discovery into market definition, power, or anticompetitive effects. Indeed, the Court’s distinction between barring proof versus barring pursuit of a claim would even suggest that arbitration clauses could baldly prohibit offering any proof in arbitration on market definition, power, or anticompetitive effects, because that would go simply to the right to prove the claim. This would leave private enforcement by US buyers even less effective than in Europe.

This development would immunize businesses against US federal antitrust enforcement by anyone who contracts with them, which is almost any private party who can sue given that federal antitrust law largely limits antitrust enforcement to direct purchasers. The main exception would be antitrust suits by rivals excluded by exclusionary conduct, who may have no contract with the defendant and thus no arbitration provision. But that is hardly an adequate substitute because:

[A]ny rival claim will be limited to the competitive profits the rival could have earned on some share of the market in the but-for world. A monopolist will generally find it profitable to pay such low competitive profits on a smaller market share out of the monopoly profits it gains on its monopoly market share.14

Further, “it is too easy to cut side deals with rivals through settlements that may satisfy the financial interests of the rivals but fail to fix (or even worsen) the anticompetitive problem.”15 Indeed, the Italian Colors decision creates incentives for them to cut side deals that include arbitration provisions that bar effective antitrust enforcement between them. And given that the Italian Colors decision allows each business to use arbitration clauses that effectively immunize them against their buyers, businesses might not have much incentive to even try to exclude each other since it is more profitable to instead collude and jointly exploit their buyers.

#### Empowering private antitrust enforcement stimulates competition in the healthcare sector.

Helm ’17 [Anne; 2017; Chief of Staff to the Chancellor & Dean at the Hastings College of the Law at the University of California; Saint Louis University Journal of Health Law and Policy, “Optimizing Private Antitrust Enforcement in Health Care,” vol. 11]

I. Introduction

Americans are paying too much for health care services and insurance, in large part due to insufficient competition among providers and payors.1 Waves of consolidation in these markets have fortified providers and insurers with market power, resulting in higher prices and lower quality for consumers.2 As antidotes, health economists and other policy advocates have proposed various legislative, regulatory, and enforcement solutions.3 Yet private antitrust enforcement is rarely recommended to remedy health care market dysfunction. Whereas public antitrust enforcement is generally touted as indispensable,4 private antitrust enforcement is often disregarded as baseless, self-serving litigation that only strains judicial resources and may even raise costs.5 But the notion that private litigation is important should not be controversial.6 Private antitrust enforcement can restore competition, deter antitrust violations, and compensate victims in the markets for health care services and insurance, and, accordingly, the United States should be looking for ways to optimize it.

When passed, the antitrust statutes envisioned private cases as a fundamental part of an overall enforcement scheme.7 Indeed, the treble damages remedy was meant to spur private litigation.8 The Supreme Court has acknowledged as much: “By offering potential litigants the prospect of a recovery in three times the amount of their damages, Congress encouraged these persons to serve as ‘private attorneys general.’”9 The Court later elaborated, “The treble-damages provision wielded by the private litigant is a chief tool in the antitrust enforcement scheme, posing a crucial deterrent to potential violators.”10 Over the last century, private cases have greatly outnumbered public enforcement actions.11 Recently, however, “private actions have caught up in the well-orchestrated, ideologically driven ‘tort reform’ movement” and have been characterized as “legalized blackmail” as opposed to a vital component of our statutory antitrust scheme.12 Private antitrust enforcement does not deserve this characterization and indeed is a much needed means to address health care pricing.

Antitrust law is premised on the notion that competition leads to lower costs, higher-quality products and services, and encourages investment and innovation. In health care, as former Federal Trade Commission (FTC) Chair Edith Ramirez stated, “The success of health care reform in the United States depends on the proper functioning of our market-based health care system.”13 Although highly regulated and somewhat complicated by the buyer and seller relationships among patients, providers, and payors, health care in the United States is nonetheless market-based. As such, the sector depends on competition to drive prices down and quality up, even after the at-risk Patient Protection and Affordable Care Act.14 There is a real need for more antitrust enforcement in health care. As to hospital mergers, a named top public enforcement priority,15 the FTC has only challenged one percent of mergers over the past decade.16 And, even with both the FTC and the Department of Justice (DOJ) enforcing the federal antitrust laws, the lower-priority cases challenging anti-competitive conduct are even more scant, and criminal cases are rarer still.17 In this void, private antitrust enforcement is essential to address market power in health care, and one that assumes a role that public enforcement cannot—or does not—presently fill.18

The insufficiency of public enforcement to address antitrust concerns in health care will likely only be exacerbated by the new presidential administration, under which at least one commentator has noted that “it is fair to expect some tempering of the level of activity that characterized the Obama administration.”19 Generally, Republican administrations are less likely to intervene in transactions and challenge the conduct of businesses, and despite some campaign rhetoric to the contrary, President Trump’s appointments seem to indicate an approach more in line with the party than with a new populism.20 Of course, political influence is not limited to the federal realm; in states, the political priorities of elected attorneys general influence antitrust policy as well. Nevertheless, even the most aggressive public enforcement scheme would be incapable of addressing antitrust issues in health care without its private cousin.

What can private antitrust enforcement accomplish? Effective enforcement achieves deterrence, compensates victims,21 and maintains or restores competition in health care markets. Private enforcement allows health care entities to police their own markets and consumers to seek relief from anticompetitive acts. But it is often said that antitrust laws are meant to protect competition and consumers, not competitors.22 The concern is that entities, acting in their own self-interest, will use the antitrust laws to try to modify contracts, redress various business torts, stifle competition, and extort settlements from rivals.23 Despite criticisms that private suits are self-interested and therefore anti-competitive, a lawsuit can be both self-interested and pro-competitive.24 Indeed, the antitrust laws were written to take advantage of private plaintiffs’ incentives and information to bring suits that benefit both themselves and consumers.

Moreover, to limit the likelihood of abuse, courts have narrowed the per se doctrine, increased standing requirements,25 and augmented the pleading standards; all of which deter frivolous, self-serving suits. In any event, studies have shown that antitrust actions by competitors in more concentrated markets, like health care markets, are more likely to be pro-competitive than they would be in more dispersed industries.26 Some would argue that these measures even overdeter and overscreen.27

#### Antitrust expands coverage quality and reverses data blocking---a consilience of research proves.

Gaynor ’21 [Martin; May 19; E.J. Barone University Professor of Economics and Public Policy at Heinz College in Carnegie Mellon University; Statement before the Committee on the Judiciary Subcommittee on Competition Policy, Antitrust, and Consumer Rights U.S. Senate, “Antitrust Applied: Hospital Consolidation Concerns and Solutions,” <https://www.judiciary.senate.gov/imo/media/doc/Gaynor_Senate_Judiciary_Hospital_Consolidation_May_19_2021.pdf>]

4 Evidence on the Impacts of Consolidation

There is now a considerable body of scientific research evidence on the impacts of hospital consolidation (see Gaynor et al., 2015; Tsai and Jha, 2014; Gaynor and Town, 2012a,b; Dranove and Satterthwaite, 2000; Gaynor and Vogt, 2000; Vogt and Town, 2006, for reviews of the evidence).

4.1 Impacts on Prices

4.1.1 Hospital Mergers

There are many studies of hospital mergers. These studies look at many different mergers in different places in different time periods, and find substantial increases in price resulting from mergers in concentrated markets (e.g., Town and Vistnes, 2001; Krishnan, 2001; Vita and Sacher, 2001; Gaynor and Vogt, 2003; Capps et al., 2003; Capps and Dranove, 2004; Dafny, 2009; Haas-Wilson and Garmon, 2011; Tenn, 2011; Thompson, 2011; Gowrisankaran et al., 2015). Price increases on the order of 20 or 30 percent are common, with some increases as high as 65 percent.6

These results make sense. Hospitals’ negotiations with insurers determine prices and whether they are in an insurer’s provider network. Insurers want to build a provider network that employers (and consumers) will value. If two hospitals are viewed as good alternatives to each other by consumers (close substitutes), then the insurer can substitute one for the other with little loss to the value of their product, and therefore each hospital’s bargaining leverage is limited. If one hospital declines to join the network, customers will be “almost as happy” with access to the other. If the two hospitals merge, the insurer will now lose substantial value if they offer a network without the merged entity (if there are no other hospitals viewed as good alternatives by consumers). The merger therefore generates bargaining leverage and hospitals can negotiate a price increase.

Overall, these studies consistently show that when hospital consolidation is between close competitors it raises prices, and by substantial amounts. Consolidated hospitals that are able to charge higher prices due to reduced competition are able to do so on an ongoing basis, making this a permanent rather than a transitory problem. Moreover, there is no difference between not-for-profit and for-profit hospitals in the extent to which they raise prices due to increased market power.

There is also more recent evidence that mergers between hospitals that are not near to each other can lead to price increases. Quite a few hospital mergers are between hospitals that are not in the same area (see Figure 4). Many employers have locations with employees in a number of geographic areas. These employers will most likely prefer insurance plans with provider networks that cover their employees in all of these locations. An insurance plan thus has an incentive to have a provider network that covers the multiple locations of employers. It is therefore costly for that insurer to lose a hospital system that has hospitals in multiple locations – their network would become less attractive. This means that a merger between hospitals in these different locations can increase their bargaining power, and hence their prices.

There are two recent papers find evidence that such mergers lead to significant hospital price increases. Lewis and Pflum (2017) find that such mergers lead to price increases of 17 percent. Dafny et al. (2019) find that mergers between hospitals in different markets in the same state (but not in different states) lead to price increases of 10 percent.

Understanding the competitive effects of cross-market hospital mergers is an important area for further investigation, and determining appropriate policy responses (Brand and Rosenbaum, 2019).

4.1.2 Hospital Acquisitions of Physician Practices

Studies that examine the impacts of hospital acquisitions of physician practices find that such acquisitions result in significantly higher prices and more spending (Capps et al., 2018; Neprash et al., 2015; Baker et al., 2014; Robinson and Miller, 2014). For example, Capps et al. (2018) find that hospital acquisitions of physician practices led to prices increasing by an average of 14 percent and patient spending increasing by 4.9 percent.

4.2 Impacts on Quality

Just as important, if not more, than impacts on prices are impacts on the quality of care. The quality of health care can have profound impacts on patients’ lives, including their probability of survival.

4.2.1 Hospital Mergers

A number of studies have found that patient health outcomes are substantially worse at hospitals in more concentrated markets, where those hospitals face less potential competition.

Studies of markets with administered prices (e.g., Medicare) find that less competition leads to worse quality. One of the most striking results is from Kessler and McClellan (2000), who find that risk-adjusted one year mortality for Medicare heart attack (acute myocardial infarction, or AMI) patients is significantly higher in more concentrated markets.7 In particular, patients in the most concentrated markets had mortality probabilities 1.46 points higher than those in the least concentrated markets (this constitutes a 4.4% difference) as of 1991. This is an extremely large difference – it amounts to over 2,000 fewer (statistical) deaths in the least concentrated vs. most concentrated markets.

There are similar results from studies of the English National Health Service (NHS). The NHS adopted a set of reforms in 2006 that were intended to increase patient choice and hospital competition, and introduced administered prices for hospitals based on patient diagnoses (analogous to the Medicare Prospective Payment System). Two recent studies examine the impacts of this reform (Cooper et al., 2011; Gaynor et al., 2013) and find that, following the reform, risk-adjusted mortality from heart attacks fell more at hospitals in less concentrated markets than at hospitals in more concentrated markets. Gaynor et al. (2013) also look at mortality from all causes and find that patients fared worse at hospitals in more consolidated markets.

Studies of markets where prices are market determined (e.g., markets for those with private health insurance) find that consolidation can lead to lower quality, although some studies go the other way. In my opinion the strongest scientific studies find that quality is lower where there’s less competition. For example, Romano and Balan (2011) find that the merger of Evanston Northwestern and Highland Park hospitals had no effect on some quality indicators, while it harmed others. Capps (2005) finds that hospital mergers in New York state had no impacts on many quality indicators, but led to increases in mortality for patients suffering from heart attacks and from failure. Hayford (2012) finds that hospital mergers in California led to substantially increased mortality rates for patients with heart disease. Cutler et al. (2010) find that the removal of barriers to entry led to increased market shares for low mortality rate CABG surgeons in Pennsylvania. Haas et al. (2018) find that system expansions (such as those due to merger or acquisition) can pose significant patient safety risks. Short and Ho (2019) find that hospital market concentration is strongly negatively associated with multiple measures of patient satisfaction.

4.2.2 Hospital Acquisitions of Physician Practice

Research on the effects of hospital ownership of physician practices does not find evidence of improved quality. McWilliams et al. (2013) find that larger hospital owned physician practices have higher readmission rates and perform no better than smaller practices on process based measures of quality. (Scott et al., 2018) find no improvement in quality of care at hospitals that acquired physician practices compared to those that did not. Koch et al. (2020) do not find significant effects of hospital ownership of physician practices on Medicare patients’ health outcomes. Short and Ho (2019) also find a limited effect of hospital ownership of physician practices on Medicare quality measures, but find that increased market concentration is strongly associated with reduced quality. Further, the testimony of Dr. Kenneth Kizer in a recent physician practice merger case (Federal Trade Commission and State of Idaho v. St. Luke’s Health System, Ltd, and Saltzer Medical Group, P.A.) documents that clinical integration is achieved with many different forms of organization, i.e., that consolidation isn’t necessary to achieve the benefits of clinical integration.8

4.2.3 Patient Referrals

There has been concern about the possible impact of hospital ownership of physician practices on where those physicians refer their patients, and whether that is in the patients’ best interests (Mathews and Evans, 2018). A number of studies have found that patient referrals are substantially altered by hospital acquisition of a physician practice. (Brot-Goldberg and de Vaan, 2018) find that if primary care physicians in Massachusetts are in a practice owned by a health system they are substantially more likely to refer to an orthopedist within the health system that owns the practice. They also estimate that this is largely due to anti-competitive steering. (Venkatesh, 2019) examines Medicare data and finds a 9-fold increase in the probability that a physician refers to a hospital once their practice is acquired by the hospital. Hospital divestiture of a practice has the opposite effect (Figure 6). A study by Walden (2017) also employs Medicare data and finds that hospital acquisitions of physician practices “increases referrals to specialists employed by the acquirer by 52 percent after acquisition”, and reduces referrals to specialists employed by competitors by 7 percent. Whaley et al. (2021) find evidence of a substantial shift of referrals to hospitals as a result of hospital ownership of physician practices, and Young et al. (2021) find that hospital acquisitions of physician practices led to increases in inappropriate referrals for diagnostic imaging.

4.2.4 Labor Market Impacts, Monopsony Power

It is also possible that health care consolidation can have impacts on labor markets. Consolidation that causes competitive harm in the output market does not necessarily cause harm to competition in the input market (monopsony power is the term for market power in buying inputs). For example, two local grocery stores may merge to monopoly in an area, but they purchase frozen food items on a national market with lots of competition. Conversely, it is possible that a merger may have no harm to competition in the output market, but cause competitive harm in an input market. For example, consider two coal mines located in the same area that merge. Coal is sold on a national market, so the merger will not cause competitive harm. However, if the coal mines are the largest (or only) employers in the area, then the merger will cause harm to competition in the labor market.

In the case of health care, however, both the output market for health care services and the input market for labor are local. As a consequence, a merger that causes harm to competition in the market for health care services has nontrivial potential to harm competition in the labor market. The extent to which such a merger will cause labor market harms depends on the alternatives that workers have in terms of the types of other jobs available and where they are located. Nonspecialized workers, such as custodians, food service workers, and security guards are less likely to be affected by a merger, since their skills are readily transferable to other employers in other sectors.9 Workers who have specialized skills that are not readily transferable to other employers in other sectors are more likely to be harmed. For example, consider a town with two hospitals, a large automobile assembly plant, and multiple retail and service establishments. If the two hospitals merge to monopoly, hospital custodians and security guards will have alternatives at the assembly plant or at the retail or service establishments. As a consequence, competition for these workers may be little affected by the merger. Nurses and medical technicians, however, have nowhere else to turn in the local market, so there will be substantial harm to competition for health care workers.

There are a number of papers that have demonstrated the presence of monopsony power in the market for nurses (see e.g., Sullivan, 1989; Currie et al., 2005; Staiger et al., 2010). These papers demonstrate that hospitals possess and exercise monopsony power in the market for nurses. They do not, however, provide direct evidence on the impacts of consolidation. A recent paper, however, looks directly at the impacts of hospital mergers on workers’ wages. Prager and Schmitt (2021) look at the impacts of 84 hospital mergers nationally between 2000 and 2010. They find that hospital mergers that resulted in large increases in concentration substantially reduced wage growth for workers with industry specific skills, but not for unskilled workers. They find that “Following such mergers, annual wage growth is 1.1 percentage points slower for skilled non-health professionals and 1.7 percentage points slower for nursing and pharmacy workers than in markets without mergers.” This suggests that hospital mergers can harm competition in the labor market for workers with skills specific to the hospital industry.

The impacts of consolidation on labor markets (and input markets generally) is an area where study is needed to understand the nature of the impacts of consolidation and evidence of those effects. Moreover, antitrust authorities need to know to what extent merger enforcement focused on output markets addresses potential input market competitive harms, and to what extent input markets require a separate focus. Further, if the agencies are to pursue enforcement in this area they need to develop economic and legal approaches to this issue.

4.3 Impacts on Costs, Coordination, Quality

It is plausible that consolidation between hospital, physician practices or insurers, in a number of combinations, could reduce costs, increase care coordination, or enhance efficiency. There may be gains from operating at a larger scale, eliminating wasteful duplication, improved communications, enhanced incentives for mutually beneficial investments, etc. However, it is important to realize that consolidation is not integration. Acquiring another firm changes ownership, but in and of itself does nothing to achieve integration. Integration, if it happens, is a long process that occurs after acquisition.

While the intuition, and the rhetoric, surrounding consolidation, has been positive, the reality is less encouraging. The evidence on the effects of consolidation is mixed, but it’s safe to say that it does not show overall gains from consolidation (Neprash and McWilliams, 2019). Merged hospitals, insurers, physician practices, or integrated systems are not systematically less costly, higher quality, or more effective than independent firms (see Burns and Muller, 2008; Burns et al., 2015; Goldsmith et al., 2015; Burns et al., 2013; McWilliams et al., 2013; Tsai and Jha, 2014).

For example, Burns et al. (2015) find no evidence that hospital systems are lower cost, Goldsmith et al. (2015) find no evidence that integrated delivery systems perform better than independents, Koch et al. (2018) find higher Medicare expenditures for cardiology practices in consolidated markets, and McWilliams et al. (2013) find higher Medicare expenditures for large hospital-based practices. In contrast, Schmitt (2017) finds evidence of significant cost savings (4-7 percent) due to hospital mergers, with the exception of mergers of hospital in the same market (and thereby likely competitors). Gaynor et al. (2021) examine the merger of two large hospital chains. They find that the acquisition led to adoption of a new electronic medical record system, and similarity of management practices, but neither the profitability of the acquired hospitals or the acquiring hospitals increased, nor did patient outcomes improve. Beaulieu et al. (2020) report that “Hospital acquisition by another hospital or hospital system was associated with modestly worse patient experiences and no significant changes in readmission or mortality rates. Effects on process measures of quality were inconclusive.”

After more than 3 decades of extensive consolidation in health care, it seems likely that the promised gains from consolidation would have materialized by now if they were truly there.

5 Anticompetitive Conduct

Firms that acquire a dominant market position usually wish to keep it. The incentive to maintain or enhance a dominant position can be beneficial when it leads the firm to deliver value to consumers in order to keep or gain their business. This can result in lower prices, higher quality, better service, or enhanced innovation. There may also be strong incentives for such firms to engage in anticompetitive practices in order to disadvantage competitors or make it difficult for new products or firms to enter the market and compete.

There are prominent instances of firms in the health care industry engaging in what appear to be anticompetitive tactics. Cooper et al. (2019) find that hospitals with fewer potential competitors are more likely to negotiate contracts with insurers that have payment forms that are more favorable to them (e.g., fee for service) and reject payment forms they dislike (e.g., DRG based payment). While this is not an anticompetitive practice, it suggests that hospitals with market power are able to negotiate contracts with insurers that contain anticompetitive elements. This indeed is the issue in some recent antitrust cases. These cases revolve around the use of restrictive clauses in hospital contracts with insurers.10

These clauses prevent insurers from using methods to direct their enrollees to less costly or better hospitals. One of these methods is called tiering - a practice where enrollees pay less out of their own pockets for care received from providers in a more favorable group (“tier”), and pay more if they see a provider in a less favorable tier. Insurers use tiering to give enrollees incentives to obtain care at less costly or higher quality providers. This system thus gives providers an incentive to do the things it takes to be in the more favorable tier, and is a way to promote competition. Another method is steering - enrollees are directed to providers who are preferred, due to lower costs or higher quality. Steering also promotes competition - providers have incentives to agree to lower prices or provide better quality or service in order to be in the preferred group. A third method employed by insurers is transparency – providing enrollees with information about the costs or quality of care at different providers. The intent is to provide enrollees with the information they need to choose the right provider, and by doing so to give providers incentives to compete on those factors.

In both of the antitrust suits mentioned above, the health systems had negotiated clauses in their contracts with insurers which prohibited the insurers from using any of these methods to try to direct patients to lower cost or better providers. The clauses prohibiting the use of these methods are called “anti-tiering,” “anti-steering,” and “gag” clauses. The concern with the use of these restrictive clauses is that they harm competition by preventing insurers by using methods that provide incentives to providers to compete to attract patients. The lawsuit by the DOJ against Carolinas Health System was settled, with the health system agreeing not to use these restrictive clauses.11 The California Attorney General’s lawsuit against Sutter Health System was also settled, with a similar outcome. 12

At present there is no systematic evidence on the extent to which anti-tiering, anti-steering, and gag clauses are being employed by health systems in their contracts with insurers, nor analysis of their impacts. This is an area which needs investigation to document the extent of the practice and its impacts.

Another practice that raises concerns is “data blocking” (Savage et al., 2019). Data blocking is a practice in which health systems impede or prevent the flow of patients’ clinical data to providers outside their system. It is also refers to a practice by electronic medical record (EMR) providers to impede the flow of data to rival EMR systems via lack of compatibility. Data blocking by providers makes it more difficult for patients to go to rival providers, locking them in, since their medical information doesn’t go with them. Reducing patient mobility across providers harms competition and benefits incumbents. While there are extensive reports of data blocking, there isn’t systematic evidence on the extent of the practice, or on its impacts. Study is needed to understand the nature of data blocking, and the extent to which it leads to harm to competition or to efficiencies.

#### Effective healthcare solves existential risks from pandemics and bioterror.

Millett ’17 [Piers and Andrew Snyder-Beattie; August 1; Ph.D. and Senior Research Fellow in the Future of Humanity Institute at Oxford University; M.S. and Director of Research in the Future of Humanity Institute at Oxford University; Health Security, “Existential Risk and Cost-Effective Biosecurity,” vol. 15]

Cost-Effective Biosecurity

How should we balance speculative risks of human extinction in a biosecurity portfolio? Here we turn to cost-effectiveness analysis, which is one method of prioritizing public projects.[29](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/#B29) Cost-effectiveness analysis is helpful if our goal is to maximize the effect of our resources to achieve a measurable aim (such as life-years saved or cases of disease averted). Here we compare the cost-effectiveness of reducing risks in the categories of incidents, events, disasters, and existential risks.

Calculating Costs

The US federal government was projected to spend almost $13 billion on health security–related programs in 2017.[59](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/#B59) To our knowledge, there has not been a quantitative assessment of how this spending has reduced the chances of bioterrorism, biowarfare, or even naturally occurring pandemics. However, the World Bank estimates that it would cost $1.9 billion to $3.4 billion per year over 5 years to bring all human and animal health systems up to minimal international standards, and it suggests that these measures would prevent at least 20% of pandemics. [60](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/#B60) Many countries do not currently have healthcare systems that meet international standards—for example, in 2014 only 33% of countries reported their national arrangements met those required under the International Health Regulations.[61](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/#B61)

These mitigation measures would be adopted to be effective regardless of whether a disease outbreak originates naturally, accidentally, or deliberately. The ability to rapidly detect and characterize the agent involved helps fast-track public health and R&D responses. Acting promptly enables basic public health measures that might decrease the likelihood of spread (such as social distancing) and track its emerging epidemiology (providing critical input for tailoring the responses). Even if we lack existing or candidate vaccines or therapeutics, having the capacity to treat symptoms can have a dramatic impact on case fatality rates.

We therefore assume that strengthening healthcare systems to meet international standards would have an impact on mitigating all types of disease risk, ranging from incidents and events to existential risks. We extend the World Bank's assumptions to include bioterrorism and biowarfare—that is, we assume that the healthcare infrastructure would reduce bioterrorism and biowarfare fatalities by 20%. We conservatively assume that existential risks will be reduced by only 1%, since any potential existential risk would likely be deliberately designed to overcome medical countermeasures.

We calculate that purchasing 1 century's worth of global protection in this form would cost on the order of $250 billion, assuming that subsequent maintenance costs are lower but that the entire system needs intermittent upgrading.[†††††](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/#fn18) To calculate the cost per life-year saved, we use the equation C/(N × L × R), where C is the cost of reducing risk, N is the number of biothreats we expect to occur in 1 century, L is the number of life-years lost in such an event, and R is the reduction in risk achieved by spending a given amount (specified by C). For non-extinction risks, we increase L 50 times over to denote 50 life-years saved per life. The denominator N × L × R denotes the total number of life-years saved. In a subsequent model we also apply a discount rate to represent policymakers concerned only about lives in the short term.

Results

Including future generations into our cost-effectiveness calculations demonstrates that reducing existential risks, even if they are improbable, can be incredibly cost-effective in expectation ([Table 2](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5576214/table/T2/)). Depending on the model used, we estimate that we can purchase 1 quality adjusted life-year in expectation for 10s of dollars (with outliers suggested around 12 cents to $1,600). Even with the most conservative estimates of existential risk, reducing the risk of human extinction is at least 100 times more cost-effective than standard biosecurity interventions, and possibly up to 1 million times more cost-effective.

#### Government agencies will overlook AI monopolization, risking widespread misinformation and algorithmic distopia.

Chakravorti ’21 [Bhaskar; July 27; Dean of Global Business in the Fletcher School at Tufts University, Ph.D. in Economics from the University of Rochester; Wired, “Biden’s ‘Antitrust Revolution’ Overlooks AI—at Americans’ Peril,” <https://www.wired.com/story/opinion-bidens-antitrust-revolution-overlooks-ai-at-americans-peril/>]

Despite the executive orders and congressional hearings of the “Biden antitrust revolution,” the most profound anti-competitive shift is happening under policymakers’ noses: the cornering of artificial intelligence and automation by a handful of tech companies. This needs to change.

There is little doubt that the impact of AI will be widely felt. It is shaping product innovations, creating new research, discovery, and development pathways, and reinventing business models. AI is making inroads in the development of [autonomous vehicles](https://www.ft.com/content/46ff4fe4-0ae6-4f68-902c-3fd14d294d72), which may eventually improve road safety, reduce urban congestion, and help drivers make better use of their time. AI recently predicted the molecular structure of almost every protein in the human body, and it helped develop and roll out a Covid [vaccine in record time](https://www.wsj.com/articles/how-ai-played-a-role-in-pfizers-covid-19-vaccine-rollout-11617313126). The pandemic itself may have accelerated AI’s incursion—in emergency rooms for [triage](https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/); in airports, where [robots spray](https://www.latimes.com/politics/story/2021-05-04/covid-automation-robots-trends-effects-on-workers) disinfecting chemicals; in increasingly automated [warehouses](https://www.uschamber.com/co/good-company/launch-pad/warehouse-robotic-automation-coronavirus-pandemic) and [meatpacking plants](https://www.wsj.com/articles/meatpackers-covid-safety-automation-robots-coronavirus-11594303535); and in our remote workdays, with the growing presence of chatbots, speech recognition, and email systems that get better at completing our sentences.

Exactly how AI will affect the future of human work, wages, or productivity overall remains unclear. Though service and blue-collar wages have lately [been on the rise](https://www.cnbc.com/2021/07/17/why-the-biggest-job-wage-boom-is-blue-collar-.html), they’ve stagnated for three decades. According to MIT’s Daron Acemoglu and Boston University’s Pascual Restrepo, [50 to 70 percent](https://www.nber.org/papers/w28920) of this languishing can be attributed to the loss of mostly routine jobs to automation. [White-collar occupations](https://www.brookings.edu/wp-content/uploads/2019/11/2019.11.20_BrookingsMetro_What-jobs-are-affected-by-AI_Report_Muro-Whiton-Maxim.pdf) are also at risk as machine learning and smart technologies take on complex functions. According to [McKinsey](https://hai.stanford.edu/ai-future-work-conference), while only about 10 percent of these jobs could disappear altogether, 60 percent of them may see at least a third of their tasks subsumed by machines and algorithms. [Some](https://siepr.stanford.edu/system/files/Artificial%20Intelligence%20and%20the%20Modern%20Productivity%20Paradox-%20A%20Clash%20of%20Expectations%20and%20Statistics.pdf) researchers argue that while AI’s overall productivity impact has been so far disappointing, it will improve; [others](http://bostonreview.net/forum/science-nature/daron-acemoglu-redesigning-ai) are less sanguine. Despite these uncertainties, most [expert](https://www.gsb.stanford.edu/insights/misplaced-fear-job-stealing-robots)s agree that on net, AI will “become more of a challenge to the workforce,” and we should anticipate a [flat to slightly negative impact](https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy) on jobs by 2030.

Without intervention, AI could also help undermine democracy–through amplifying misinformation or enabling mass surveillance. The past year and a half has also underscored the impact of algorithmically powered social media, not just on the health of democracy, but on health care itself.

The overall direction and net impact of AI sits on a knife's edge, unless AI R&D and applications are appropriately channeled with wider societal and economic benefits in mind. How can we ensure that?

A handful of US tech companies, including Amazon, Alibaba, Alphabet, Facebook, and Netflix, along with Chinese mega-players such as Baidu, are responsible for $2 of every $3 spent globally on AI. They’re also among the top AI patent holders. Not only do their outsize budgets for AI dwarf others’, including the [federal government](https://www.nationaldefensemagazine.org/articles/2021/2/10/federal-ai-spending-to-top-$6-billion)’s, they also emphasize building internally rather than buying AI. Even though they buy comparatively little, they’ve still cornered the AI [startup](https://www.cbinsights.com/research/report/ai-in-numbers-q1-2020/) acquisition market. Many of these are [early-stage](https://www.bloomberg.com/news/articles/2020-03-16/big-tech-swallows-most-of-the-hot-ai-startups) acquisitions, meaning the tech giants integrate the products from these companies into their own portfolios or take IP off the market if it doesn’t suit their strategic purposes and redeploy the talent. According to research from my [Digital Planet](https://sites.tufts.edu/digitalplanet/the-shifting-geography-of-talent/) team, US AI talent is intensely concentrated. The median number of AI employees in the field’s top five employers—Amazon, Google, Microsoft, Facebook, and Apple—is some 18,000, while the median for companies six to 24 is about 2,500—and it drops significantly from there. Moreover, these companies have near-monopolies of data on key behavioral areas. And they are setting the stage to become the primary suppliers of AI-based products and services to the rest of the world.

Each key player has areas of [focus](https://www.manceps.com/ai-examples) consistent with its business interests: Google/Alphabet spends disproportionately on natural language and image processing and on optical character, speech, and facial recognition. Amazon does the same on supply chain management and logistics, robotics, and speech recognition. Many of these investments will yield socially beneficial applications, while others, such as [IBM’s Watson](https://www.nytimes.com/2021/07/16/technology/what-ever-happened-to-ibms-watson.html?action=click&module=In%20Other%20News&pgtype=Homepage)—which aspired to become the go-to digital decision tool in fields as diverse as health care, law, and climate action—may not deliver on initial promises, or may fail altogether. Moonshot projects, such as level 4 driverless cars, may have an excessive amount of investment put against them simply because the Big Tech players choose to champion them. Failures, disappointments, and pivots are natural to developing any new technology. We should, however, worry about the concentration of investments in a technology so fundamental and ask how investments are being allocated overall. AI, arguably, could have more profound impact than social media, online retail, or app stores—the current targets of antitrust. Google CEO Sundar Pichai may have been a tad overdramatic when he declared that AI will have [more impact on humanity than fire](https://www.marketwatch.com/story/artificial-intelligence-is-more-profound-than-fire-electricity-or-the-internet-says-google-boss-11626202566), but that alone ought to light a fire under the policy establishment to pay closer attention.

Biden's antitrust revolutionaries need a four-step plan to confront the AI revolution.

Antitrust authorities must first be forward-looking. They must recognize that the AI chess pieces being moved today will shape tomorrow’s endgame–particularly in a tech industry with high barriers to entry and early moves that are hard to reverse after scale. Tech antitrust action often occurs after it’s too late. Policymakers should also trace the outlines of multiple future AI scenarios, including a dystopian one. They must imagine, for example, a society that suffers from “algorithmic poverty,” in which users generate data as unpaid “labor,” which is used to train algorithms that in turn displace wage-producing labor.

#### Technological misinformation is an existential threat.

Lin, 19—senior research scholar for cyber policy and security at the Center for International Security and Cooperation and Hank J. Holland Fellow in Cyber Policy and Security at the Hoover Institution, both at Stanford University (Herbert, “The existential threat from cyber-enabled information warfare,” Bulletin of the Atomic Scientists, 75:4, 187-196, dml) [language modifications denoted by brackets]

Corruption of the information ecosystem has become an existential threat to civilization as we know it because prosperity and advancement depend on a secure information infrastructure and environment that provides human beings with contextualized, reliable, trustworthy information when and where it is needed. Information is as much a part of human ecology and the essence of being human as DNA (itself a form of information!) is a part of the evolutionary processes in biological systems.

Today, chaos reigns in much of the information ecosystem on which societies depend. In many forums for political and societal discourse, national leaders shout about fake news, by which they mean information they do not like. These same leaders lie shamelessly, calling their lies truth, or perhaps “truthful hyperbole.” Acting across national boundaries, these leaders and their surrogates exacerbate existing divisions, creating rage and diminishing confidence in elections and democratic institutions. Using unsupported anecdotes and sketchy rhetoric, denialists undermine well-established science about climate change and other urgent issues. Established institutions of the government, journalism, and education – institutions that have traditionally provided stability – are under attack precisely because they have provided stability.

The founding of the Bulletin predates by several decades the widespread availability of computers, the Internet, smart phones, search engines, and social media. Few could imagine in 1945 a technological environment that affords today’s high-speed and widespread connectivity, high degrees of anonymity, insensitivity to distance and national borders, easy and customized information searches, democratized access to publishing capabilities, inexpensive production and consumption of information content (including and increasingly importantly emotionally evocative video and audio content), disintermediation of established information sources, and ubiquitous, always-on, always-available access to information sources through mobile devices.

Such advances in information technology have heralded the arrival of the information age, a world in which taking near-immediate advantage of information opens up enormous opportunities in both the private and public sectors for improved delivery of existing products and services and, perhaps more important, the creation of entirely new products and services. Products and services can be customized to individual needs and preferences on a large scale and at more affordable costs. Transactional friction can be tremendously reduced. Through the Internet of Things, actuators and sensors can be connected to process control computers to optimize the behavior and function of physical systems. Everywhere that information can be used to create and improve new and existing functionality (that is, essentially everywhere), one can find or imagine new information technologies to do so.

At the same time, advances in information technology have a dark side. The same increases in the volume and velocity of information have created a louder and more chaotic information environment that stimulates fast, angry, reflexive, intuitive, and visceral thinking, reaction, and action in people and thus displaces more complex, reflective, and rational thought. In a chaotic environment of information overload, people are more likely to use mental shortcuts as a way to reduce the cognitive burden that such an environment places on their thinking.

In recent years, we have seen how the Internet, social media, and mobile devices (and other technologies) can be used by foreign adversaries to interfere in elections and to disrupt the democratic process. We have seen:

● Social media exploitation of cognitive biases to increase their impact and reach – short messages of 280 characters and emotionally evocative video/ audio clips are nearly ubiquitous and much more the norm than they ever were two decades ago.

● Disintermediation of established information sources that reduces the role and influence of those previously responsible for providing factual information and proliferates information sources. The US Supreme Court noted in Associated Press v. US (1945) that “the widest possible dissemination of information from diverse and antagonistic sources is essential to the welfare of the public, that a free press is a condition of a free society.” Today, modern information technology has enabled the creation of a larger number of information sources than the 1945 US Supreme Court could possibly have imagined.

● Search engines that return highly visible results for queries based in large part on the popularity of those results and the inferred desires of the user for specific information rather than their actual importance to those queries. Such functionality also makes it easier than ever for people to find information online “by doing their own research,” thus indulging in their confirmation biases by selectively finding and attending only to information that confirms one’s beliefs. Search engine optimization techniques enable gaming of search algorithms to promote the visibility of false, misleading, or worthless information. ● Many-to-many connectivity that enables the formation of echo chambers and media bubbles that reinforce pre-existing beliefs.

● Large-scale data mining that allows adversaries to sift huge amounts of personal data on individuals to identify and target those most susceptible to customized, inflammatory, false, malign, or misleading messages – and also to keep such messages away from public view.

● Near-immediate data transfer, which enables propaganda and other malign information to spread far and wide quickly, while efforts to correct false information are more expensive, often fall short, and frequently fail altogether.

● Inauthentic voices that are largely indistinguishable from authentic ones. Macedonian entrepreneurs discovered ways to monetize an affinity of Trump voters for fake news (Subramanian 2017). Paid human employees of the Internet Research Agency created and spread false information on behalf of the Russian government prior to the 2016 U.S. election (MacFarquhar 2018). And automated “bots”–accounts purportedly associated with human users but in fact managed entirely or mostly by machines – add further chaos to the information environment.

Is this state of information affairs really new? Haven’t adversaries of all stripes always employed propaganda and lies – otherwise known as information warfare (or at least a big part of it) – to advance their interests?

Yes. Information warfare indeed has a long pedigree that reaches into the past for at least the three millennia since the Trojan Horse enabled Greek warriors to breach the walls around the city of Troy. Much more recently, the rise of the Nazi regime in Germany relied on propaganda. As Hitler (1925, 155–56) wrote:

[I]ts purpose must be . . . to attract the attention of the masses and not by any means to dispense individual instructions to those who already have an educated opinion on things or who wish to form such an opinion on grounds of objective study – because that is not the purpose of propaganda, it must appeal to the feelings of the public rather than to their reasoning powers. . . . The art of propaganda consists precisely in being able to awaken the imagination of the public through an appeal to their feelings, in finding the appropriate psychological form that will arrest the attention and appeal to the hearts of the national masses. . . . The receptive powers of the masses are very restricted, and their understanding is feeble.

But more so today than at any earlier point in human history, human beings are vulnerable to information warfare. At the same time that new information technologies have led to an increase in the volume and velocity of information available on Earth by many orders of magnitude in the past few decades, the cognitive architecture of the human mind is more or less unchanged on the time scale of centuries or even millennia.

On human cognition

Research in the fields of cognitive and social psychology has formalized what Hitler knew intuitively. We now understand that human cognitive processing capability is not unlimited; humans have finite cognitive resources that can be “used up” under mentally stressful circumstances. Findings from the same cognitive psychology that has transformed neoclassical economics into behavioral economics (and resulted in three Nobel Prizes in economics) have made clear the “bounded rationality” of human thought and the simultaneous existence in every individual of the capability to engage in two types of cognitive processing.

Specifically, heuristic dual-system cognitive theory posits that human beings have two systems for cognitive processing – an intuitive, reflexive, and emotionally driven mode of thought (often designated as System 1) and a slower, more deliberate, analytical mode of thought (often designated as System 2). Kahneman (2011) provides a primer on System 1 and System 2 thinking. (See Petty and Cacioppo 1986; Chaiken 1987 for other variants of dual-system cognitive theory; see Kruglanski and Thompson 1999 for a contrary view on dual-system cognitive theory.)

System 1 is designed to operate rapidly, but it can do so because it does not take account of all available information and is thus more prone to error (also called bias). System 2 operates more slowly but is more likely to take into account the available information and is less prone to error. People engaging in System 1 information processing respond more emotionally and less rationally or critically than in System 2 processing.

Most important, System 1 thinking is the default mode of thought for human beings – it uses smaller amounts of cognitive resources, relies on simple gutbased judgments, and is used more often when humans are under stress. For most situations encountered in everyday life, System 1 thinking is adequate and produces mostly valid and useful outcomes, but it often fails when a situation requires complex inferences for understanding. For such situations, System 2 thinking, which is effortful and consumptive of cognitive resources, is more often appropriate – and when individuals fail to use System 2 when it is appropriate to do so, they are easily misled.

Most individuals are capable of both System 1 and System 2 thinking; thus, the important operative question is the circumstances under which they select one or the other type of thinking. Psychology has accumulated considerable evidence relevant to this question.

For example, Taber and Lodge (2006) show that an individual tends to be less critical of information that is favorable to his or her position than of information that is not favorable – that is, he or she is more likely to engage in System 1 thinking for favorable information. People have a confirmation bias in their information seeking and processing behavior – they preferentially seek out information that is consistent with their beliefs and they are highly critical of (or ignore) information that contradicts their beliefs. In a meta-analysis of 91 studies, Hart et al. (2009) considered two motivations for how an individual might select information to consume – the desire to gain an accurate understanding of reality and the desire to feel validated in his or her beliefs. These two motivations conflict when an accurate understanding of reality does not validate one’s beliefs, and such a situation motivates the question of which of these motivations is more powerful. Hart et al. concluded that both motivations drive human informationseeking behavior, thus moderating each other to a certain extent, but that on balance, humans do exhibit a tendency towards the validation of their beliefs. People are also subject to belief perseverance (a.k.a. a continuing influence effect) – a cognitive bias through which individuals do not revise beliefs based on erroneous information even when they know for sure that such information is erroneous (Lewandowsky et al. 2012).

Maintenance of an individual’s social identity is an important influence on his or her invocation of System 1 or System 2 thinking. Evidence suggests that individuals tend to adopt the views of the peer groups that are most salient to them, even if the “objective” or “factual” information available to them contradicts those views. (Asch 1951 performed the classic “conformity experiments” that demonstrated this phenomenon in the early 1950s.) Uncritical System 1 thinking is active in processing information that is consonant with the beliefs and attitudes of those peer groups. Critical and skeptical System 2 thinking is active in processing information that is dissonant to those groups’ beliefs. These effects (that individuals tend to accept salient group norms) are even more pronounced in an anonymous environment, such as that which characterizes much online interaction (Postmes et al. 2001).

Lastly, there is evidence that emotion and motivation affect cognition. For example, people who are angry tend to rely more heavily on simple heuristic cues (suggestive of System 1 thinking) than those who are not angry (Bodenhausen, Sheppard, and Kramer 1994). Individuals are more likely to stereotype people (a form of System 1 thinking) when that stereotype is consistent with their desired impression of those people; conversely, when the stereotype is inconsistent with their desired impression, individuals tend to inhibit the use of this stereotype (Kunda and Sinclair 1999). Negative emotions (such as those induced by the receipt of information incongruent with a person’s prior beliefs) can improve the ability of a person to reason logically, thus enabling him or her to negate or discount that information (Goel and Vartanian 2011).

In the new information environment, exploitation of human cognitive architecture and capabilities – which are largely unchanged from what existed millennia ago – provides the 21st century information warrior with cyber-enabled capabilities that Hitler, Stalin, Goebbels, and McCarthy could have only imagined. By exploiting cognitive limitations, the perpetrators of cyber-enabled information warfare have learned to exacerbate prejudices, biases, and ideological differences; to add heat but no light to political discourse; and to spread widely believed “alternative facts” in advancing their political positions.

Russian interference in the 2016 US presidential election has dominated news headlines ever since. But interference by authoritarian countries in the elections of democratic states – as undesirable and threatening as it may be – is hardly the only negative consequence of cyber-enabled information warfare. The problems of nuclear war and climate change are hard enough to solve even when well-intentioned, well-informed parties on all sides share a basic understanding and knowledge of the relevant facts. Yes, they may have different values and different priorities, may act under different constraints, and be able to bring to bear different levels of resources to these problems.

But without shared, fact-based understandings of the blast, thermal, and radiation effects of nuclear explosions, what hope is there for national leaders to reach agreements to reduce the threat of nuclear holocaust [war] or to make good decisions about nuclear weapons use in times of crisis? Without shared, fact-based understandings that rising atmospheric carbon dioxide concentrations caused by human beings result in corresponding increases in global temperature and climatic disruption, what hope is there for national leaders to reach agreements to begin serious efforts at decarbonizing their economies?

Climate change denialism

Climate change denialism can be fairly characterized as cyber-enabled information warfare against the reality of large-scale anthropogenically-induced climate change. In the responses of people resistant to taking action to mitigate climate change, we see a number of psychological factors at work (Zaval and Cornwell 2016). For example, one key element of System 1 thinking is the availability heuristic, with which individuals tend to associate the likelihood of an event with the ease with which they can remember similar events in the past. But the long-term consequences of climate change are unprecedented in recorded human history and obviously people have no personal memories of unprecedented events.

Moreover, climate change is a long-term process whose inexorable progression is easily masked by short-term fluctuations in local weather conditions. For example, public concerns about climate changes correlate with local weather conditions (Krosnick et al. 2006). Climate change deniers are also quick to flag for public attention days that are particularly cold as “evidence” that global warming is not occurring and thus, they claim, discrediting theories of climate change. This illustrates a bias known as attribute substitution, as Kahneman and Frederick (2002) describe, through which individuals substitute salient information (such as the cold temperature today) for information that is more relevant but harder to understand (such as information about global climate change).

People are also subject to a loss-aversion bias, in which they place greater weight on losses than gains of equal value. In 1992, the United States committed itself to the United Nations Framework Convention on Climate Change, although President George HW Bush also stated that “the American way of life is not up for negotiation” – and in 2018, the United States withdrew from the Paris Agreement (which was based on the convention). The argument? That the United States would have to give up too much if it kept to the agreement.

To close this (merely illustrative) exploration of biases relevant to climate change denialism, the optimism bias suggests that people consider themselves exceptions when considering the likelihood of a negative event occurring. That is, bad things may happen to other people, but they won’t happen to me, even though I and those other people are similar in important and relevant ways. In a climate context, the bad things may involve sea level rise or heat waves – and the misperception that “others may suffer from such problems but I won’t” diminishes the power of personal concern as a driver for rational decision making.

Connecting the operation of these cognitive biases to the affordances of modern information technologies is not difficult. For example, Roxburgh et al. (2019) demonstrate how the characteristics of specific weather events (e.g. hurricanes or snowstorms) and “short-term socio-political context can play a critical role in determining the lenses through which climate change is viewed.” Note especially the importance of “short-term socio-political context” – precisely the context that social media shapes.

Elsasser and Dunlap (2013) noted the influential role of a variety of newspaper columnists in advancing denialist arguments and thus amplifying these arguments to a broad segment of the American public. Fewer in number then, essentially all columnists today (of all political leanings) have a social media presence that they use to publicize their work, and in many instances their online presence is driven in significant part by social media and reach many more readers online than in print. Furthermore, subtleties and nuances in their extended written pieces are likely to be lost when they are represented in social media.

Another important element of climate change denialism is the easy accessibility of seemingly-authoritative information that casts doubt on the well-established science of climate change. As reported by The Guardian, a variety of largely secret funding sources distributed $118 million to 102 denialist organizations (Goldenberg 2013). Oreskes and Conway (2011) provide the definitive work on deliberate information campaigns to obscure the scientific truth on a range of issues from smoking to climate change. These denialist organizations have generated a variety of products for public and policy consumption (but – unsurprisingly – not many peer-reviewed scientific articles) that are easily accessible to the public, mainstream media outlets, and policy makers. Their products are broadly disseminated through social media and easily found through customized search, and they are sought by reporters who seeking to cover “both sides” of a controversy that is intellectually equivalent to a “controversy” about whether the earth is round or flat.

Nuclear conflict

On the risks of nuclear conflict, theories and approaches to nuclear deterrence and strategic stability developed prior to the collapse of the Soviet Union in the late 1980’s and early 1990’s rest on the presumption of rationality in national decision makers. In particular, they assume that adversaries are deterred from attacking by a threat of retaliation that would impose costs on the adversary that would outweigh any conceivable benefits that it would gain from an attack (Morgan 2003). Central to this assumption is a rational adversary that can and does make a calculation of expected costs and benefits, compares them, and then acts accordingly.

But the psychologically informed understanding of realworld decision making described above was not accepted widely in the scientific literature until approximately the same time as the collapse of the Soviet Union, and the seminal work in such understanding occurred only in the decade previous to that. What a psychologically-informed understanding of real-world decision making tells us is that the rationality assumption at the base of much traditional thinking on deterrence and strategic stability is untenable, given that humans have evolved to rely on intuitive, reflexive, heuristic System 1 thinking to make decisions, particularly when faced with time pressures, surprise and other obstacles to the deliberate calculation implied by System 2 thinking (Kahneman 2011). Psychology tells us that – more often than not – the fast, intuitive judgements of System 1 often take precedence over the slower, more analytical thinking of System 2.

The challenges posed by reflexive reliance on System 1 thinking are greatly accentuated by characteristics of today’s information environment. Social media networks in particular are optimally designed to stimulate System 1 thinking – emotional, reflexive, immediate – and they rapidly transmit content among like-minded individuals, creating the ideal conditions for public polarization and divisiveness to occur (Pfeffer, Zorbach, and Carley 2014). Multiple narratives rapidly emerge around complex events; citizens splinter into their own informational universes and are unable to agree on an underlying reality. Political leaders themselves are subject to these conflicting narratives and may even be active and influential participants in one or another of them.

It is thus easy to posit that in this information environment, manipulated information – either artificially constructed or adopted by a strong grassroots base – could be used by interested parties to generate pressure on leaders to act. At the same time, leaders themselves are likely to be facing information overload and less able to distinguish analyzed information from their own intelligence sources and other, unvetted information originating from their constituencies.

The coming information dystopia

Nuclear war and climate change are arguably the most important existential challenges today that are compounded by the corruption of the information ecosystem. But even if a single miraculous stroke the laws of physics were changed to make nuclear weapons impossible to build and operate and to immediately eliminate anthropogenic emissions at zero cost, cyber-enabled information warfare could still can lead to an information dystopia. Here are some possible elements:

● Adversaries manufacture numerous graphic videos of American soldiers (complete with sound effects) committing battlefield atrocities, and spread them widely through the Internet. Once upon a time, highquality video forgeries were difficult and expensive to make. AI-based technologies will bring this socalled deepfake capability to the masses, and anyone with imagination, a modicum of technical skill, and a personal computer will be able to distribute reasonably realistic forgeries. Denials will be issued but of course will also not be believed by large fractions of viewers. Even if proof of inauthenticity can be provided, such evidence will not affect the responses of many viewers.

● Political campaigns conduct similar efforts to discredit political opponents (e.g. “showing” an opponent making controversial or disqualifying remarks before an election). But they also use the existence of deepfake technologies to deflect attention from authentic and real evidence of their own political and personal misdeeds. For example, a real video of a candidate punching an old lady who supports his opponent will be dismissed as “one of those deepfakes that anyone could have produced.”

● Financial markets are disrupted by falsified videos of CEOs making announcements regarding company prospects that are much more pessimistic than expected. Attempts to correct the record are drowned out in a subsequent flood of contradictory information, all of which appear at first glance to be authentic.

● Public safety is compromised by reports of local disasters (e.g. explosions of chemical plants that result in the release large amounts of toxic gases). These reports, along with “authentic” video of people choking amidst locally familiar locations (e.g. well-known fields or sport stadiums), cause spontaneous mass evacuations. Contradictory directions for evacuation broadcast using social media result in chaos on the streets and highways.

● Public health is placed at risk when the safety and efficacy of medical treatments known to be safe and effective are publicly questioned through active disinformation campaigns conducted on the Internet and in bookstores. Attempts to provide valid information are met with responses such as “that’s what the pharmaceutical companies and medical establishment want you to think, but just look at what’s happened to our children.”

● Children in schools are threatened by online campaigns to spread rumor, innuendo, and positive or negative information about various students. Conducting such campaigns for pay becomes the business model of entrepreneurs who advertise that they can guarantee admission to selective colleges, boost the social standing of the children of their clients, or take revenge on those who have harmed such children, all in anonymous and untraceable ways.

● Journalists, political leaders, and judges are compromised by artfully forged emails and alterations to other documents that are mixed with entirely authentic leaked emails and documents and are indistinguishable from them.

A world with these elements – and many more comparable ones – will be the inevitable result if and when deployment and use of the tools of cyber-enabled information warfare become widespread. And even more troubling is the fact that not every bit of information needs to be corrupted for this dystopian outcome to occur – it will require only a fraction of it to be corrupted for people to lose faith entirely in “objective” and “trustworthy” sources of information, the result of which will be that people will fractionate into their own information realities.

Fearing the end of the enlightenment

The Enlightenment established reason and reality as the foundational pillars of civilized discourse. In such discourse, logic matters, and a logical contradiction between statement A and statement B means that at least one of those statements is false. The truth of a statement about the world is tested by its correspondence to objective reality rather than by how many people believe it; that is, empirical data are influential. Furthermore, statements known to be wrong or false do not affect conclusions or choices between alternative courses of action.

Cyber-enabled information warfare provides the tactics, tools, and procedures – in short, the means – to replace the pillars of logic, truth, and reality with fantasy, rage, and fear. In a world of ubiquitous cyber-enabled information warfare, communication and information inflame passions rather than informing reason, play to the worst in people’s cognitive architectures rather than the best, and divide rather than unify. Deliberate corruption of the information ecosystem could be seen as an analog of poisoning water supplies that can be done remotely, inexpensively, and anonymously. All of this is just another way of saying that today it is possible to see glimmerings of an anti-Enlightenment that can possibly take root and that would indeed be the end of civilization as we know it.

#### Anticompetitive effects of AI are nascent but enabled by government inflexibility and arbitration---facilitating private litigation solves.

Banicevic et al. ’18 [Anita, Gabrielle Kohlmeier, Dajena Pechersky, and Ashley Howlett; Fall 2018; Coalition of attorneys working under the technology compliance working group, including legal experts from Davies Ward Phillips & Vineberg LLP, and the working group chair; Compliance and Ethics Spotlight, “Algorithms: Challenges and Opportunities for Antitrust Compliance,” p. 8-12]

III. Anticompetitive Effects of Algorithms

Although algorithms provide significant competitive benefits for firms and consumers, they present novel challenges for competition authorities—and by extension, compliance teams. In particular, there is a rising concern that algorithms increase the potential for both tacit and overt collusion.29

Collusion occurs when competing firms coordinate or act in a manner to set prices above the market equilibrium with the objective of increasing profits. While this may result in short-term profits to the market actors involved in the arrangement, collusion harms consumers, is anticompetitive, and leads to poor long-term economic outcomes.30

There are two forms of collusion, overt and tacit. In a free market, firms are free to act to maximize their profits, as long as they are acting independently. Overt collusion refers to anticompetitive conduct that is facilitated through explicit agreements, written or oral. It is illegal in most jurisdictions around the world. Tacit collusion, also known as conscious parallelism, refers to anticompetitive conduct that can be achieved without explicit coordination between competing firms. Although competing firms decide their profit-maximizing strategies independently, they are able to arrive at and maintain a noncompetitive outcome. Conscious parallelism is generally not illegal because there is no underlying illegal behavior or action to sanction.31

Algorithms that facilitate overt collusion present no real new legal challenges to competition authorities. If a firm is found to have coordinated with other market actors through algorithms, provided there is evidence of direct or indirect contact showing there to be an explicit agreement between market actors, competition authorities have the necessary tools to discipline the actors involved in the arrangement.32

The challenge lies with tacit algorithmic collusion. Some scholars believe that algorithms could become capable of facilitating collusion without the need for any communication or coordination between market actors. Dynamic pricing algorithms, for instance, can assess and adjust prices for thousands of products and services in milliseconds in response to changes in competitor prices. The concern is that, as a result, firms might choose not to discount products and services if they perceive the benefit to be short-lived. Some have argued that this leads to tacit collusion since firms are seen as able to “set” and maintain market prices through algorithms without explicitly coordinating.33

In most algorithms, market actors can input the general rules that map inputs to outputs or they can at least decrypt the algorithm to understand why an algorithm behaves a certain way. Some academics have argued that deep learning algorithms create further issues for competition authorities. They have posited that the development of deep learning algorithms means that market actors may not necessarily know how or why a particular algorithm arrives at particular outputs as such algorithms would essentially act autonomously towards potentially anticompetitive outcomes.34

The concept of tacit algorithmic collusion is disputed, however, with prominent scholars presenting data showing that, currently, such collusion is no more than an unproven theory.35 Professor Salil Mehra commented before the Federal Trade Commission (FTC) that “[t]here has been scaremongering based on fears that artificial intelligence will somehow destroy competition as we know it.” In his view, “these fears are premature . . . because technological development is still far from creating some sort of autonomous algorithmic cartel robot.”

As economist Ai Deng explains, although some algorithms may do a better job than humans at establishing cooperative relationships, a number of assumptions underlying antitrust fears about algorithms have little, if any, empirical support.36 Real world competitive decisions are highly complex, presenting significant computational and technical challenges, and current research shows that designing algorithms capable of learning to cooperate in such an environment is very difficult. In fact, a growing body of theoretical and experimental economic literature posits that algorithms have to learn to communicate with one another and react to communications to achieve a collusive outcome in a market with more than two market actors. Although this may be theoretically possible, the development of algorithms that are able to communicate in this way is still in its very early stages.37

The debate on whether tacit algorithmic collusion is a possibility is far from settled. However, the existence of that debate does not detract from the fact that algorithms have arrived and have materially influenced the way in which market actors execute their functions.

IV. Algorithms and Antitrust Compliance in the United States

A. Collusion

Section 1 of the Sherman Act, 15 U.S.C. § 1, prohibits all agreements that unreasonably restrain trade. When discussing the application of the traditional analysis of illegal collusion in the algorithmic pricing context, U.S. antitrust officials have emphasized the importance of the existence of an agreement, which is often missing or harder to detect when algorithms are used. As Maureen Ohlhausen has explained, “[t]he type of technology used to communicate with competitors is wholly irrelevant to the legal analysis. Whether it is phone calls, text messages, algorithms or Morse code, the underlying legal rule is the same—agreements to set prices among competitors are always unlawful.”38

The U.S. Supreme Court has defined an agreement as “a unity of purpose or a common design and understanding, or a meeting of the minds in an unlawful arrangement.” 39 This broad definition includes not only explicit agreements to set prices collusively, but also exchanges of competitively sensitive, nonpublic information between competitors for anticompetitive purposes. The antitrust laws do not, however, prohibit companies from engaging in conscious parallelism or interdependent pricing, where an agreement is not present.40 The mere fact that competitors monitor, and even match, each other’s public pricing information is not sufficient to establish a Section 1 violation.

During a hearing before the Senate Subcommittee on Antitrust, Competition Policy, and Consumer Rights on October 3, 2018, Assistant Attorney General Makan Delrahim stated that the use of algorithms to create anticompetitive effects is an “important issue” that antitrust enforcers are struggling with both in the U.S. and in Europe.41 For the most part, antitrust enforcers in the U.S. agree that allegations of collusion involving algorithms should be analyzed under the same legal standards as any other collusive conduct.42 Barry Nigro, a deputy assistant attorney general in the Antitrust Division of the Department of Justice (DOJ), has reportedly stated that “when analyzing whether conduct constitutes collusion, an observer should ‘take out’ the fact that an algorithm was involved[.]”43 Nonetheless, some officials have noted that while U.S. antitrust laws generally provide the tools necessary to address anticompetitive conduct facilitated by algorithms, the use of algorithms may make it harder to identify collusion and may require antitrust enforcers to employ new investigative techniques.44 With respect to the difficulty of detecting algorithm-enabled collusion, Maureen Ohlhausen, former Acting Chairman of the FTC, has commented that, as with other types of price-fixing conduct, the DOJ’s leniency program and the threat of criminal penalties should incentivize self-detection and cooperation with enforcers where external detection is not enough.45

The DOJ and private plaintiffs have brought cases involving price-fixing using algorithms or other nontraditional electronic tools. For example, in U.S. v. Topkins, 46 the first antitrust e-commerce criminal prosecution, DOJ charged David Topkins and his coconspirators with using pricing algorithms to engage in a conspiracy to fix the prices of posters sold in the Amazon Marketplace. Specifically, the conspirators “agreed to adopt specific pricing algorithms for the sale of the agreed upon posters with the goal of coordinating changes to their respective prices.”47 Topkins pleaded guilty to price fixing in violation of Section 1 and agreed to pay a $20,000 criminal fine.

In Meyer v. Kalnick, 48 private plaintiffs alleged that drivers agree with Uber to charge certain fares with the understanding that all other Uber drivers are agreeing to charge the same fares. They further alleged that this agreement is facilitated by Uber’s pricing algorithm, which all drivers use to set their fares. The plaintiffs alleged that this arrangement amounted to a hub-and-spoke conspiracy in violation of Section 1.49 The court denied the defendant’s motion to dismiss, but Uber subsequently moved the matter to arbitration.

In October 2018, Assistant Attorney General Makan Delrahim stated that the DOJ had a criminal case that he expected would come to a conclusion in the next two weeks related to the use of search algorithms by competitors to effectuate price fixing, which he added would be the “first of its kind.” 50 Further details on the case that he was referring to have not yet been released.

B. Price Discrimination

Pricing algorithms can be used to discriminate among buyers. For example, a seller may use an algorithm to charge different prices or offer different discounts to customers located in different states. The DOJ and FTC have said that algorithmic price discrimination should be analyzed using the same legal framework as price discrimination in other contexts, and “[a]lgorithmic pricing that leads to price discrimination—without incorporating competitor data—is unlikely to raise competition concerns.”51

A seller who charges competing buyers different prices for the same product may be liable for price discrimination under the Robinson-Patman Act, 15 U.S.C. § 13(a). The U.S. antitrust agencies have not actively enforced the act in decades.52 Some enforcers have noted, however, that algorithmic pricing may lead to new and more sophisticated methods of price discrimination.53 Private parties continue to sue under the act’s provisions.54 Companies using algorithmic pricing may face litigation from private plaintiffs if the algorithm facilitates actionable price discrimination.

As algorithms are increasingly used by businesses to narrowly target specific types of customers, their use might enhance the risk of a merger challenge where markets can be defined by differential pricing. The 2010 Horizontal Merger Guidelines55 authorize the agencies to challenge a proposed merger if it is likely to be anticompetitive in any relevant market, no matter how small. The Merger Guidelines instruct that the antitrust agencies will evaluate the possibility of price discrimination against targeted customers and further provide that “[w]hen discrimination is reasonably likely, the agencies may evaluate competitive effects separately by type of customer.”56

Businesses are empowered by sophisticated pricing algorithms to process large amounts of data and analyze more granular data regarding consumer characteristics than was previously possible. Algorithms enable sellers to segment their customers into smaller and smaller groups and engage in more targeted price discrimination methods.57 The more differentiated the algorithmic pricing program is, the narrower the antitrust agencies may define the relevant markets for purposes of evaluating the competitive effects of a potential merger.58 A proposed merger that has no anticompetitive effects in one market segment may thus be found to substantially decrease competition in a smaller “price discrimination market.”

#### Enforcing antitrust against digital collusion optimizes corporate governance of artificial intelligence rollout---extinction from multiple interlocking crises.

Critch ’20 [Andrew and David Krueger; June 11; Research Scientist and Ph.D. at the Center for Human-Compatible AI at the University of California, Berkeley; Executive Director of the Dialogue Institute, Ph.D. from Temple University; Arxis, “AI Research Considerations for Human Existential Safety,” p. 33-85]

Hazardous deliverables. Supposing humanity develops highly advanced AI systems, those systems could aid humans in developing other technologies which would themselves pose significant global risks to humanity. Nuclear weapons, chemical weapons, and bioweapons are examples of such hazardous technologies that have been developed in the past, without the aid of AI technology.

Risks arising from the development of more such hazardous technologies in the future—with or without the assistance of AI in the development process—are not explicitly addressed by the technical directions of this report. However, such risks could be addressed by related principles of safe and ethical oversight.

Suboptimal futures. More generally, it has been argued that futures where humans exist, but are not flourishing to the degree one would hope, should be considered existential risks or at least be treated with the same degree of severity as human extinction risks. For example, Bostrom (2013) considers “permanent stagnation” and “flawed realization” scenarios, wherein human civilization respectively either “fails to reach technological maturity” or “reaches technological maturity in a way that is dismally and irremediably flawed”. These scenarios are excluded from this report for two reasons. The first reason is to avoid debate in this report the issue of what constitutes a suboptimal future, as discussed somewhat in Section 2.9. The second reason is that these other risks do not naively belong under the heading “existential”, so most readers are not likely to be confused by their omission.

4 Flow-through effects and agenda structure

Sections 5, 6, 8, and 8 of this report may be viewed as a very coarse description of a very long-term research agenda aiming to understand and improve interactions between humans and AI systems, which could be viewed as ongoing throughout the full historical development of artificial intelligence, multi-agent systems theory, and human-computer interaction.

How can one begin to account for the many ways in which progress in different areas of AI research all flow into one another, and how these flowthrough effects relate to existential risk? The task is daunting. To organize and reduce the number of possible flow-through effects one would need to consider, the research directions in this report have been organized under the subsections of Sections 5, 6, 8, and 9, which themselves are related by a lattice structure depicted in Figure 7.

4.1 From single/single to multi/multi delegation

Research on single/single delegation can be expected to naturally flow through to a better understanding of single/multi and multi/single delegation, and which will in turn flow through to a better understanding of multi/multi delegation.

4.2 From comprehension to instruction to control

Sections 5, 6, and 8 are each divided into subsections regarding the human ability to either comprehend AI systems, instruct AI systems, or control AI systems, as defined in Section 2.7. Within each section, comprehension research can be expected to benefit but not subsume instruction research, and comprehension and instruction research can be expected to benefit but not subsume control research.

4.3 Overall flow-through structure

Put together, the flow-through effects discussed above combine to yield the lattice depicted in Figure 7 below. This lattice defines the overall organizational structure for Sections 5, 6, 8, and 9, and summarizes the bulk of the “discovery flow-through” effects that should be expected between research directions in this report. Whenever a research direction would contribute to multiple corners of this subsection lattice, it is discussed under the earliest relevant subsection, leaving its usefulness to subsections further down in the lattice to be implied from the document structure.

4.4 Research benefits vs deployment benefits

Suppose that a major breakthrough is made in single/single delegation, but that multi/multi delegation remains poorly understood. If the breakthrough leads to the release of several AI systems each intended to serve a different human stakeholder, then a multi/multi interaction scenario immediately results. In such an event, the R&D process that designed the AI systems will not have accurately accounted for the interaction effects between the multiple humans and systems. Hence, many errors are likely to result, including safety issues if the AI systems are sufficiently impactful as a collective.

In the preceding scenario, single/single research flows through to a harm, rather than a benefit, in a multi/multi deployment setting. Such scenarios can make it very confusing to keep track of whether earlier developments will help or hinder later developments. How can one organize one’s thinking about such flow-through effects? One way to reduce confusion is to carefully distinguish research benefits from deployment benefits. While research on earlier nodes can be reasonably expected to benefit research on later nodes, the opposite effect can hold for deployment scenarios on later nodes. This happens when research on an earlier node results in a premature deployment event in a setting where research on a later node was needed to ensure proper functioning. For instance, Figure 8 summarizes a causal pathway whereby research on single/single delegation could robustly lead to realworld errors in multi/multi delegation.

Of course, it is common sense that the premature distribution of a powerful new technology can be hazardous. However, combined with the observation that single/single systems can easily be replicated to yield a multi/multi interaction scenario, the potential for premature deployment implies that an understanding of multi/multi delegation for powerful systems may be needed in short order after the development of any powerful single/single delegation solutions. For any AI technology with the potential for global impact, this observation should not be taken lightly. Society may typically learn to correct premature deployment errors through experience, but an error that yields a human extinction event is not one that we humans can learn from and correct later.

4.5 Analogy, motivation, actionability, and side effects

In the next few sections, the reader may soon notice a series of repeated subheadings, intended to suggest a methodology for thinking about long-term risks. The intended meaning behind these subheadings will be as follows:

* “Social analogue”. These subsections are post-hoc analogies for introducing each research direction by comparing desired AI system properties with typical human properties. The analogies can only be fitting to the extent that AI systems might be designed to operate according to similar principles as humans. Hence, the motivation and actionability subsections (below) aim to give more precise illustrations that are intended to expand, clarify, and supersede these analogies.
* “Scenario-driven motivation”. These subsections explain the final causal pathway through which a given research direction could be used to reduce existential risk. In aggregate, this content is intended to illustrate just some of the many technical and social mechanisms through which AI research and existential safety are intertwined. Motivations for some sections may be directly at odds with other sections. At best this suggests a hedged portfolio of approaches to existential safety; at worst, some approaches may need to be cut short if they present serious negative externalities.
* “Instrumental motivation”. These subsections explain how a given research direction could be steered and applied to benefit other research directions in this report.
* “Actionability”. These subsections aim to provide illustrative examples of existing work relevant to a given research direction. This report falls woefully short of providing fair and comprehensive overviews of the large corpora of work relevant to each direction, and for this the authors apologize in advance.
* “Consideration of side effects”. These subsections examine ways in which particular research ideas could be taken in directions that would be problematic from an existential safety perspective. The fact that many research directions are “dual purpose” in this way seems unavoidable: when examining capabilities relevant to existential risk, there is always the possibility that poor judgments about how to intervene on those capabilities could make matters worse.

5 Single/single delegation research

This section begins our examination of research directions relevant to existential safety in the delegation of tasks or responsibilities from a single human to a single AI system.

Consider the question: how can one build a single intelligent AI system to robustly serve the many goals and interests of a single human? Numerous other authors have considered this problem before, under the name “alignment”. For a diversity of approaches to AI alignment, see Soares and Fallenstein (2014); Taylor et al. (2016); Leike et al. (2018).

The AI alignment problem may be viewed as the first and simplest prerequisite for safely integrating highly intelligent AI systems into human society. If we cannot solve this problem, then more complex interactions between multiple humans and/or AI systems are highly unlikely to pan out well. On the other hand, if we do solve this problem, then solutions to manage the interaction effects between multiple humans and AI systems may be needed in short order.

(Despite the current use of the term “alignment” for this existing research area, this report is instead organized around the concept of delegation, because its meaning generalizes more naturally to the multistakeholder scenarios to be considered later on. That is, while it might be at least somewhat clear what it means for a single, operationally distinct AI system to be “aligned” with a single human stakeholder, it is considerably less clear what it should mean to be aligned with multiple stakeholders. It is also somewhat unclear whether the “alignment” of a set of multiple AI systems should mean that each system is aligned with its stakeholder(s) or that the aggregate/composite system is aligned.)

Social analogue. As a scenario for comparison and contrast throughout our discussion of single/single delegation, consider a relationship between a CEO named Alice who is delegating responsibilities to an employee named Bob:

* (comprehension) In order to delegate effectively to Bob, Alice needs some basic understanding of how Bob works and what he can do— Alice needs to comprehend Bob to some degree.
* (instruction) Alice also needs to figure out how to explain her wishes to Bob in a way that he will understand—to instruct Bob.
* (control) If Bob genuinely wants to enact Alice’s wishes as she intends them, that is a good start, but he can still falter, perhaps catastrophically. Perhaps he might ignore or severely misinterpret Alice’s instructions. So, Alice also needs some systems in place to control Bob’s involvement in the company if he begins to behave erratically. For instance, she should be able to revoke his computer system or building access if needed. As Bob’s employer, Alice also maintains the legal authority to fire him, at which point other company employees will typically stop accommodating his plans.

Consideration of side effects. There are a number of potentially negative side effects of developing single/single delegation solutions in general, which are included here to avoid repetition:

* (racing) If near-prepotent AI systems are eventually under development by competing institutions, single/single delegation solutions might increase the willingness of the systems’ creators to move forward with deployment, thereby exacerbating Type 2a risk (unsafe development races).
* (enfeeblement) Widespread consumer dependence on single/single AI systems could lead to Type 2c risk (human enfeeblement) if the systems take on so many mental and physical tasks that human capabilities begin to atrophy.
* (misleading safety precedents) Single/single delegation solutions that only work for non-prepotent AI systems could create a false sense of security that those solutions would scale to near-prepotent and prepotent systems, increasing Type 1c risk (unrecognized misalignment). For instance, “just turn it off when it’s malfunctioning” is a fine strategy for many simple machines, but it won’t work if the AI system is too pervasively embedded in key societal functions for shutting it down to be politically viable (e.g., food distribution), or if the system will develop and execute strategies to prevent humans from shutting it down even when they want to.
* (premature proliferation) If single/single delegation solutions are deployed broadly without sufficient attention to the multi/multi delegation dynamics that will result, the resulting interaction between multiple humans and/or multiple AI systems could be destabilizing to society, leading to as-yet unknown impacts. This general concern was discussed in Section 2.8.1.

5.1 Single/single comprehension Comprehending a human employee is quite different from comprehending an AI system. Humans have many cognitive features in common, due to some combination of common evolutionary and societal influences. Therefore, a human may use an introspective self-model as a stand-in for modeling another person—to “put oneself in someone else’s shoes”. By contrast, artificial intelligence implementations are by default quite varied and operate very differently from human cognition. A recent and salient illustration of the difference between machine and human intelligence is the vulnerability of present-day image classifiers to the perturbations that are imperceptible to humans Szegedy et al. (2013), due the many degrees of freedom in their high dimensional inputs Goodfellow et al. (2014). For instance, Su et al. (2017) trained an All Convolutional Network to achieve 86% accuracy on classifying images in the CIFAR-10 database of 32 × 32 images, and found that 68.36% of the images could be transformed into a misclassified image by modifying just one pixel (0.1% of the image), with an average confidence of 73.22% assigned to the misclassification. As well, Athalye et al. (2017) developed a method for constructing physical objects that are deceptive to machine vision but not to human vision. The method was used to construct a toy replica of a turtle that was misclassified as a rifle from almost all viewing angles, by TensorFlow’s standard pre-trained InceptionV3 classifier (Szegedy et al., 2016), an image classifier with a 78.0% success rate of classifying ImageNet images using the “top-1” scoring rule. The fact that the image classifier networks in these experiments tend to fail outside their training sets means that the networks themselves have difficulty generalizing. This alone is not a problem with human/AI comprehension. However, the fact that the networks fail in ways that humans find surprising means that our own understanding of their capabilities is also prone to generalizing poorly. In particular, humans are unlikely to be able to comprehend AI systems by generalizing from simple analogies to other humans. As such, research specifically enabling human/AI comprehension will likely be needed to achieve and maintain a reasonable level of understanding on the part of human users and even AI developers. 5.1.1 Direction 1: Transparency and explainability One approach to improving human/AI comprehension is to develop methods for inspecting the inner-workings of the AI system (transparency), or for explaining the counterfactual dependencies of its decisions (explainability). These techniques can then be used guide R&D by helping engineers to better understand the tools they are building. Perhaps good metrics for transparency and/or explainability could be used as objectives to guide or constrain the training of complex systems. Together, transparency and explainability are sometimes called “interpretability”. Social analogue. Businesses are required to keep certain records of decisions made and actions taken in order to remain amenable to public oversight, via government agencies such as the IRS. This makes the expenditure of business resources on illegal activities at least somewhat difficult. If one views an AI system as somewhat analogous to a corporation—a non-human entity which nonetheless pursues an objective—one might hope to impose analogous internal record-keeping requirements that could be used by humans to detect undesirable cognitive patterns before they would manifest in harmful actions. Doing so would require a degree of transparency to the humans imposing the requirements. Scenario-driven motivation. The decision to deploy a powerful AI system should come with a high degree of confidence that the system will be safe, prior to system being deployed. In particular, the researchers and developers responsible for the system should have enough insight into the its inner workings to determine that it is not misaligned and prepotent. Just as business tends to move faster than governance, powerful AI systems will likely eventually operate and make decisions on a time scale that is too fast for humans to oversee at all times. The more we are able to understand how such systems work, the less likely they will be to surprise us. Thus, AI transparency improves our ability to foresee and avert catastrophes, whether it be with a powerful AI system or a rudimentary one. Explainability, or after-the-fact transparency, also serves to improve human predictions about AI systems: aside from explanations informing humans’ future predictions about what the system will do, if we impose explainability as a constraint on the system’s behavior, we might avert at least some behaviors that would be surprising—to the point of being inexplicable—to the human. Hence, this direction could apply to reducing Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment), by helping us to understand and predict the prepotence and/or misalignment of a system before its deployment. Transparency and explainability techniques could also be used to reduce Type 1d risks (involuntary MPAI deployment), such as by enabling the inspection any AI-dependent computer security infrastructure in use by AI development teams. Actionability. There is already active research working to make the decisions of modern machine learning systems easier to explain, for instance, Yosinski et al. (2015) and Olah et al. (2017) have created visualization tools for depicting the inner workings of a neural network. While the decisions made by a neural network routinely combine thousands of variables under intricate rules, it is in principle possible to locally approximate arbitrarily complex decisions by identifying a small number of critical input features that would most strongly affect the output under relatively small changes. This can be used to provide tractable “local” explanations of AI decisions that might otherwise be difficult or impossible for humans to comprehend (Ribeiro et al., 2016). Modifying the objective function or architecture of a machine learning system to require a degree of explainability to human inspectors could result in systems that are more legible to human overseers (Zhang et al., 2018). One might hope to achieve better generalizability than most earlier work on explainability for AI systems, such as Van Lent et al. (2004). Perhaps quantitative models of pragmatic communication (Goodman and Stuhlmüller, 2013), wherein speakers and listeners account for one another’s goals to communicate and thereby cooperate, could be useful for representing objective functions for explainability. Or, perhaps sparse human feedback on the understandability of a self-explaining ML system could be augmented with frequent feedback from an automated dialogue state-tracking system, e.g., as studied by Henderson et al. (2014). This would mean repurposing the dialogue state-tracking system to give quantitative feedback on the understandability of the outputs of the self-explaining system, based on the state-tracker’s experience with understanding human dialogue. Explanations in natural language are an active area of exploration, e.g., by Hendricks et al. (2016). The use of natural language is promising because it is in principle infinitely expressive, and thus opens up a wide space of possible explanations. However, their technique currently produces afterthe-fact “rationalizations” that do not always correspond to the decision procedure actually employed by the AI system in each classification instance. Further work on producing natural language explanations should focus on ensuring faithfulness to the underlying reasoning of the system in each decision instance. As Hendricks et al. remark, future models could “look ‘deeper’ into networks to produce explanations and perhaps begin to explain the internal mechanism of deep models”. This objective is critical: the goal of explainability should be to inform human users, never to appease or convince them. By contrast, if explanations are optimized merely to convince the human of a foregone conclusion, the system is essentially being trained to deceive humans in situations where it has made a mistake. Starting down the path of developing such deceptive AI systems might exacerbate Type 1b, 1c, and 1d risks (unrecognized prepotence, unrecognized misalignment, and involuntary MPAI deployment). Robotic motion planning is another area of application for transparency. Using a simple model that treats humans as Bayesian reasoners, robots can adjust their motion using that model to more legibly convey their goal to a human collaborator (Dragan et al., 2013), and plan action sequences that will be easier for humans to anticipate (Fisac et al., 2016). Studies of mutual adaptation in human-robot collaboration seek to account for humans’ ability to infer and conform to the robot’s plan while also expecting it to reciprocate (Nikolaidis et al., 2016). To guide progress in any application area, it would be useful to understand the features of transparency and explanation that (1) humans instinctively prefer, and (2) aid in improving human judgment. For example, humans tend to prefer certain features in the explanations they receive, including simplicity (Lombrozo, 2007) and “exportable dependence”, i.e., usability of the explanation for future predictions and interventions (Lombrozo and Carey, 2006; Lombrozo, 2010). These principles could be quantified in objective functions for training prototypical “explainable AI” systems. Consideration of side effects. One possible source of negative side effects could occur if transparency and explaiability (T&E) tools are developed which enable engineers to build much more complex systems than they would otherwise be able to construct, and if AI systems nearing prepotence turn out to be beyond the reach of the T&E methods. So, if T&E methods are developed which hasten tech development but for whatever reason cannot be applied to ensure the safety of near-prepotent systems, the result would be a precarious situation for humanity. 5.1.2 Direction 2: Calibrated confidence reports This research direction is concerned with developing AI systems which express probabilistic confidence levels that roughly match their success rates in answering questions or choosing good actions. For instance, among statements that a knowledge database system assigns a 89%-91% probability of truth, roughly 90% of those statements should turn out to be true. Expressing calibrated confidence to accompany decisions can be seen as a subproblem of transparency or explainability, but has other applications as well. Social analogue. Suppose Bob sells Alice an investment promising her a 99% chance of doubling her money by the end of the year. However, Alice also learns that among many other investments that Bob has sold claiming “over a 95% chance of doubling”, only 65% actually doubled. Therefore, even though Bob’s “99%” recommendation claims a very good expected value, Alice does not end up believing Bob’s explicit claims about the likelihood of success. Suppose Alice also receives an investment tip from Charlie, who claims a 99% chance of doubling in value. When Alice investigates Charlie’s past performance, he has no prior record of either success or failure rates on which to base her judgment. Alice also investigates Charlie’s reasons for claiming the investment will double, and finds that Charlie has done almost no market research, and knows very little about the investment. Even without a track record, Alice is able to reason that Charlie is probably not very well calibrated, and does not end up believing his claim. Scenario-driven motivation. Ultimately, the decision to deploy a powerful AI system should come with a well-calibrated prediction that the system is non-prepotent and/or aligned, prior to its deployment. A working methodology for producing calibrated confidence reports could be used for this, in conjunction with well-codified notions of prepotence and/or misalignment. That is to say, one could ask a confidence reporting system for the probability that a given AI system is aligned and/or non-prepotent. Hence, this direction could help to address Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment). In addition, reliable confidence reports could be used to temper an AI system’s online behavior. For instance, a powerful AI system could be required to shut down or act conservatively when its confidence in the humanalignment of in its decision-making is low, thereby reducing the probability of catastrophes in general. Instrumental motivation. •Direction 10 (corrigibility). Well-calibrated uncertainty could help an AI system to recognize situations where shutdown or repair is needed. •Direction 11 (deference to humans). Calibrated confidence reports could be used to trigger increased human oversight when an AI system’s confidence in its own good performance is low (Hadfield-Menell et al., 2016b). •Direction 17 (hierarchical human-in-the-loop learning (HHL)). Correctly identifying its uncertainty also allows an AI system to make better use of a limited supply of human feedback. For instance, an RL agent can specifically request feedback about human preferences or rewards when it is less certain (Christiano et al., 2017) or when the information is expected to help it improve its policy (Krueger et al., 2016). Thus, to make marginal improvements to scalable oversight, improvements to calibration need only lead to better-than-random decisions about what kind of feedback is useful. Actionability. Efforts to represent model uncertainty in deep learning (Gal and Ghahramani, 2016; Kendall and Gal, 2017) are directly applicable to developing well-calibrated confidence reports from AI systems. There are many recent papers focussed on improving calibration for machine learning models used to make uncertain predictions or classifications (Guo et al., 2017; Lakshminarayanan et al., 2017; Lee et al., 2017; Liang et al., 2017; DeVries and Taylor, 2018; Hafner et al., 2018; Kuleshov et al., 2018). Because of the inevitability of some model misspecification in any system one might build, perfectly accurate calibration may be impossible to achieve in reality. Thus, it is important to determine when and how one can reliably achieve precise calibration, and when and how awareness of imperfect calibration (in a sense, “meta calibration”) can be leveraged to improve active learning and corrigibility. For instance, Liu et al. (2015) propose an active learning approach that accounts for a model’s inductive bias and thereby outperforms random selection of queries. Meanwhile, understanding the implications of miscalibration can motivate future work by suggesting applications of calibration solutions. As a case study, Carey (2017) provides examples of how misspecification of an RL agent’s priors in an “off-switch” game (Hadfield-Menell et al., 2016b) can lead to incorrigibility of the RL agent, via miscalibration about when to defer to the human. Consideration of side effects. The potential negative side effects of this work are similar to those of Direction 1 (transparency and explainability), i.e., the risk that these methods might accelerate tech development without scaling to apply to near-prepotent systems. One way this could occur is if calibrated safety reports are fundamentally more difficult to produce for a system with the capacity for developing a plan to deceive the safety assessment protocol. Perhaps this issue, if it arose, could be mitigated with other transparency techniques for detecting if the system is planning to deceive the safety assessment. 5.1.3 Direction 3: Formal verification for machine learning systems For any safety criterion that one could hope for a powerful AI system to meet, a combination of empirical (experiment-driven) and formal (proof/argument-driven) verification methods might be relevant and useful. This direction is about bolstering formal methods. Social analogue. When a venture capital (VC) firm chooses to invest in a start-up, they look for formal legal commitments from the company regarding how and when the VC firm will be entitled to redeem or sell its shares in the company. Suppose instead the start-up offered only a word-of-mouth agreement, appealing to fact that the VC firm has never been swindled before and are hence unlikely to be swindled now. The VC firm would likely be unwilling to move forward with the actual transfer of funds until a formal, legally enforceable agreement was written and signed by the start-up. With the written agreement, the firm can develop a greatly increased confidence that they will eventually be entitled to liquidate their investment. Scenario-driven motivation. At the point of deploying any powerful AI system or system component that could result in prepotence and/or misalignment, reliance entirely on empirical tests for alignment and/or controllability is likely to be unsatisfying and perhaps even reckless. Indeed, the test “will this system overthrow human society after it is deployed?” is not an experiment one would like to actually run. But how can one know the outcome of an experiment before running it? In other high-stakes engineering endeavors, such as building a bridge or launching a rocket, one is never satisfied with merely testing the components of the bridge or rocket, but also use formal arguments from well-established principles of physics to establish bounds on the safety of the system. Such principled analyses serve as a guide for what can and cannot be concluded from empirical findings, e.g., “if force X amounts to less than 100 Newtons and force Y amounts to less than 200 Newtons, then in combination they will amount to less than 300 Newtons”. Laying out such arguments in an explicit form allows for the identification of key assumptions which, if violated, could result in a system failure (e.g., a bridge collapse, or a rocket crash). As AI systems become more powerful, persons and institutions concerned with risks will expect to see similarly rigorous formal arguments to assess the potential impacts of the system before deployment. Some would argue that such assessments should already have been carried out prior to the deployment of widespread social media technology, given its pervasive impact on society and potential to affect the outcome of national elections. Techniques and tools for automatically generating formal assessments of software and its interaction with the real world will thus be in increasing demand as more powerful AI systems are developed. Actionability. Since many present-day AI systems involve deep learning components, advances in scalable formal verification techniques for deep neural networks could be potentially very valuable. For instance, Dvijotham et al. (2018) have developed an anytime algorithm for bounding various quantities definable from network weights, such as robustness to input perturbations. Katz et al. (2017) have adapted the linear programming simplex method for verifying or refuting quantifiable statements about ReLU networks. Akintunde et al. (2018) and Lomuscio and Maganti (2017) have begun developing methods for reachability analysis of feed-forward ReLU neural networks. Selsam et al. (2017) have developed an automated proof assistant for generating machine-checkable proofs about system performance as a step in the engineering process. Their training system, Certigrad, performed comparably to Tensorflow. For even more rigorous verification, one must also consider assumptions about the so-called trusted computing base (TCB), the core software apparatus used to interpret and/or compile code into binaries and to write and verify proofs about the code. Kumar et al. (2018) argue that verification with a very small TCB is possible with appropriate adjustments to the programmer’s workflow, and that such workflows are already possible in systems such as CakeML (Kumar et al., 2014) and Œuf (Mullen et al., 2018). In order to formally specify societal-scale safety criteria that formal verification tools would go on to verify for powerful AI systems, input may be needed from many other research directions, such as Directions 8, 7, and 13 (human cognitive models, human belief inference, and rigorous coordination models). Consideration of side effects. There is an interesting duality between design and verification in the creation of AI systems by human developers, that can be seen as analogous to the duality between training and testing in the creation of image classifiers by supervised learning algorithms. Specifically, when some fraction of formal verification specs for an AI system are withheld from the human developers who design and build the system, the withheld specs can serve as an independent test of the system’s performance (and hence also the quality of the developers’ design process). This is similar to how, after a classifier has been “built” from a training dataset by a supervised learning algorithm, a separate testing dataset typically serves as an independent test of the classifier’s accuracy (and hence also the quality of the learning algorithm). Such independent tests are important, because they reveal “overfitting” tendencies in the learning algorithm that make past performance on the training data an overly optimistic predictor of future performance on real data. Conversely, using the entirety of a supervised learning dataset for training and none of the data for testing can result in a failure to detect overfitting. The analogue for human developers designing AI systems is that including too many automated verifications for the developers to use throughout the design processes enables the developers to fix just the automatically verifiable issues and not other issues that may have been overlooked. Thus, if one publishes all of one’s available formal verification methods for testing an AI system’s performance, one impoverishes one’s ability to perform independent tests of whether the developers themselves have been sufficiently careful and insightful during the design process to avoid “over-fitting” to the specs in ways that would generalize poorly to real-world applications. This potential side effect of making too many formal verification specs publicly available can be viewed as an instance of Goodhart’s Law (Manheim and Garrabrant, 2018): “When a measure becomes a target, it ceases to be a good measure.” Simply put, if all known proxy measures for safety are made publically available in the form of automated tests, it could become too easy for reseachers to accidentally or intentionally learn to “cheat” on the test. What this means for formal verification methods is that once a useful formal safety verification standard is developed, a non-trivial decision needs to be made about whether to publish reproducible code for running the safety test (making it a “target”), or to keep the details of the test somewhat private and difficult to reproduce so that the test is more likely to remain a good measure of safety. For very high stakes applications, certain verification criteria should always be withheld from the design process and used to make final decisions about deployment. 5.1.4 Direction 4: AI-assisted deliberation Another approach to improving human/AI comprehension is to improve the human’s ability to analyze the AI system’s decisions or recommendations. In this report, AI-assisted deliberation (AIAD), refers to the capability of an intelligent computer system to assist humans in the process of reflecting on information and arriving at decisions that the humans reflectively endorse. In particular, this might involve aiding the human to consider arguments or make observations that would be too complex for the human alone to discover, or even to fully reason about after the point of discovery. AIAD can be viewed as being closely complementary with transparency and explainability (T&E): while T&E methods aim to present information in a 45 form amenable to human comprehension, AIAD would assist the humans in directing their own thoughts productively in analyzing that information. Social analogue. A busy executive can benefit greatly from the assistance of employees and expert advisors who make it easier for them to evaluate important choices. At the same time, reliance on deliberative assistance leaves the executive prone to accidental or intentional manipulation by the assistant. Scenario-driven motivation. It is possible that humanity will collectively insist on relatively simple constraints for any powerful AI system to follow, that would ensure the humans are unlikely to misunderstand its reasoning or activities. Absent such constraints, humans can be expected to struggle to understand the discoveries and actions of systems which by design would exceed the humans’ creative abilities. The better guidance one can provide to the human overseers of powerful systems, the less likely they will be to overlook the misalignment or prepotence of an AI system. Hence, AIAD could be used to address Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment). At the same time, if AIAD technologies are eventually developed, caution may be needed to prevent their use in ways that would accidentally or intentionally deceive or distract humans away from key safety considerations, especially for high-stakes applications that could be relevant to existential risk. (For instance, presentday social media services employ a plethora of interactive AI/ML systems to capture and maintain user attention, and many people report that these services distract them in ways they do not endorse.) Instrumental motivation. Improved human deliberation would be directly useful to safety methods that rely on human feedback. This includes Directions 6, 17, and 23 (preference learning, hierarchical human-in-the-loop learning (HHL), and moderating human belief disagreements) Actionability. There is also evidence that automated systems can be used to aid human deliberation on non-technical topics. The delivery of cognitive behavioral therapy (CBT) by automated conversational agents over the internet has been found to be somewhat effective for reducing some symptoms of general psychological distress, in comparison with reading an e-book (Twomey et al., 2014) or simply awaiting an in-person therapist (Fitzpatrick et al., 2017). One might therefore hypothesize that automated problem-solving agents could assist in the making of stressful or otherwise difficult decisions. Christiano (2017) has proposed a recursive framework for decomposing problems assisting deliberation, recursively named “Humans Consulting HCH (HCH)”. This method has undergone some empirical testing by a new research group called Ought.org (2017a,b). Consideration of side effects. Widespread use of AIAD could lead to unexpected societal-scale effects. For example, if humans come to rely on AIAD more than their fellow humans to help them deliberate, perhaps trust between individual humans will gradually become degraded. As well, providing AIAD without accidentally misleading or distracting the human 46 may remain an interesting and important challenge. To avoid this, it may be necessary to develop an operationalized definition of “misleading”. 5.1.5 Direction 5: Predictive models of bounded rationality Both humans and AI systems are subject to bounds on their computational abilities. These bounds will likely need to be accounted for, explicitly or implicitly, in predicting what independent and collaborative behaviors the humans and AI systems can or will exhibit. Ideally, a good model of a boundedly rational decision-making system should be able to predict what sorts of the decisions the are too hard, or sufficiently easy, for the system to make correctly with its given computational resources. Social analogue. When a law school student with a poor memory and slow reading speed fails a final examination, it is apt to attribute their failure to a lack of ability rather than a lack of desire to pass. On the other hand, if a student known to have a prodigious memory and a fast reading speed is seen to fail such an exam, it may be more appropriate to infer that they are insufficiently motivated to pass. Thus, observing the same behavior from two different humans—namely, failing an exams—lead us to different conclusions about their desires (trying to pass and failing, versus not caring much about passing). In this way, thinking informally about a person’s mental capabilities is key to making inferences about their desires. Conversely, suppose you know your attorney has the best of intentions, but nearly failed out of law school and required numerous attempts to pass the bar exam. If a serious lawsuit comes your way, you might be inclined to find a more skilled attorney. These situations have at least three analogues for AI systems: (1) humans accounting for the limitations of AI systems, (2) AI systems accounting for the limitations of humans, and (3) AI systems accounting for the limitations of other AI systems. Scenario-driven motivation. See the instrumental motivations. Instrumental motivation. Numerous directions in this report would benefit from the ability to calculate upper and lower bounds on a given cognitive capacity of a system, as a function of the computational resources available to the system (along with other attributes of the system, which are always needed to establish non-trivial lower bounds on performance): • Direction 6 (preference learning). Inferring the preferences of a human from their words and actions requires attributing certain failures in their behavior to limitations of their cognition. Some such limitations could be derived from resource bounds on the human brain, or even better, on relevant cognitive subroutines employed by the human (if sufficient progress in cognitive science is granted to identify those subroutines). • Direction 17 (hierarchical human-in-the-loop learning (HHL)). The degree of oversight received by an AI system should be sufficient to overcome any tendency for the system to find loopholes in the judgment of an overseer(s). A precise model of how to strike this balance would benefit from the ability to predict lower bounds on the cognitive abilities of the overseer and upper bounds on the abilities of the AI 47 system being overseen, accounting for their respective computational resources. • Direction 28 (reimplementation security). Upper bounds on the collective capabilities of malicious hackers could be used to estimate whether they have sufficient resources to re-train, re-program, or otherwise compromise a powerful AI system or the security protocols surrounding it. It would be informative if such bounds could be derived from estimates of the hackers’ total computational resources. (Although this would not protect against flaws in the assumptions of the designers of the system to be protected, which are the main source of real-world security breaches.) • Direction 29 (human-compatible equilibria). Suppose some sufficiently sharp upper bounds on the collective capabilities of the non-human-agents in a multi-agent system could be predicted as a function of their computational resources. These bounds could be used to set limits on how much computation the non-human agents are allowed to wield, so as to ensure a sufficient degree of control for the humans while maintaining the usefulness of the non-human agents to the collective. • Direction 26 (capacity oversight criteria). Bounds on the capabilities of both AI systems and humans could be used to determine whether an AI system is sufficiently computationally endowed to be prepotent. This could lead to more definable standards for when and when not to worry about Type 1b risks (unrecognized prepotence). • Direction 8 (human cognitive models). Griffiths et al. (2015) have argued that computational limitations should be accounted for in human cognitive models. A better understanding of how an ideal bounded reasoner manages computation for rational decision-making could lead to better predictive and interactive models of humans, which could flow through to work on Directions 1, 4, 7, and 11 (transparency and explainability, AI-assisted deliberation, human belief inference, and deference to humans). Actionability. Most experimental work in the field of machine learning is concerned with assessing the capabilities of AI systems with limited computation. Therefore, it could be fruitful and straightforward to begin experimental approaches to each bullet point in the instrumental motivation section above. However, to bolster experimental approaches, it would help to develop a rigorous framework for planning and evaluating such experiments in advance. Currently, no satisfactory axiomatic theory of rational thinking under computational limitations—such as the hardware limitations inherent in a human brain, or any physical computer system—is known. One essential difficulty is that probability estimates calculated using bounded computational resources cannot be expected to follow the laws of probability theory, which require computation in order to satisfy (see the historical note below). For example, it can take a great deal of computation to prove that one statement is logically equivalent to another, and therefore to deduce that the statements should be assigned the same probability. Agent models which assume agents’ beliefs follow the rules of probability 48 theory—which assign equal probability to logically equivalent statements— are therefore unrealistic. Another difficulty is that it is unclear what rules the beliefs of reasoners in a multi-agent system should be assumed to satisfy, especially when the reasoners are in competition with one another. Competition means the agents may have an incentive to deceive one another; when one agent deceives another, should the deceived agent be blamed, or the deceiver, or both? On one hand the deceived agent is failing to protect itself from deception; on the other hand the deceiver is failing to uphold a basic principle of good faith communication that might be fundamental to effective group-scale interactions. Garrabrant et al. (2016) have made some effort to resolve these difficulties by developing a model of a bounded reasoner called a “logical inductor”, along with a suite of accompanying theorems showing that logical inductors satisfy a large number of desirable properties. A logical inductor’s capabilities include converging toward satisfying the laws of probability over time, making well-calibrated predictions about other computer programs including other logical inductors, the ability to introspect on its own beliefs, and self-trust. Logical inductors also avoid the fallacy of treating the outputs of deterministic computations as random events, whereas past models of bounded reasoners tend to assume the reasoner will implicitly conflate uncertainty with randomness (Halpern et al., 2014). However, the logical inductor theory as yet provides no upper bounds on a bounded reasoner’s capabilities, nor does it provide effective estimates of how much computation the reasoner will need for various tasks. Thus, progress on bounded rationality could be made by improving the Garrabrant model in these ways. Consideration of side effects. A working predictive theory of bounded rationality would eliminate the need to run any machine learning experiment whose outcome is already predicted by the theory. This would make machine learning research generally more efficient, hastening progress. The theory could also inspire the development of new and more efficient learning algorithms. It is unclear whether such advancements would reduce or increase existential risk overall. Historical note. Chapters 1 and 3 of Do the Right Thing (Russell and Wefald, 1991) contain a lengthy discussion of the challenge of treating bounded rationality axiomatically. Some excerpts: “[...] computations are treated as if they were stochastic experiments, even when their outcomes are completely deterministic. [...] Given the absence of a satisfactory axiomatic system for computationally limited agents, our results have only a heuristic basis, strictly speaking.” (p. 25) “These time-limited estimates, which Good (1977) called dynamic probabilities and utilities, cannot obey the standard axioms of probability and utility theory. Just how the axioms should be revised to allow for the limited rationality of real agents without making them vulnerable to a charge of incoherence is an important open philosophical problem, which we shall not attempt to tackle here. [...] the formulae here and in chapters 4 and 5 have as yet only a heuristic justification, borne out by practical results.” (pp. 60-61) 49 Despite this, many attempts to axiomatize bounded rationality since then, such as by Halpern and Pass (2011), continue to prescribe that the agent should model the outputs of unfinished computations using probability. 5.2 Single/single instruction 5.2.1 Direction 6: Preference learning Preference learning is the task of ensuring that an AI system can learn how to exhibit behavior in accordance with the preferences of another system, such as a human. Social analogue. When a CEO asks her employee to help increase their company’s profits, she implicitly hopes the employee will do so without conspiring to have her fired from the company in order to replace her with someone more effective, or by engaging in immoral acts like hacking a competitor’s bank account. The CEO’s preferences are thus quite a bit more complex than the statement “help us increase profits” alone might suggest. Moreover, because she cannot easily specify the innumerable things she hopes the employee will not do, the employee must exercise some independent judgment to infer the CEO’s preferences from surrounding social context. Scenario-driven motivation. Preference learning is mainly relevant to mitigating Type 1c risks (unrecognized misalignment), and requires striking a balance between literal obedience and independent judgment on the part the AI system. If a superintelligent factory management system is instructed with the natural language command, “make as many paperclips as possible this year”, one of course hopes that it will not attempt to engineer nanotechnology that fills a sphere two light-years in diameter with paperclips Bostrom (2014, Chapter 8, “Infrastructure Profusion”). At the same time, if it does not make any paperclips at all, it will tend to be replaced by another system which does. Without a satisfactory procedure for striking a balance between literal obedience and independent judgment, we humans may be unable to instate our preferences as governing principles for highly advanced AI systems. In particular, the continued existence and general well-being of human society—a highly complex variable to define—would be placed at risk. Actionability. Specifying an AI system’s objectives directly in terms of a score function of the environment to be maximized can lead to highly unpredictable behavior. For an example, programming a cleaning robot to maximize the amount of dirt it picks up could result in the robot continually spilling out dirt for itself to clean (Russell et al., 2003, Chapter 17.1). Similarly, a reinforcement learning system trained to maximize its score in a boat racing game learned to drive in circles to collect more points instead of finishing the race (Amodei and Clark, 2016). One approach to this problem is to use preference learning, i.e., to design AI systems to adjust their model of human preferences over time. Human preference learning is already an active area of research with numerous past and present applications, for example in product recommendation systems or automated software configuration. New commercial applications of 50 preference learning, such as personal assistant software, will surely become more prevalent over the coming decade. There are numerous mathematical formulations of preference learning problem; see Braziunas (2006) for a review. In a sequential decision-making setting, the problem can be expressed as a POMDP, where the human’s preferences are encoded as information about the environment determining which states are desirable (Boutilier, 2002). This formulation involves not only learning human preferences, but taking actions that satisfy them. This is the full problem of preference alignment: aligning an AI system’s behavior with the preference a user. Preference learning is further complicated in a cooperative setting, where the human is also taking actions directly toward their goal. Here, success for the AI system is defined as the combined efficacy of a human/AI team working toward a common objective that is understood primarily by the human. This setting can also been represented as a POMDP, where the human’s actions are part of the environment’s transition function (Fern and Tadepalli, 2010). The human’s actions can then be taken as evidence about their preferences, such as using inverse reinforcement learning (IRL), also known as inverse optimal control (Kalman, 1964). This approach was introduced by Javdani et al. (2015). Somewhat concurrently, Hadfield-Menell et al. (2016a) introduced cooperative inverse reinforcement learning (CIRL), a problem framing where a human and an AI system share common knowledge that the AI system is attempting to learn and optimize the human’s objective. The CIRL framing been used to explore the possibility of “pragmatic” robots that interpret human actions with an awareness that the human is attempting to teach them (Fisac et al., 2017). Using similar but slightly different assumptions from CIRL (in particular, using limited levels of metacognition on the part of the human and robot, yielding non-equilibrium strategies), Milli and Dragan (2019) show that non-pragmatic robots are more robust than pragmatic robots, even when humans are in fact trying to teach them about their preferences. In these experiments, joint performance is improved when the robot takes a literal interpretation of the human, even when the human is not attempting to be literal. There are some concerns that present-day methods of preference learning may not suffice to infer human preferences in a form sufficiently detailed to safely direct the behavior of a prepotent or near-prepotent AI system. Thus, in order to be marginally valuable for the purpose of reducing existential risk, a focus on approaches to preference learning that might scale well for directing more advanced systems (as in Tier 1 risks) may be needed. For this, heuristics for minimizing the unintended side effects of the system’s operation (Amodei et al., 2016; Krakovna et al., 2018), avoiding taking optimization to extremes (Taylor, 2016b), or taking optimization instructions too literally, also known as “reward hacking” (Amodei et al., 2016; Ibarz et al., 2018)), could be useful to codify through theory or experiment. Absent an approach to single/single delegation that would address such issues implicitly and automatically, heuristics could be helpful as transient rules of thumb to guide early AI systems, or to provide inspiration for rigorous and scalable long-term solutions to preference alignment. As well, preference learning methods that account for idiosyncrasies of human cognition may also be needed to avoid interpreting errors in judgement as preferred outcomes. For instance, Evans and Goodman (2015) explore preference learning methods accounting for bounded cognitive capacity in the humand, and (Evans et al., 2016) account for biases in the 51 human’s judgement. An alternative approach would be to ascertain how humans themselves infer and convey preferences (Baker and Tenenbaum, 2014; Lucas et al., 2014; Meltzoff, 1995), and develop AI systems to use the same methods. This approach is being investigated by Stuart Armstrong, in as-yet unpublished work. Consideration of side effects. If AI systems or human institutions use preference learning to develop a highly precise understanding of human preferences, that knowledge could be used in ways that are harmful to the humans. For instance, satisfying the short-term preferences of the humans in question could be used as part of a longer-term strategy to gain and exploit their trust in ways that they will later regret. Thus, to respect the wishes of the persons or institutions whose preferences are being learned, certain measures may be needed to ensure that preference learning capabilities are usually or always deployed within a preference alignment methodology. Historical note. The challenge of clearly specifying commands to an intelligent machine was also remarked by Norbert Wiener (Wiener, 1960); see the historical note in Section 2.2 for a direct quote. 5.2.2 Direction 7: Human belief inference An AI system that is able to infer what humans believe about the factual state of the world could be better suited to interact with humans in a number of ways. On the other hand, it might also allow the system to acquire a large amount of human knowledge by inferring what humans believe, thereby enabling prepotence. As such, this research direction is very much “dual use”. Social analogue. Suppose Alice is a doctor, and Bob is her intern. A hospital patient named Charlie has previously experienced severe allergic reactions to penicillin. One day, Charlie gets an ear infection, and Alice prescribes penicillin for the treatment. Now suppose Bob is nearby, and knows about Charlie’s allergy. What should Bob do about Alice’s decision? If Bob assumes Alice’s beliefs about the world are correct, this would mean either Alice wishes to harm Charlie, or that that Charlie is in fact no longer allergic to penicillin. However, the pragmatic thing is for Bob to infer something about Alice’s beliefs: in this case, that Alice is not aware of Charlie’s allergy. This inference will likely lead Bob to ask questions of Alice, like whether Charlie’s allergy has been accounted for in the decision. Scenario-driven motivation. See the instrumental motivations. Instrumental motivation. Progress on the theory and practice of belief inference could improve our understanding of • Direction 4 (AI-assisted deliberation). This may require AI systems to model human beliefs, implicitly or explicitly, in order to decide when and how to assist in their deliberation. • Direction 6 (preference learning). Suppose a model describing humans does not account for potential errors in a human’s beliefs when observing the human. Then, when the human fails at a task due to 52 erroneous beliefs, the model will interpret the human as wanting to the fail at the task. Hence, belief inference is important for preference inference and thereby preference learning. • Direction 11 (deference to humans). A number of protocols for AI systems deferring to humans could involve inferring the beliefs of the human. For instance, “defer to the human’s beliefs when the human is more likely to be correct than me”, or “defer to the human in situations where the human will believe I should have deferred to them”. These protocols behave very differently when the human’s beliefs are incorrect but the human wants to be deferred to anyway, say, for policy-level reasons intended to maintain human control. Nonetheless, they both take inferred human beliefs as inputs. • Direction 24 (resolving planning disagreements). Humans with differing beliefs may come into disagreements about what policy a powerful AI system should follow. An AI system that is able to infer the nature of the differing beliefs may be able to help to resolve the disagreement through dialogue. Actionability. Human beliefs should likely be inferred through a variety of channels, including both natural language and demonstrations. Bayesian methods specifically for extracting human priors (Griffiths and Kalish, 2005) have been explored to determine human priors on variables such as box office earnings and the lengths of poems (Lewandowsky et al., 2009). For learning human beliefs from demonstrations of human actions, a generalization of Inverse Reinforcement Learning (Abbeel and Ng, 2004) could be viable, such as by modeling the human as solving a POMDP. There is a small amount of quantitative evidence that humans model other agents (and presumably other humans) in this way, i.e., by assuming the other agent is solving a POMDP and figuring out what the agent’s beliefs and desires must be to explain the agent’s behavior (Baker et al., 2011). If humans indeed make use of this “POMDP inversion” method in order to model each other, perhaps AI systems could use POMDP inversion to model humans. Differentiable MDP solvers and POMDP solvers can be used for gradient descent-based approaches to maximum-likelihood estimation of the MDP or POMDP an agent believes it is solving. This would enable a learner to simultaneously infer the prior, transition rule, and reward function in the mind of a demonstrator. Empirical testing could then assess the efficacy of this approach for assessing the beliefs of humans from their demonstrations. Reddy et al. (2018) has explored this methodology in a user study with 12 human participants. Consideration of side effects. There are several major concerns about AI systems that are able to infer human beliefs. • (rapid acquisition of human knowledge) If an AI system can infer human beliefs in a usable form, it can acquire human knowledge. For instance, if an AI system is capable of reading and understanding natural language corpora, perhaps all of the knowledge of the internet could be made available to the system in an actionable form. The ability to absorb human knowledge at scale would eliminate one of the main barriers to prepotence, namely, that human society has accumulated wisdom over time that is not by default usable to a powerful AI system. Belief inference methods, especially through natural 53 language processing that could be repurposed to process natural language corpora, could therefore enable prepotence and exacerbate all Tier 1 risks (MPAI deployment events). • (deception of humans) A related issue is that any sufficiently detailed model of a human person could be used to deceive that person, by reverse-engineering what they would need to see or hear in order to become convinced of a certain belief. If an AI system is able to deceive all of human society, this could enable prepotence via social acumen, thereby exacerbating all Tier 1 risk (MPAI deployment events). Alternatively, if an AI system is already prepotent via non-social means, but only sufficiently skilled in deception that it can can deceive a small number of individuals humans, it might trick its creators into deploying it prematurely, which would also increase Type 1b and 1c risks. These issues would need to be averted somehow to ensure that the net impact of human-modeling technology is a reduction in existential risk. 5.2.3 Direction 8: Human cognitive models Models of human cognition that are representable in a mathematical or otherwise digital form could be useful for designing human/AI interaction protocols for addressing other problems in this report. On the other hand, they could also be abused to manipulate humans. This research direction, like many, is “dual use”. Social analogue. Suppose Alice is the CEO of a law firm, and Bob is her assistant. Alice has been hoping for some time that her firm would take on CharlieCorp as a client. Once day, CharlieCorp sends Alice a long email, cc’ing Bob, which ends with “... we are therefore seeking legal counsel. We assume from your past cases that you would not be interested in taking us as a client, but thought it would be a good idea to check.” Alice, having a busy week, fails to read the last line of the email, and replies only with “Thanks for the update.” Luckily, Bob realizes that Alice might have overlooked the ending, and sends her a ping to re-read it. Alice rereads and responds with “Looking at your situation, we’d actually be quite interested. Let’s set up a meeting.” Here, Bob is implicitly modeling not only Alice’s desire to work with CharlieCorp, but also Alice’s attentional mechanism. In particular, Charlie thinks Alice’s attention was not directed toward the end of the email. Later, CharlieCorp asks Bob a question about a very long document. That day, Alice’s schedule is clear, and knowing Alice is a fast reader who is familiar with the subject matter of the document, Bob forwards the question to Alice for her to think about. Here, Bob is modeling Alice’s attentional capacity, her written language comprehension, as well as the contents of her memory. Scenario-driven motivation. See the instrumental motivations. Instrumental motivation and actionability. Progress on the theory and practice of human cognitive modeling could improve our understanding of 54 • Direction 4 (AI-assisted deliberation). To the extent that AI systems may eventually be needed to assist humans in safety assessments of other AI systems, understanding the quirks and limitations of human thinking may be helpful in designing a system that helps humans to reach a sound conclusion. To this end, Ought.org (2017b) have attempted to generate datasets of examples of human deliberative output. Collecting more data of this sort could help to train and/or validate models of human cognitive functions involved in deliberation. • Direction 6 (preference learning). To infer a person’s preferences from their behavioral outputs, it would help to understand the mapping B from preferences to behavior, including speech. Then, preference inference amounts to inverting that mapping: given observed behavior b, we seek to find preferences p that would satisfy B(p) = b. Direction 7 (human belief inference) has already discussed how the person’s beliefs play a role in defining the map B. However, B is parametrized by other features of human cognition aside from beliefs and preferences, such as planning, attention, memory, natural language production, and motor functions. Isolating or at least narrowing our uncertainty about those variables could thus help us to reduce uncertainty in the “behavior equation” B(p) = b that we are solving when performing preference inference. As an example of early work in this direction, Steyvers et al. (2006) models the interaction of inference and memory. • Direction 11 (deference to humans). Suppose an AI system plans to defer to humans to take over from certain confusing situations, but those situations would either be too complex for humans to reason about, or too prone to the influence of particular human biases for humans to handle the situation responsibly. This means that even routine applications of AI technology, in situations where the AI hands off control or decision-making to a human, will likely need to account explicitly or implicitly for human cognitive peculiarities aside from preferences. Developing principled and generalizable hand-off procedures that will scale with the intelligence of the AI system may require better models of human cognition. As a simple present-day example, self-driving car technology must account for human reaction time when handing control over to a human driver (Dixit et al., 2016). • Direction 24 (resolving planning disagreements). Disagreements between humans might sometimes be due to different tendencies in more basic cognitive functions like attention and memory. For example, if Alice has a great memory and Bob has a terrible memory, Alice might disagree with Charlie on the nature of their unrecorded verbal agreements, and Bob—if he knows he has a bad memory—might not trust Alice to be the arbitrator of those disagreements. Thus, an AI system that offers compromises that humans are likely to accept may need a working model of humans’ cognitive capacities aside from their preferences. Identifying and explaining these differences could be helpful in dispute resolutions, and hence in facilitating agreements to continue sharing ownership of powerful AI systems. For example, Taber and Lodge (2006) shows that political disagreements arise to some extent from motivated skepticism, and Griffiths et al. (2008) show that cultural disagreements should be expected to arise from inherited inductive biases. Such nuances may also prove essential in Direction 22 55 (modeling human committee deliberation). Consideration of side effects. There are a number of potentially dangerous and wide-reaching side effects to developing high-fidelity human cognitive models. • Manipulation of humans. Human cognitive models can be used to manipulate humans. This can already be seen in social media platforms that develop user models to generate addictive features to keep users engaged. If sufficiently detailed, perhaps human cognitive models could be used by an AI system to manipulate all of human society in a goal-directed fashion. In principle this could enable prepotence through social acumen, thereby exacerbating all Tier 1 risks (MPAI deployment events). • Impoverished third-party safety testing. If detailed human models are made publicly available, we impoverish our ability to perform “hold-out” safety testing and verification for powerful AI systems, as in Direction 3 (formal verification for machine learning systems). Specifically, if precise human models are not made publicly available, and instead withheld by a independent AI safety testing institution, then the models could be used to design simulation-based safety tests as a regulatory safety check for AI systems built by private corporations or the public. However, if the human models used in the safety tests were released, or derivable by institutions other than the safety testers, then the models could be used by corporations or individuals deploying AI systems to “game” the regulatory testing process (Taylor, 2016c), the way a student who knows what questions will be on exam doesn’t need to learn the rest of the course material. In particular, this could lead to an increase in Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment). Thus, a judicious awareness of how and when to apply human-modeling technology will be needed to ensure it is shared appropriately and applied beneficially. See also Direction 7 (human belief inference) for a consideration of side effects of modeling human beliefs specifically. 5.3 Single/single control 5.3.1 Direction 9: Generalizable shutdown and handoff methods As with any machine, it remains important to maintain safe shutdown procedures for an AI system in case the system begins to malfunction. One might operationalize “shutdown” as the system “no longer exerting control over the environment”. However, in many situations, ceasing to apply controls entirely may be extremely unsafe for humans, for example if the system is controlling a self-driving car or an aircraft. In general, the sort of shutdown procedure we humans want for an AI system is one that safely hands off control of the situation to humans, or other AI systems. Hence, the notion of a handoff can be seen as generalizing that of a shutdown procedure. In aviation, the term “handoff” can refer to the transfer of control or surveillance of an aircraft from one control center to another, and in medicine the term is used similarly for a transfer of responsibilities from one doctor to another. This research direction is concerned with the development of generalizable shutdown and handoff techniques for AI systems. 56 Social analogue. Suppose AliceCorp hires Betty to take on some mission-critical responsibilities. In case Betty ever becomes ill or uncooperative and can no longer perform the job, other employees must be ready to cover off Betty’s responsibilities until a replacement can be found. Such handoffs of responsibility can be quite difficult to coordinate, especially if Betty’s departure is a surprise. For instance, any documented instructions for performing Betty’s responsibilities may need to be documented in a manner that is readable to other employees, given their more limited context and perhaps experience. Therefore, many companies will go to great lengths to maintain detailed documentation of responsibilities and handoff procedures. Similar procedures are often needed but missing on the scale of industries: when certain companies become “too big to fail”, governments are left with no means of replacing them with better versions when they begin to malfunction. Scenario-driven motivation. Generalizable shutdown and/or handoff procedures could reduce the risk of Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment) by making it easier for humans to regain control of a situation where an AI system is malfunctioning or behaving drastically. In general, future applications of powerful AI systems may pose risks to society that cannot be simulated in a laboratory setting. For such applications to be responsible, general principles of safe shutdown and safe handoff procedures may need to be developed which are known in advance to robustly generalize to the high-stakes application. Somewhat orthogonally, perhaps the involvement of many humans in training and/or drills for AI→human handoffs could create a source of economic involvement for humans to reduce Type 2b risk (economic displacement of humans), and/or cognitive stimulation for humans to reduce Type 2c risk (human enfeeblement). Actionability. Practically speaking, almost any existing computer hardware or software tool has a custom-designed shutdown procedure, including AI systems. However, there has not been much technical work on generalizable strategies for shutting down or handing over control from an AI system. In human–robot interaction literature, there is a body of existing work on safe handovers, typically referring to the handoff of physical objects from robots to humans. For instance, Strabala et al. (2013), have studied both robot-to-human and human-to-robot handovers for a variety of tasks. Moon et al. (2014) showed that using humanlike gaze cues during human-robot handovers can improve the timing and perceived quality of the handover event. For self-driving cars, Russell et al. (2016) show that human motor learning affects car-to-driver handovers. For unmanned aerial vehicles, Hobbs (2010) argue that “the further development of unmanned aviation may be limited more by clumsy human–system integration than by technological hurdles.” Each of these works contains reviews of further relevant literature. For coordination with multiple humans, Scerri et al. (2002b) put forward a fairly general concept called transfer of control for an AI system coordinating with multiple humans, which was tested in a meeting-planning system called Electric Elves (E-Elves). The E-Elves system was used to assist in scheduling meetings, ordering meals, and finding presenters, over a 6-month period by a group of researchers at the University of Southern California. Scerri et al. describes the mathematical model underlying 57 the system, which used an MDP formulation of the human/AI interaction problem to express coordination strategies and assess their expected utility in terms of “the likely relative quality of different entities’ decisions; the probability of getting a response from an entity at a particular time; the cost of delaying a decision; and the costs and benefits of changing coordination constraints”. Perhaps similar general principles could be used to design shutdown and/or handover processes in other settings. In any task environment, one might try to operationalize a safe shutdown as “entering a state from which a human controller can proceed safely”. As a cheaper proxy to use in place of a human controller in early prototyping, another AI system, or perhaps a diversity of other AI systems, could be used as a stand-in during training. Suites of reinforcement learning environments such as OpenAI Gym (Brockman et al., 2016) could be used to ascertain the generality of any given safe handover technique. Consideration of side effects. As with any safety methodology, if safe handover methods are developed for near-term systems and erroneously presumed to generalize to more powerful systems, they could create a false sense of security. For instance, suppose generalizable solutions are developed for handing off control from a single AI system to a single human, such as from a self-driving car to a human driver. The same principles might not work to hand off control from an automated air traffic control system to human air traffic controllers, which might require solving a coordination problem between the humans who receive the control in the event of a shutdown. Or, a simple “suspend activity and power down” procedure might be used to shut down many simple AI systems, but then someday fail to effectively shut down a powerful misaligned system that can build and execute copies of itself prior to the shutdown event. Thus, to apply ideas from this research direction responsibly, one must remain on the lookout for unique challenges that more complex or capable AI systems will present. Historical note. Wiener has also remarked on the difficulty of interfering with a machine which operates on a much faster time scale than a human. “We have seen that one of the chief causes of the danger of disastrous consequences in the use of the learning machine is that man and machine operate on two distinct time scales, so that the machine is much faster than man and the two do not gear together without serious difficulties. Problems of the same sort arise whenever two operators on very different time scales act together, irrespective of which system is the faster and which system is the slower.” (Wiener, 1960) 5.3.2 Direction 10: Corrigibility An AI system is said to be corrigible if it “cooperates with what its creators regard as a corrective intervention, despite default incentives for rational agents to resist attempts to shut them down or modify their preferences” (Soares et al., 2015). In particular, when safe shutdown procedures are already designed and ready to execute, a corrigible AI system will not work against its human operator(s) to prevent being shut down. Social analogue. A person is said to be “corrigible” if they are capable of being corrected, rectified, or reformed. An “incorrigible” person is one 58 who does not adjust their behavior in response to criticism. If an employee behaves in an incorrigible manner, an employer may rely on the ability to terminate the employee’s contract to protect the company. Imagine, however, an incorrigible employee who is sufficiently crafty as to prevent attempts to fire them, perhaps by applying legal technicalities or engaging in manipulative social behaviors. Such a person can cause a great deal of trouble for a company that hires them. Scenario-driven motivation. As AI systems are developed that are increasingly capable of social intelligence, it becomes increasingly important to ensure that those systems are corrigible. An incorrigible AI system whose goals or goal inference instructions are mis-specified at the time of its initial deployment poses a Type 1c risk (unrecognized misalignment) to humans if it is able to prevent us from modifying or disabling it. Actionability. Hadfield-Menell et al. (2016b) have shown that a reinforcement learning system can be given uncertainty about its reward function in such a way that human attempts to shut it down will tend to cause it to believe that being shut down is necessary for its goal. This is not a full solution to corrigibility, however. Carey (2017) shows that incorrigibility may still arise if the AI system’s uncertainty about the reward function is not appropriately specified. Moreover, Milli et al. (2017) point out that too much reward uncertainty can lead an AI system to underperform, so there is a balance to be struck between expected performance and confidence that shut-down will be possible. As a potential next step for resolving these issues, experiments could test other mechanisms aside from reward uncertainty for improving corrigibility. For example, see Direction 20 (self-indication uncertainty) below. A different approach to corrigibility for reward-based agents is to somehow modify their beliefs or reward function to make them more amenable to shutdown or modification. Armstrong and O’Rourke (2017) provides an overview of attempts in this direction. Consideration of side effects. Progress on the problem of corrigibility does not seem to present many negative side effects, other than the usual risk of falsely assuming that any given solution would generalize to a highstakes application without sufficient testing. 5.3.3 Direction 11: Deference to humans Deference refers to the property of an AI system actively deferring to humans on certain decisions, possibly even when the AI system believes it has a better understanding of what is right or what humans will later prefer. Social analogue. Suppose Allan is a patient and Betty is his doctor. Allan is bed-ridden but otherwise alert, and Dr. Betty is confident that Allan should receive a dose of anesthetic to help Allan sleep. Suppose also that the Dr. Betty is bound by law to ask for the patient’s consent before administering this particular anesthetic, and that she expects the patient to say “no”. Even if Dr. Betty is very confident that she knows what’s best for the patient, the doctor is expected to defer to the patient’s judgment in this case, rather than, say, administering the anesthetic in secret along with the patient’s other medications. That is, the doctor is sometimes required 59 to defer to the patient, even when confident that the patient will make the wrong choice. Instrumental motivation. Theoretical models and/or training procedures for deference to humans could help directly with • Direction 10 (corrigibility). In order to preserve the corrigibility of an AI system over time, we will need AI systems to not only respond to corrective interventions, but to seek them out as a matter of policy, particularly on decisions that could lead to a loss of corrigibility. • Direction 17 (hierarchical human-in-the-loop learning (HHL)). A generic deference capability may allow AI systems to serve as useful delegates in a chain of command including humans and other AI systems. • Direction 29 (human-compatible equilibria). A notion of deference to humans that is stable as AI systems evolve and replicate over time might constitute an important class of Direction 29 (humancompatible equilibria). Actionability. Simulated experiments where one AI system is required to seek out and defer judgment to another AI system could be fruitful for developing and testing protocols for deferring to outside judgment. Milli et al. (2017) show that performance trade-offs are to be expected when requiring direct obedience to commands. Experiments to ascertain an appropriate balance between deference and autonomy for minimizing tail risks arising from system mis-specification could be highly informative. Consideration of side effects. Too much deference to humans could lead to catastrophic errors. For instance, if a powerful AI system responsible for managing the electrical grid of a city were to defer to a single human on the decision to shut it down, perhaps many people could suffer or die as a result. In the future, perhaps larger systemic failures of this sort could present existential risks. 5.3.4 Direction 12: Generative models of open-source equilibria AI systems are in principle completely inspectable to humans, in that their execution can create a perfect log of every internal state that occurs. The degree to which the internal “thought processes” of such machines will be understandable to humans will likely depend on the success of future research on Direction 1 (transparency and explainability). Whatever degree of transparency and/or explainability can be achieved, its implications of the game-theoretic relationship between systems and humans should be explored. But, so far, very little game theory research has been carried out to ascertain, either analytically or by simulation, what equilibria arise between agents when one agent is assumed to be partially or fully transparent to another. Social analogue. Suppose Alice is very good at reading Bob’s body language, such that if Bob tries to deceive her or make plans that she would dislike, Alice will notice. His thoughts, in addition to his outward actions, have a direct impact on his interactions with Alice. Thus, Bob has an incentive to think differently than he would if he were less transparent to Alice. 60 This changes the space of actions Bob can take, because actions that would require planning will produce side effects in Alice’s awareness. For example, if Bob begins to formulate a plan to deceive Alice, she might notice and try to shut him down and/or simply see through the deception. Similarly, imagine two nations which have a large number of spies investigating one another. If Nation A begins to plan a trade embargo against Nation B, spies may leak this information to Nation B and trigger early responses from Nation B prior to Nation A’s instatement of the embargo. The early response could range from submissive behavior (say, conceding to Nation A’s expected demands) to preemptive counter-embargoes, depending on the situation. Scenario-driven motivation. Could a powerful AI system someday learn or infer how to deceive its own developers? If possible, it could constitute a Type 1b or 1c risk (unrecognized prepotence or unrecognized misalignment). If not possible, it would be reassuring to have a definite answer as to why. This is a question for “open source game theory”, the analysis of interactions between decision-making entities that are partially or fully transparent to one another. More broadly, deception is only one important feature of a human/AI equilibrium in which mutual transparency of the human and the AI system could play a key role. Another might be intimidation or corruption: is it possible for the mere existence of a particular powerful AI system—in a partially or fully transparent form—to intimidate or corrupt its creators to modify or deploy it in ways that are harmful to the public? In a diffuse sense, this might already be happening: consider how the existence of social media platforms create an ongoing incentive for their developers to make incremental updates to increase user engagement. While profitable for the company, these updates and resulting increases in engagement might not be beneficial to the overall well-being of individual users or society. To understand the dynamics of these mutually transparent relationships between humans and AI systems, it might help to begin by analyzing the simplest case of a single human stakeholder interacting with a single relatively transparent AI system, and asking what equilibrium (long-run) behaviors are possible to arise. Instrumental motivation. Generative models of machine learning agents reaching equilibria in open-source games could be helpful toward understanding • Direction 17 (hierarchical human-in-the-loop learning (HHL)). In scenarios where one AI system is tasked with assisting in the oversight of other AI systems, it might make sense for the overseer system to be given access to the sources codes or other specifications of the systems being overseen. By contrast, classical game theory assumes that players are capable of private thoughts which determine their actions. Hence, the relationship between an AI system and a system overseeing its source code is outside the assumptions of classical game theory. • Direction 29 (human-compatible equilibria). An AI system’s source code will likely be visible to the humans who engineered it, who will likely use that code to run simulations or other analyses of the system. This relationship is also outside the assumptions of classical game theory. 61 Actionability. Halpern and Pass (2013) have already remarked that “translucency” rather than opacity is a more realistic assumption when modeling the interaction of human institutions, or humans who can read one another’s body language. Moreover, remarkably different equilibrium behavior is possible when agents can read one another’s source code. Tennenholtz (2004) developed the notion of program equilibrium for a pair of programs playing a game which, when given access to one another’s source code, have no positive incentive to be replaced or self-modified. Strikingly, it turns out that open-source agents can achieve certain cooperative (or defective) equilibria that are in principle not possible for closed-source agents (Critch, 2019). Understanding whether and how such equilibria could arise amongst advanced AI systems (and how various design choices might affect these outcomes), or between AI systems and humans, is an important question for understanding how multi-agent AI systems will equilibrate with humans. Consideration of side effects. This direction could be problematic from an existential risk perspective if models of open-source equilibria are later used to preferentially develop AI/AI/AI coordination methods in the absence of human/AI coordination methods or multi-human multi-AI coordination methods. Such methods could lead to Type 2b and 2c risks (economic displacement of humans and human enfeeblement) and/or Type 2c risk (human enfeeblement) if they result in too much human exclusion from economically productive work. 6 Single/multi delegation research This section is concerned with delegation from a single human stakeholder to multiple operationally separated AI systems (defined below). As powerful AI systems proliferate, to diminish Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment), it might help to have ways of predicting and overseeing their collective behavior to ensure it remains controllable and aligned with human interests. Even if serving a single human or human institution, coordination failures between large numbers of interacting machines could yield dangerous side effects for humans, e.g., pollutive waste, or excessive consumption of energy or other resources. These could constitute Type 1c risks (unrecognized misalignment). Conversely, unexpectedly well-coordinated interactions among multiple AI systems could constitute Type 1b risk (unrecognized prepotence), for instance, if a number of cooperating AI systems turned out to be capable of collective bargaining with states or powerful corporations. To begin thinking clearly about such questions, we must first decide what to count as “multiple AI systems” versus only a single AI system: Operational separation. Roughly speaking, for the purposes of this report, when we say “multiple AI systems” we are referring to a collection of AI-based algorithms being executed on physically or virtually separated computational substrate units, with each unit having a relatively high-bandwidth internal integration between its sensors, processors, and actuators, but only relatively low-bandwidth connections to other units. We say that such units are operationally separated. It might be tempting to simplify the number of concepts at play by viewing the collective functioning of operationally separate units as a single 62 “agent” to be aligned with the human stakeholder. However, this perspective would elide the mathematical and computational challenges involved in balancing the autonomy of the individual units against the overall functioning of the group, as well as the non-trivial task of dividing up responsibilities between the units. Dec-POMDPs. The concept of a Decentralized Partially Observable Markov Decision Process, or Dec-POMDP (Oliehoek et al., 2016), is a useful formalism for describing the problem faced by multiple AI systems (i.e., multiple operationally separated units) working to serve a common purpose. Variants of Dec-POMDPs can also be considered, such as by adding uncertainty to the reward function or transition dynamics, or more refined assumptions on computational limitations. 6.1 Single/multi comprehension If companies and governments deploy “fleets” of AI systems to serve specific objectives—be they in physical or virtual environments—humans will likely seek to understand their collective behavior in terms of the individual units and their relationships to one another. From one perspective, a fleet of AI systems might be viewed as “just a set of parallel processing units.” But, when the systems are engaged in interactive intelligent decision-making based on objective-driven modeling and planning, new tools and abstractions may be needed to organize our understanding of their aggregate impact. This section is concerned with research to develop such tools and abstractions. Single/multi delegation seems poised to become increasingly relevant. Modern computer systems, and machine learning systems in particular, already make increasing use of parallel computation. This is in part because the speed of individual processors has started to encounter physical limits, even though the cost of a FLOP has continued to decline rapidly. However, there are also increasingly relevant physical limits to communication bandwidth between processes; thus future large-scale computer systems will almost certainly employ a high degree of operational separation at some scale of organization. 6.1.1 Direction 13: Rigorous coordination models The Von Neumann-Morgenstern utility theorem and resulting utility theory (Morgenstern and Von Neumann, 1953; Von Neumann and Morgenstern, 2007) provides a principled framework for interpreting the actions of a single agent: optimizing an expected value function conditioned on a belief distribution over the state of the world. Can an analogous theory be developed for a cooperative multi-agent system to serve a single goal or objective? In addition to utilities and beliefs, the model should also include mathematical representations of at least two other concepts: • Communications: packets of information exchanged between the agents. These could be modeled as “actions”, but since communications are often designed specifically to directly affect only the internal processes of the agents communicating, they should likely receive special treatment. • Norms: constraints or objective functions for the policies of individual agents, which serve to maintain the overall functioning of the group rather than the unilateral contributions of its members. 63 Social analogue. Humans, of course, communicate. And our reliance upon norms is evident from the adage, “The ends do not justify the means”. An individual person is not generally expected to take actions at all costs to unilaterally optimize for a given objective, even when the person believes the objective to serve “the greater good”. Instead, a person is expected to act in accordance with laws, customs, and innate respect for others, which ideally leads to improved group-scale performance. Scenario-driven motivation. If there is any hope of proving rigorous theorems regarding the collective safety of multi-agent systems, precise and accurate mathematical definitions for their components and interaction protocols will be needed. In particular, theorems showing that a collective of AI systems is or is not likely to become prepotent or misaligned will require such models. Hence, this direction applies to the reduction of Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment). Moreover, common knowledge of problems and solutions in this area may be necessary to motivate coordination to reduce Type 1a risks (uncoordinated MPAI development), or to avoid dangerous interactions with powerful AI systems that would yield Type 1d risk (involuntary MPAI deployment). Actionability. The framework of Dec-POMDPs introduced by Bernstein et al. (2002) provides a ready-made framework for evaluating any architecture for decentralized pursuit of an objective; see Oliehoek et al. (2016) for an overview. As such, to begin proving a theorem to support the use of any given coordination protocol, one could start by stating conjectures using the language of Dec-POMDPs. Protocols could be tested empirically against existing machine learning methods for solving Dec-POMDPs. In fact, any given Dec-POMDP can be framed as two distinct machine learning problems: • Centralized training for decentralized execution. This is the problem of producing—using a centralized training and/or learning system— a suite of decentralized “agents” (sensor/actuator units) that collectively pursue a common objective. As examples of recent work in this area: – Sukhbaatar et al. (2016) treat a system of decentralized agents undergoing centralized training as a single large feed-forward network with connectivity constraints representing bandwidthlimited communication channels. The authors find that on four diverse tasks, their model outperforms variants they developed with no communication, full-bandwidth communication (i.e., a fully connected network), and models using discrete communication. – Foerster et al. (2016) propose two approaches to centralized learning of communication protocols for decentralized execution tasks. The first, Reinforced Inter-Agent Learning (RIAL), has each agent learn its communication policy through independent deep Q-learning. The second, Differentiable Inter-Agent Learning (DIAL), allows the training system to propagate error derivatives through noisy communication channels between the agents, which are replaced by discrete (lower bandwidth) communication channels during execution. 64 – Foerster et al. (2017) explore, in a collaborative multi-agent setting with no communication at execution time, two methods for making use of experience replay (the re-use of past experiences to to update a current policy). Each method aims to prevent the learners from confusing the distant-past behavior of its collaborators with their more recent behavior. The first method treats replay memories as off-environment data (Ciosek and Whiteson, 2017). The second method augments past memories with a “fingerprint”: an ordered tuple comprising the iteration number and exploration rate, to help distinguish where in the training history the experience occurred. • Decentralized training for decentralized execution. This is the problem of a decentralized set of learners arriving at a collective behavior that effectively pursues a common objective. As examples of recent related work: – Matignon et al. (2012) identify five qualitatively distinct coordination challenges—faced by independent reinforcement learners pursuing a common (cooperative) objective—which they call “Pareto-selection”, “nonstationarity”, “stochasticity”, “alter-exploration” and “shadowed equilibria”. – Tampuu et al. (2017) examine decentralized Q-learners learning to play variants of Pong from raw visual data, including a cooperative variant where both players are penalized equally when the ball is dropped. The variety of problems and methods in recent literature for training collaborative agents shows that no single architecture has been identified as universally effective, and far from it. None of the above works is accompanied by a rigorous theoretical model of how coordination ought to work in order to be maximally or even sufficiently effective. Hence the motivation for more rigorous foundations: to triage the many potential approaches to learning for single/multi delegation. Consideration of side effects. In order for research enabling multiagent coordination to eventually lead to a decrease rather than an increase in existential risk, it will need to be applied in a manner that avoids runaway coordination schemes between AI systems that would constitute a Type 1a, 1b, 1c, or 1d risk (uncoordinated MPAI development, unrecognized prepotence, unrecognized misalignment, or involuntary MPAI deployment). In particular, coordination-learning protocols compatible with a human being serving as one of the coordinating agents may be considerably safer in the long run than schemes that exclude humans. Present methods do not seem particularly suitable for explicitly including humans in the mix. 6.1.2 Direction 14: Interpretable machine language Just as today we seek more enlightening explanations for the actions of a neural network in order to improve our ability to evaluate and predict its behavior, in the not-too-distant future we will likely find ourselves seeking to understand the content of communications between AI systems. 65 Social analogue. Business regulations that generate legible, auditable communications within and between companies increase the difficulty for those companies to engage in corrupt business practices. This effect is of course only partial: despite the significant benefits of auditing requirements, it is usually still possible to find ways of abusing and/or circumventing legitimate communication channels for illegitimate means. Scenario-driven motivation. As we humans delegate more of our decisions to AI systems, we will likely require those systems to communicate with each other to achieve shared goals. Just as transparency for an individual AI system’s cognition benefits our ability to debug and avoid systematic and random errors, so too will the ability to interpret communications between distinct decision-making units. This benefit will likely continue to scale as the scope and number of AI systems grows. For AI capabilities approaching prepotence, interpretability of communications between AI systems may be needed to avoid Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment). The more broadly understandable the interpreted communications are made, the better developer coordination can be enabled to diminish Type 1a risk (uncoordinated MPAI development). Since interpretable communications are more easily monitored and regulated, interpretable communication standards may also be helpful for regulating communicative interactions with powerful deployed AI systems, including communications that could precipitate Type 1d risk (involuntary MPAI deployment). Actionability. As techniques develop for machine learning transparency and interpretability, similar techniques may be adaptable to ensure the interpretability of machine–machine communications in multi-agent settings; see Direction 1. Or, there may arise entirely novel approaches. Bordes et al. (2016) explore the use of end-to-end trained dialog systems for issuing and receiving API calls, as a test case for goal-oriented dialogue. In this setting, one could consider a dialogue between two machines, Machine A and Machine B, where A treats B as a machine+human system in which the human on rare occasions attempts to understand messages from A to B and penalizes the system heavily if they are not understandable. As an alternative or complement to sparse human feedback, perhaps machine–machine language could be constrained or regularized to be similar to human language, as in Lewis et al. (2017). Or, perhaps frequent automated feedback on the understandability of the A/B communication channel could be provided by a dialog state-tracking system (DSTS). As DSTS normally attempts to understand human dialogue (Henderson et al., 2014), but perhaps one could be repurposed to give automated feedback on whether it can understand the communication between A and B. Consideration of side effects. Any attempt to design or select for interpretability could lead to accidentally deceiving humans if one optimizes too much for human satisfaction with the communications rather than the accuracy of the human’s understanding. A particular concern is “steganography”, where information is “hidden in plain sight” in a way that is invisible to humans; demonstrate steganography in cycleGANs (). 66 6.1.3 Direction 15: Relationship taxonomy and detection In any attempt to train a multi-agent system to perform useful tasks like delivery services and waste collection, it is already clear that our choice of training mechanism will tend to affect whether the individual agents end up exhibiting cooperative or competitive relationships with one another. Aside from “cooperative” and “competitive”, what other descriptors of relationships between agents in a multi-agent system can be quantified that would allow us to better understand, predict, and perhaps improve upon the system’s behavior? Social analogue. Alice and Bob work together on a team whose responsibility is to send out a newsletter every week. Alice always asks to see the newsletter before Bob sends it out. Bob has expressed that he thinks Alice’s review is an unnecessary step, however, Alice continues to advocate for her review step. Are Alice and Bob in a competitive or cooperative relationship here? The answer could be somewhat complex. Perhaps Alice and Bob both really have the newsletter’s best interests at heart, and know this about each other, but Alice just doesn’t trust Bob’s judgment about the newsletters. Or, perhaps she doubts his loyalty to their company, or the newsletter project specifically. Perhaps even more complicatedly, she might trust Bob’s judgment about the content entirely, but prefer to keep the reviews in place to ensure that others know for sure that the newsletter has her approval. This scenario illustrates just a few ways in which disagreements in working relationships can arise from a variety of different relationships between beliefs and values, that do not always involve having different values. Scenario-driven motivation. To avert Type 1b and 1c risks (unrecognized prepotence and unrecognized misalignment), any single institution deploying multiple powerful AI systems into the real world will need to have a sufficient understanding of the relationships that would arise between those systems to be confident their aggregate behavior would never constitute an MPAI. To avoid Type 1a and 1d risks (uncoordinated MPAI development and involuntary MPAI deployment), development teams will collectively need to maintain an adequate awareness of the potential interactions between their own AI systems and AI systems deployed by other teams and stakeholders. For instance, consider the possibility of a war between AI systems yielding an unsurvivable environment for humanity. • If the warring AI systems were developed by warring development teams, the aggregate AI system comprising the interaction between the warring systems would be an MPAI. This would constitute a Type 1a risk (uncoordinated MPAI development), or a Type 1e risk (voluntary MPAI deployment) if one of the teams recognized that their involvement in the war would make it unsurvivable. Such cases could perhaps be made less likely by other “peacekeeping” AI systems detecting the violent relationship between the conflicting systems, and somehow enforcing peace between them to prevent them from becoming an MPAI in aggregate. • If the war or its intensity was unexpected or unintended by the developers of the AI technology used in the war, it could constitute a 67 Type 1b, 1c, or 1d risk (unrecognized prepotence, unrecognized misalignment, or involuntary MPAI deployment). Such cases could perhaps be made less likely by detecting and notifying developers when violent relationships are arising between the systems they develop and deploy, and allowing developers to recall systems on the basis of violent usage. On the other hand, an unexpected coalition of AI systems could also yield a runaway loss of power for humanity. If the coalition formation was expected by everyone, but human institutions failed to work together to stop it, then it would constitute a 1a or 1d. Developing a methodology for identifying and analyzing relationships between AI systems might be among the first steps to understanding and preventing these eventual possibilities. Crucially, there may be many more complex relationships between powerful AI systems that we humans would struggle to define in terms of simple war or peace, furthering the need for a systematic study of machine relationships. In any case, both positive and negative results in research on relationship taxonomy and detection could be beneficial to making negative outcomes less likely: • Benefits of negative results. If the relationships between nearprepotent AI systems begin to appear too complex to arrange in a manner that is legibly safe for humanity, then researchers aware of this issue can advise strongly for policies to develop at most one very powerful AI system to serve human civilization (or no such system at all, if multi/single delegation also proves too difficult). In other words, advanced warning of unsurmountable difficulties in this research area might help to avoid heading down a so-called “multi-polar” development path for powerful AI technologies. • Benefits of positive results. If the relationships between nearprepotent AI systems appear manageable, perhaps such systems could be used to keep one another in check for the safety of humanity. In other words, positive results in this area might help to optimize a “multi-polar” development pathway to be safer on a global scale. Actionability. One approach to this research area is to continually examine social dilemmas through the lens of whatever is the leading AI development paradigm in a given year or decade, and attempt to classify interesting behaviors as they emerge. This approach might be viewed as analogous to developing “transparency for multi-agent systems”: first develop interesting multi-agent systems, and then try to understand them. At present, this approach means examining the interactions of deep learning systems. For instance, Leibo et al. (2017) examine how deep RL systems interact in two-player sequential social dilemmas, and Foerster et al. (2018) explore the consequences of agents accounting for one another’s learning processes when they update their strategies, also in two-player games. Mordatch and Abbeel (2018) examine the emergence of rudimentary languages from a centralized multi-agent training process, giving rise to a variety of interactive behaviors among the agents. Consideration of side effects. This sort of “build first, understand later” approach will become increasingly unsatisfying and unsafe as AI technology improves, especially if AI capabilities ever approach prepotence. As remarked by Bansal et al. (2017), “a competitive multi-agent environment 68 trained with self-play can produce behaviors that are far more complex than the environment itself.” As such, it would be useful to develop a methodology for relationship taxonomy and detection that not only makes sense for current systems but will generalize to new machine learning paradigms in the future. For this, a first-principles approach rooted in the language of game theory and/or economics may be necessary as a complement to empirical work. 6.1.4 Direction 16: Interpretable hierarchical reporting This research direction is concerned with arranging hierarchies of AI systems that report to one another and to humans in a manner that resembles a present-day human business, and that would be legible to human overseers. Hierarchy is a natural solution to the problem of “scalable oversight” (Amodei et al., 2016) for teams of AI systems and/or humans, because hierarchies often lead to exponential gains in efficiency by reducing the complexity of problems and systems to smaller parts. In a hierarchical reporting paradigm, AI systems could be developed for the express purpose of “middle management”, to provide intelligible reports and questions either directly to humans, or other AI systems. By involving human overseers at more levels of the hierarchy, perhaps a greater degree of interpretability for the aggregate system can be maintained. Social analogue. Imagine the CEO of a large corporation with thousands of employees. The CEO is responsible for making strategic decisions that steer the company towards desirable outcomes, but does not have the time or expert technical knowledge to manage all employees and operations directly. Instead, she meets with a relatively small number of managers, who provide her with summarized reports on the company’s activities that are intelligible to the CEO’s current level of understanding, with additional details available upon her request, and a limited number of questions deferred directly to her judgment. In turn, each manager goes on to review other employees in a similar fashion. This reporting structure is enriched by the ability of the CEO to ask questions about reports from further down in the “chain of command”. Scenario-driven motivation. Consider a world in which autonomous, nearly-prepotent AI systems have become capable of interacting to generate a large number of business transactions that generate short-term wealth for their users and/or trade partners. Who or what entity can oversee the net impact of these transactions to avoid negative externalities in the form of catastrophic risks, e.g., from pollution or runaway resource consumption? Historically, human governments have been responsible for overseeing and regulating the aggregate effects of the industries they enable, and have benefited from human-to-human communications as a source of inspectable documentation for business interactions. If no similar report-generation process is developed for AI systems, human businesses and governments will face a choice: either to stifle the local economic gains obtainable from autonomous business transactions in favor of demanding more human involvement to generate reports, or to accept the risk of long-term loss of control in favor of the short-term benefits of more autonomy for the AI systems. If and when any nation or corporation would choose the latter, the result could be: 69 • An increase in Type 1c, 1b, and 1d risks (unrecognized misalignment, unrecognized prepotence, and involuntary MPAI deployment) due to the inability of the companies releasing AI systems to monitor their potential prepotence or misalignment through reporting mechanisms, and • An increase in Type 1a risks (uncoordinated MPAI development) due to the inability of human authorities such as governments and professional organizations to recognize and avert decentralized development activities that could pose a risk to humanity in aggregate. Thus, it would makes sense to find some way of eliminating the pressure to choose low-oversight regulatory regimes and business strategies, by making high-oversight strategies cheaper and more effective. Hierarchical reporting schemes would take advantage of exponential growth of the amount of supervision carried out as a function of the depth of the hierarchy, and may become a key component to scaling up supervisory measures in a cost-effective manner. One potential approach to this problem would be to deploy AI systems in “middle management” roles that curate reports for human consumption. One can imagine chains of command between submodules that oversee one another for safety, ethics, and alignment with human interests. Just as communication between employees within a company can be made to produce a paper trail that helps to some degree with keeping the company aligned with governing authorities, perhaps teams of AI systems could be required to keep records of their communications that would make their decision-making process more inspectable by, and therefore more accountable to, human overseers. Such an approach could serve to mitigate Tier 1 risks (MPAI deployment events) in full generality. Actionability. The interpretability aspect of this research direction would benefit directly from work on Directions 1 (transparency and explainability). The concept of hierarchical learning and planning is neither new nor neglected in reinforcement learning (Dayan and Hinton, 1993; Kaelbling, 1993; Wiering and Schmidhuber, 1997; Sutton et al., 1999; Dietterich, 2000; Kulkarni et al., 2016; Vezhnevets et al., 2016; Bacon et al., 2017; Tessler et al., 2017). The conception of different levels of the planning hierarchy as separate agents is also familiar (Parr and Russell, 1998). By viewing levels of hierarchical planning as separate learning agents, one can ask how to improve the transparency or interpretability of the subagents to the superagents, along the lines of Direction 1 (transparency and explainability). Ideally, the “reports” passed from subagents to superagents would be human-readable as well, as in Direction 14 (interpretable machine language). Hence, work on building interpretable hierarchical reporting structures could begin by combining ideas from these earlier research directions, subject to the constraint of maintaining and ideally improving task performance. For instance, one might first experiment with unsupervised learning to determine which ‘report features’ should be passed from a sub-agent to a superagent, in the manner learned by the agents in Mordatch and Abbeel (2018). One could then attempt to impose the constraint that the reports be human-interpretable, through a combination of real human feedback and artificial regularization from natural language parsers, although as discussed in Direction 1 (transparency and explainability), it is unclear how to ensure such reports would reflect reality, as opposed to simply offering “rationalizations”. 70 Consideration of side effects. If the humans involved in interpreting the system were insufficiently concerned with the safety of the public, they might be insufficiently vigilant to avert catastrophic risk from rare or unprecedented events. Or, if the humans individually cared about catastrophic risks, but were for some reason uncomfortable with discussing or reporting the potential for rare or unprecedented catastrophes, their individual concerns would not be enough to impact the collective judgment of the system. Hence, Type 2d risk (ESAI discourse impairment) might undermine some of the usefulness of this research direction specifically for existential risk reduction. Finally, if the resulting systems were interpretable to humans, but the institutions deploying the systems chose not to involve enough humans in the actual task of interpreting the systems (say, to operate more quickly, or to avoid accountability), then advancements in this area would accrue mostly to the capabilities of the resulting systems rather than their safety. 6.2 Single/multi instruction This section is concerned with delivering instructions to N operationally separated decision-making units to serve the objectives of a single human stakeholder. This problem does not reduce to the problem of instructing N separate AI systems to each serve the human on their own. This is because coordination solutions are needed to ensure the units interact productively rather than interfering with one another’s work. For instance, given multiple “actuator” units—each with the job of taking real-world actions to affect their physical or virtual environments—a separate “coordinator” unit could be designed to assist in coordinating their efforts. Conveniently, the role of the coordinator also fits within the Dec-POMDP framework as a unit with no actuators except for communication channels with the other units. 6.2.1 Direction 17: Hierarchical human-in-the-loop learning (HHL) Just as reports will be needed to explain the behavior of AI systems to humans and other AI systems, queries from subsystems may be needed to aid the subsystems’ decision-making at times when they have insufficient information or training to ensure safe and beneficial behavior. This research objective is about developing an AI subsystem hierarchy in a manner compatible with real-time human oversight at each level of the hierarchy. Social analogue. Many companies are required to undergo financial audits on a regular basis. For example, the California Nonprofit Integrity Act requires any charity with an annual gross revenue of $2 million or more to have their financial statements audited, on an annual basis, by an independent certified public accountant. This ensures that the taxpayer has a representative—the auditing firms—involved in the management of every tax-exempt company of a sufficient size. Suppose instead that California’s Franchise Tax Board attempted to audit every company itself; the FTB would quickly become overwhelmed by the amount of information to process. Hence, the notion of an auditing firm is a replicable and hence scalable unit of organization that allows for more pervasive representation of taxpayer interests, at a scale of authority that is intermediate between the employees of individual companies on the low end and the California Franchise Tax Board on the high end. 71 Scenario-driven motivation. Active learning—that is, machine learning driven by queries from the machine to a human about areas of high uncertainty—seems potentially necessary for ensuring any AI system makes economical use of the human labor involved in training it. It is likely possible to arrange AI systems into a hierarchy, as in Direction 16 (interpretable hierarchical reporting), where lower-level systems make queries to higherlevel systems. In such a set-up, human beings could be involved in answering the queries, either • only at the topmost level of the hierarchy, or • at all or most levels of the hierarchy. The latter option would seem better from an employment perspective: more roles for humans in the hierarchy means a reduction of Type 2b risk (economic displacement of humans), and if the roles involve maintaining valuable human skills, a reduction of Type 2c risk (human enfeeblement). Involving a human at each node of the hierarchy also seems better from the perspective of accountability and governance. Many human laws and accountability norms are equipped to deal with hierarchical arrangements of responsibilities, and hence could be applied as soft constraints on the system’s behavior via feedback from the humans. In particular, human-checked company policies could be implemented specifically to reduce Type 1b, 1c, and 1d risks (unrecognized prepotence, unrecognized misalignment, and involuntary MPAI deployment), and nation-wide or world-wide laws could be implemented to reduce Type 1a and 1e risks (uncoordinated MPAI development and voluntary MPAI deployment). The weight of these laws could derive in part from the accountability (or less euphemistically, the punishability) of the individual humans within the system if they fall short of their responsibilities to instruct the system according to safety guidelines. Such a system of accountability might feel daunting for whatever humans would be involved in the system and therefore accountable for global safety, but this trade-off could well be worth it from the perspective of existential risk and long-term human existence. Actionability. Engineering in this area would benefit from work on Direction 16 (interpretable hierarchical reporting) because of the improved understanding of the aggregate system that would accrue to the engineers. After deployment, in order for each human in the HHL system to oversee their corresponding AI system in a time-efficient manner, techniques would be needed to train each AI system to take a large number of actions with only sparse feedback from their human supervisor on which actions are good. Amodei et al. (2016) identify this issue as a problem in what they call “scalable oversight”, and propose to approach it via semi-supervised reinforcement learning (SSRL). In SSRL, where a managing or training system (which might involve a human) provides only sparse input into the decisionmaking of a reinforcement learner. They outline six technical approaches to scalable oversight, and potential experiments to begin work in this area. Sparse rewards are merely one piece of the puzzle needed to be solved to enable HHL. Abel et al. (2017) aim to develop a schema for “Humanin-the-Loop Reinforcement Learning” that is agnostic to the structure of the learner. Scaling up human-in-the-loop interaction models in a principled and generalizable manner is a rich technical challenge. To reduce confusion about whether solutions would be applicable for more complex or civilization-critical tasks, it is recommended that authors include in their 72 publications some discussion of the scalability of their solutions, e.g., as in Saunders et al. (2017). Consideration of side effects. Hierarchical decision-making structures present a clear avenue for general AI capabilities advancements. These advancements may fail to reduce existential risk if any of the following problems arise: • The institutions deploying the resulting AI systems choose not to involve enough humans in the hierarchy. For instance, the institution might prefer this outcome to speed up performance, or avoid accountability. • The AI systems in the hierarchy are insufficiently legible to the humans, i.e., if progress on Direction 16 (interpretable hierarchical reporting) has been insufficient, or not applied to the system. • The humans involved in the hierarchy are insufficiently individually motivated to think about and avert unprecedented catastrophic risks. • The humans in the hierarchy are uncomfortable discussing or reporting their concerns about unprecedented catastrophic risks. 6.2.2 Direction 18: Purpose inheritance As AI systems are used increasingly in the development of other AI systems, some assurance is needed that the deployment of a putatively “aligned” system will not lead to the creation of dangerous systems as a side effect. To begin thinking about this dynamic informally, if an AI system A takes actions that “create” another AI system B, let us say that B is a “descendant” of A. Descendants of descendants of A are also considered to be descendants of A. Given a satisfactory a notion of “creating a descendant”, we say that A has a heritable purpose to the extent that there is some purpose—that is, some internally or externally specified objective— which A’s own actions directly benefit, and which the collective actions of A’s descendants also benefit. This research direction is concerned with the challenge of creating powerful AI systems with any particular heritable purpose, with human survival being a purpose of special interest. While the precise definition of “creating a descendant” is interesting to debate, the relevant definition for this report is whatever notion can best guide our efforts to reduce existential risk from useful real-world AI systems. In particular, our notion of “creation” should be taken fairly generally. It should include cases where A creates B • “intentionally”, in the sense of being directed by a planning process internal to A which represents and selects a series of actions for their utility in creating B; • “subserviently”, in the sense of being directed by a human or another AI system with an intention to use use A as a tool for the creation of B; or • “accidentally”, in the sense of not arising from intentions on the part of A or other systems directing A. Whatever the definition, safety methods applicable for broader definitions of “descendant” will be able to cover more bases for avoiding existential risks from descendant AI systems. 73 Social analogue. A human corporation may be viewed as having a heritable purpose if it only ever creates subsidiary companies that effectively serve the parent corporation’s original purpose. To the extent that subsidiaries might later choose to defect against the parent’s mission, or create further subsidiaries that defect, the parent’s purpose would not be considered perfectly heritable. When a human institution builds an AI system, that system can be viewed as a “descendant” of the institution. So, if an AI system brings about human extinction, it could be said that human civilization itself (as an institution) lacks the survival of the human species as a heritable purpose. Scenario-driven motivation. An AI system with the potential to create prepotent descendants presents a Type 1b risk (unrecognized prepotence). As an unlikely but theoretically enlightening example, an AI system performing an unconstrained search in the space of computer programs has the potential to write an AI program which is or becomes prepotent. In general, it may be difficult to anticipate which AI systems are likely to instantiate descendants, or to detect the instantiation of descendants. At the very least, a powerful AI system that is not itself an MPAI, but which lacks human survival as a heritable purpose and is used to develop other AI systems, could constitute a Type 1c risk (unrecognized misalignment). For instance, an automated training system for developing machine learning systems could be used as a tool to develop an MPAI, and hence the training system would lack human survival as a heritable purpose. Actionability. Lack of technically clear definitions of “instantiate a descendant” and “heritable purpose” are obstructions to this research direction. Some definitions would be too restrictive to apply in reality, while others would be too permissive to imply safety results even in theory. Hence, next actions could involve developing clearer technical conceptions of these ideas that are adequate for the purposes of guiding engineering decisions and reducing existential risk. There are at least two distinct approaches one might consider: • Approach 1: Avoidance techniques. This approach develops an adequate definition of “instantiating a descendant”, and uses the resulting concept to design AI systems that entirely avoid instantiating descendants, thus obviating the need for purpose inheritance. There has not been much research to date on how to quantify the notion of “instantiating a descendant”, though a few attempts are implicit in literature on agents that “copy”, “teleport”, or “tile” themselves (Yudkowsky and Herreshoff, 2013; Orseau, 2014a,b; Soares, 2014; Fallenstein and Soares, 2015). One problem is that current theoretical models of AI systems typically assume a well-defined interface between the AI system and its environment, receiving inputs only via well-defined sensors and making outputs only via well-defined actuators. Such models of AI systems are sometimes called dualistic, after mind-body dualism. In reality, AI systems are embedded in the physical world, which they can influence and be influenced by in ways not accounted for by the leaky abstraction of their interface. Orseau and Ring (2012) consider a fully embedded version of AIXI (Hutter, 2004; Everitt and Hutter, 2018) and conclude that in this setting: “as soon as the agent and environment interact, the boundary between them 74 may quickly blur or disappear” (Orseau and Ring, 2012), but these works do not attempt to resolve the questions this raises about identifying descendants. Thus, a more general and real-world applicable notion of “instantiating a descendent” is needed. Alternatively, one could imagine a “know it when we see it” approach to defining the concept. However, such an approach might not scale well to regulating systems that could find ways of replicating and/or engineering new systems that humans would not easily recognize as cases of replication and/or engineering. Thus, a characterization of “instantiating descendants” that is simultaneously rigorous and realworld applicable is missing. The reader is invited ponder potential approaches to formalizing this problem. • Approach 2: Heritability results. Develop an adequate definition for “instantiating a descendant”, as well has “heritable purpose”, and use these conceptions in one of two ways: (a) Possibility results: Develop AI systems with the heritable purpose to serve and protect humanity as a whole, in particular by avoiding existential risks and MPAI deployment events; or (b) Impossibility results: Develop demonstrations or arguments that Approach 2(a) is too difficult or risky and that Approach 1 is better. These approaches are more difficult than Approach 1 because they involve more steps and concepts. Nonetheless, some attempts in this direction have been made. Yudkowsky and Herreshoff (2013); Fallenstein and Soares (2015) and others consider AI systems reasoning about the heritable properties of their descendants using logic, which remains a topic of ongoing research. One remaining challenge is to maintain the strength of descendants’ reasoning in the face of self-reference issues, which is addressed to some extent—at least asymptotically—by Garrabrant et al. (2016). It could also be valuable to empirically evaluate the propensity of agents based on current machine learning techniques to create descendants. For instance Leike et al. (2017) devise a toy grid-world environment for studying self-modification, where they consider the behavior of reinforcement learning algorithms. Considering more complex environments where descendants are still easy to identify by construction would be a good next step. Learning to predict which behaviors are likely to instantiate descendants in such settings would be also be useful. Consideration of side effects. Progress on possibility results in Approach 2(a) would be dual purpose, in that the results would likely create the theoretical capability for other purposes aside from “serve and protect humanity” to be inherited and proliferated. As well, progress on defining the notion of descendant in Approach 1 could be re-purposed for a better understanding of heritability in general, and could thereby indirectly contribute to dual purpose progress within Approach 2(a). 6.2.3 Direction 19: Human-compatible ethics learning It is conceivable that human-favorable behavior norms for a powerful AI system interacting with human society could be derived from some more 75 fundamental ethical abstraction, such as loyalty or reciprocity of an agent toward other agents that have allowed its existence, which would include humans. This research direction involves investigating that possibility. Social analogue. Many individuals experience a sense of loyalty to the people and systems that have empowered them, for example, their parents and teachers, their country of origin, the whole of human civilization, or nature. As a result, they choose to align their behavior somewhat with their perceptions of the preferences of those empowering systems. Scenario-driven motivation. It is conceivable that many peculiarities of human values will not be easily describable in terms of individual preferences. There may be other implicit constraints on the behavior of individual humans that would violate the von Neumann-Morgenstern rationality axioms for individual agents, but might be valuable at the scale of group rationality. For example, a person might reason “I won’t do X because if everyone did X it would be bad, even though if only I did X it might be slightly good.” Failing the development of an explicit theory for learning “nonpreferential” human values, a fallback option might be to discover cooperative ethical principles from scratch, and then test to see if they suffice for sustainable cooperation with humans. This would add another potential pathway to alignment, thereby reducing Type 1c risk (unrecognized misalignment). Perhaps the ethic “avoid acquiring too much power” could be among the ethical principles discovered, leading to a reduction in Type 1b risk (unrecognized prepotence). In principle, preference learning and ethics learning could be complementary, such that partial progress on each could be combined to build more human-aligned systems. Instrumental motivation. In addition to posing an complementary alternative to preference learning, work on human-compatible ethics learning could yield progress on • Direction 6 (preference learning) and Direction 24 (resolving planning disagreements). It is conceivable that a single basic principle, such as loyalty or reciprocity, would be enough to derive the extent to which an AI system should not only achieve preference learning with the human customer who purchases the system, but also with the engineers who designed it, and other individuals and institutions who were passively tolerant of its creation, including the public. The system could then in theory be directed to exercise some of its own judgment to determine the relative influence various individuals and institutions had in its creation, and to use that judgment to derive appropriate compromises between conflicts in their preferences. • Limited instances of Direction 28 (reimplementation security). A system which derives its loyalties implicitly from the full history of institutions and people involved in its creation—rather than from a simple “whom to serve” attribute—might be more difficult to redirect to serve the purposes of a delinquent individual, thus addressing certain instances of reimplementation security. Actionability. This direction could benefit from progress on Direction 13 (rigorous coordination models), to the extent that human-compatible ethics 76 will involve cooperation with humans. Decentralized learning of cooperation is more likely to be applicable than centralized learning of cooperation: when an AI system learns to cooperate with a human, the human’s beliefs and policies are not being controlled by the same training process as the AI system’s. That is, any group that includes humans and AI systems working together is a decentralized learning system. Implicit progress and insights might also be drawn from working on other research directions in this report, such as Directions 1, 6, 11, 20, and 24 (transparency and explainability, preference learning, deference to humans, self-indication uncertainty, and resolving planning disagreements). AI researchers will likely encounter disagreements with each other about how to operationalize ethical concepts such as loyalty or reciprocity to humanity, just as developing technical definitions of concepts like cause, responsibility, and blame have also been topics of debate among AI researchers (McCoy et al., 2012; Halpern, 2015). Hence operationalizing these concepts may need to go through numerous rounds of discussion and revision before researchers would converge on satisfactory definitions of what constitutes ethics learning, and what ethics are human-compatible. Consideration of side effects. In order to selectively advance technology that would enable human/machine cooperation rather than only machine/machine cooperation, studies of decentralized machine/machine cooperation will need to be thoughtful about how humans would integrate into the system of cooperating agents. Otherwise, these research directions might increase the probability of runaway economies of AI systems that cooperate well with each other at the exclusion of human involvement, increasing Type 2b and 2c risks (economic displacement of humans and human enfeeblement). 6.2.4 Direction 20: Self-indication uncertainty AI systems can be copied, and can therefore be implemented in numerous distinct environments including test environments, deployment environments, and corrupted environments created by hackers. It is possible that powerful AI systems should be required to be built with some awareness of this fact, which we call “self-indication uncertainty”. Social analogue. Self-indication uncertainty is not a matter of necessary practical concern for most humans in their daily life. However, suppose a human named Alice awakes temporarily uncertain about whether she is still dreaming. Alice may be viewed as being uncertain about whether she is “Real Alice” or “Dream Alice”, a kind of self-indication uncertainty. To put it another way, Alice is uncertain about whether her current perceptions and actions are taking place in the “real world” or the “dream world”. A more familiar but perhaps more tenuous analogy is the following. Suppose Alex is a supporter of a certain political party is considering staying home instead of voting, because he expects his candidate to win. He might find himself thinking thoughts along the lines of “If I stay home, does that mean many other supporters of my party will also stay home? And if so, doesn’t that mean we’ll lose?” Now, consider the mental subroutines within Alex that are deciding whether he should stay home, and generating the above question in his mind. These subroutines may be viewed as uncertain about whether they are deciding just for the one voter (Alex), or for a large number of “copies” of the same decision-making procedure inside the 77 minds of many other supporters of her party. In other words, the voteor-stay-home subroutine has self-indication uncertainty about who (and in particular, how many party members) it is operating within. Scenario-driven motivation. See instrumental motivations. Instrumental motivation. Progress on modeling or training selfindication uncertainty could be useful for some instances of: • Direction 10 (corrigibility). Ensuring that an AI system that is able to wonder if it is a misspecified version of its “true self” could aid in motivating the system to seek out corrections for those misspecifications. For example, consider an AI system which, after real-world deployment, maintains some degree of uncertainty about whether it is operating in a pre-deployment test environment. Such a system might be more likely to comply with shut-down commands if it believes noncompliance in the test environment would result in non-deployment and therefore no opportunity to pursue its real objective in the real world. It may even be the case that some degree of self-indication uncertainty of this form is needed for an AI system to exhibit the degree of “humility” that humans naturally exhibit and would like to see exhibited in AI systems. That is to say, it remains an open question whether implicit or explicit self-indication uncertainty is a necessary condition for corrigibility. • Direction 11 (deference to humans), Direction 17 (hierarchical humanin-the-loop learning (HHL)), and Direction 29 (human-compatible equilibria). A computerized decision algorithm that knows it is being implemented on many different machines at once might reason, when making a decision, about the consequence of all of its copies making that same decision, rather than fallaciously assuming that only one of its copies will do it. This could allow an individual AI system acting within a collective or hierarchy of other AI systems to derive and follow principles that are appropriate for the entire group, playing a role in the multi-agent dynamics of HHL and human-compatible equilibria. Perhaps deference to humans is such a principle. • Direction 12 (generative models of open-source equilibria). For each decision a game-playing agent takes, it could be made uncertain about whether the resulting action is taking place in its own reasoning (its true self) or in another player’s model of the agent’s reasoning (an approximate representation of its true self). • Direction 28 (reimplementation security). An AI system which knows it might be running on a hacker’s computer might be intrinsically motivated to take extra precautions before a highly-impactful action like disabling a portion of the electrical grid. Actionability. Any particular method for representing indexical uncertainty for an agent can be readily tested for efficacy in solving Decentralized POMDPs (Dec-POMDPs), a well-known problem class for stating decentralized coordination problems between agents with a common objective (Bernstein et al., 2002). For inspiration, Dec-POMDPs can also be related to older work on decision problems with imperfect memory (Piccione and Rubinstein, 1997), by observing that solving a Dec-POMDP is equivalent 78 to solving a “memoryless POMDP”, i.e., a POMDP where the solution policy is prohibited from passing information from one time step to the next. Piccione and Rubinstein (1997), and later Taylor (2016a) using more modern notation, have shown that any memoryless POMDP solution may be viewed as a Nash equilibrium, by viewing each memory loss as the creation of a new agent with the same utility function but with a certain form of uniform uncertainty about which agent it is. The form of uncertainty used in the equilibrium, which Bostrom calls the “Self-indication assumption” (Bostrom, 2007), could be incorporated in decentralized Dec-POMDP solvers and tested for scalability. Future work would need to account for the possibility that copies of an AI system can be made with only minor variations, in which case perhaps a non-Boolean concept will be needed for whether to view two such systems as sharing the same identity. Consideration of side effects. How to ideally implement self-indication uncertainty for an AI system is a subtle question, because it might enable the system to reason about whether it is operating within a test environment or a real environment. If the system is successful in determining whether or not it is undergoing a test, it could choose to exhibit different behavior in the test than in reality, thus invalidating tests used to assess the system’s safety. Hence, there are ways in which self-indication uncertainty as a capability could undermine certain approaches to safety. 6.3 Single/multi control Analogues of Direction 9 (generalizable shutdown and handoff methods), Direction 10 (corrigibility), and Direction 11 (deference to humans) all exist for a multi-agent system serving a single human principal. Achieving solutions to these problems in a single/multi control setting will likely involve challenges that did not exist in the single/single paradigm. This section aims to identify some such problems. For instance, what precisely does it mean to shut down an AI system if it is able to transmit a copy of its source code to another existing AI system prior to being turned off? Unlike in the other research sections of this report, actionable research directions are not identified in this section, only problems. Network design and administration: a present-day analogue. The design and administration of communications and computing networks may be viewed as solving present-day analogues of single/multi control problems, to the extent that network components can be viewed as very rudimentary AI systems. For instance, maintaining communications networks requires methods for modifying and shutting down network components. The patent literature includes techniques for upgrading a programmable device in communication with a server (San Martin et al., 2002), transmitting an emergency shutdown signal to a malfunctioning device in the network (Litwin Jr and Ramaswamy, 2006), and gracefully shutting down intermediate network nodes Scudder et al. (2008). Similarly, maintaining computing networks involves somewhat analogous methods, with patents including techniques for allowing a job on a node in a computing cluster to shut down that node if it is malfunctioning (Block et al., 2005), and putting to sleep or terminating idle virtual machines to sleep to conserve cloud computing resources (Huang et al., 2014; Rigolet, 2017). 79 However, each of the above methods seems to depend on components following pre-defined protocols, rather than learned protocols. Moreover, irrespective of the method, most of the available literature on the shutdown and maintenance protocols for communications and computing networks is contained in patents—which generally only contain enough detail to enforce ownership disputes—as opposed to research literature that is intended to convey knowledge. As such, there might be considerable room for academic progress in this area. Single/multi delegation will likely present novel control problems. Just as novel administrative challenges arise for present-day computing networks that are not needed for operating a single computer, single/single control solutions are not likely to be entirely sufficient to solve single/multi control problems. Consider the problem of safe shutdown for a multi-agent system. Any operationalization of the command “safely hand off control to a human or other system” from Direction 9 could be deployed in parallel to hand off control from each agent to another “backup” agent assigned to it, such as a human overseer. However, novel failure modes might then arise, including the following: • Unsafe composition of local shutdown protocols. Safe protocols for shutting down single agents in a multi-agent system might not be safe when applied to all the agents at once, e.g., if the resulting disruption to overall system performance would be unsafe. To give a human social analogy: while it might be relatively safe for one doctor at a hospital to take a sick day when they’re not feeling well, it would not be safe for all the doctors in the hospital to do so at the same time. • Malfuctioning of local shutdown protocols. If most agents in a multi-agent system successfully shut down as a result of a global shutdown command, but some agents remain active, the actions of the remaining agents might be highly unsafe outside of the context of rest of the system. To give a human social analogy: the action of a human pilot taking off an airplane is normally a safe action to take, but would be an incredibly unsafe action if air traffic controllers around the world were on strike. Thus, any procedure that takes air traffic controllers off the job had better take pilots off the job as well. What present-day AI research directions could be undertaken that could begin to address these issues? The task of identifying concrete next actions for single/multi control research, beyond the repeated local application of single/single control solutions, is a challenge left to the reader and future researchers. 7 Relevant multistakeholder objectives Before proceeding to discuss research directions on multi/single and multi/multi delegation, this section outlines some objectives that Sections 8 and 9 will build upon in their scenario-driven motivations. These objectives may also serve as general, high-level guidelines in the furtherance of multi/single and multi/multi delegation research. A diagram of the objectives and their most direct relationships is provided in Figure 10. Note on the meaning of “misalignment”. In a setting involving multiple stakeholders with diverse values, what should be considered an “aligned” AI system? While there is much room for debate about what constitutes alignment from the perspective of all of humanity, by contrast there is a great deal of agreement among people that the world becoming unsurvivable to humanity would be a bad outcome. More generally, there may be many outcomes that nearly everyone would agree are worse than the status quo, such that the concept of misalignment might be more agreeably meaningful than alignment in many multi-stakeholder scenarios of interest. In any case, for the purpose of this report, MPAI will continue to refer to AI systems whose deployment would be unsurvivable to humanity, as it was defined in Section 2.3.

7.1 Facilitating collaborative governance

As time progresses and the impacts of AI technology increase, existential safety concerns and other broadly important issues will likely lead to an increased pressure for states and companies to collaborate in the governance of AI technology.

What is collaborative governance? For the purposes of this report, collaboration between stakeholders in the oversight of AI technology refers to the exchange of reliable information and commitments between the stakeholders. Collaborative governance of AI technology refers to collaboration between stakeholders specifically in the legal governance of AI technology. The stakeholders could include representatives of governments, companies, or other established groups.

Making the governance of AI technology more collaborative, i.e., involving more exchange of information and commitments in the governance process, is not guaranteed to be safer or more effective, as elaborated somewhat below.

Moreover, the technical properties of AI systems themselves can add to or detract from the options available for multiple stakeholders to collaborate in the oversight of the systems’ activities. We therefore adopt the following objective:

Objective 7.1 (facilitating collaborative governance) is to make it easier for diverse stakeholders to collaborate in the oversight of powerful AI technologies, by the co-development of AI technology and accompanying governance techniques that will capture the benefits of collaboration in certain aspects of governance while avoiding forms of collaboration that would be unsafe or unnecessarily costly relative to independent governance.

This objective may be somewhat complex to achieve, because the potential benefits collaborative governance may also come with a variety of pitfalls that need to be avoided, as follows.

Potential benefits of collaborative governance. Consider a scenario where some powerful new AI capability is being implemented by multiple human institutions, collaboratively or independently, to pursue one or more purposes, such as:

* efficient distribution of electricity from power plants in a safe and equitable manner;
* global health research requiring difficult-to-negotiate privacy policies for patients;
* education tools that might enable the spread of cultural values that are difficult to agree upon; or
* environmental monitoring or protection systems that might require difficult-to-negotiate economic policies.

There are a number of reasons why the developing institutions might be motivated to collaborate in the governance of this technology, including:

1. to ensure fair representation of diverse views and other objectives in governing their system(s);
2. to pool the collective knowledge and reasoning abilities of the separate institutions; or
3. to ensure sufficient weight is given to other objectives that are of interest to everyone involved (such as existential safety), relative to objectives only of interest to one person or institution.

Items B and C here point to an existential safety argument for collaboration in the governance of AI systems: a committee of representatives from different institutions of would be less likely to accidentally (by B) or intentionally (by C) take risks that a single institution might be willing to take. This consideration is elaborated further in Objective 7.3 (reducing idiosyncratic risk-taking).

Pitfalls of collaborative governance. In pursuing collaborative governance for AI systems, it is important to be mindful that collaborative governance does not guarantee better outcomes than independent governance. In general, too much collaboration or the wrong kinds of collaboration between institutions can in general lead to a variety of problems:

* Fragility: if the institutions become more dependent upon one another through collaboration, a failure of one institution risks failure of the other.
* Interference: the institutions’ operations could become entangled in unexpected ways, leading to unexpected errors.
* Collusion: by collaborating, the institutions could gain too much power or influence relative to other institutions or the public; antitrust and competition laws exist to prevent these outcomes.
* Groupthink: membership in a group can sometimes cloud the judgement of individuals, by a process known as groupthink (Janis, 1971; Hart, 1990; Janis, 2008; Esser, 1998; Janis, 2008; Bénabou, 2012). In groupthink, individual beliefs are warped to match the prevailing group consensus. Collaboration between institutions might reduce groupthink within each institution by exposing individuals to views from outside their institution, but it could also increase groupthink if the institutions begin to view themselves as a single large group.

Innovations in collaborative governance for powerful AI systems should aim to account for these and other failure modes of collaborative decisionmaking that would be harmful to many objectives, including safety.

How and when should governance be collaborative? When, and in what ways, can collaborative governance of AI systems be more effective than independent governance by essentially separate institutions? This is a daunting and multi-faceted question that is beyond the scope of this report to resolve. However, we do aim instigate some technical thinking in this area, particularly as pertaining to existential safety.

Sources of historical lessons. Absent a satisfying theory of how and when to collaborate in the governance of powerful AI systems, studies of successes and failures in the oversight of safety-critical technologies could yield informative lessons with implications at various scales of governance.

On the failure side, Sasou and Reason (1999) have developed a broad taxonomy of team decision-making failures in the oversight of safety-critical systems, through examining case studies in aviation, nuclear power, and the shipping industry. Charles Perrow’s widely cited book Normal Accidents (Perrow, 1984)—written partially in response to Three Mile Island nuclear accident of March 1979—predicts catastrophic failure in hazardous systems when those systems involve “complex and tightly coupled” interactions. Subsequent technological disasters are also considered in the 1999 edition (Perrow, 1999), such as the Bhopal industrial chemical leak in India in December 1984 (Shrivastava, 1992), the explosion of the US space shuttle Challenger in January 1986 (Vaughan, 1996), and the Chernobyl nuclear accident in Russia in April 1986 (Meshkati, 1991). Perrow contrasts these events with “normal accidents”, concluding that they involved serious managerial failures and were not inevitable consequences of the underlying technological systems.

On the success side, positive lessons can be taken from human institutions with strong track records for the safe provision of highly valued services in hazardous industries. This point has also been argued somewhat by Dietterich (2019). There is an existing corpus of academic studies examining so-called high reliability organizations (HROs), i.e., “organizations that operate beneficial, highly hazardous technical systems at high capacity with very low risk, for instance, the effective management of physically (and often socially) very hazardous production processes with very low incidents of operational failure” (LaPorte and Thomas, 1995). Examples of organizations identified and studied closely as HROs by organizational researchers include

* two nuclear-powered aircraft carriers (Rochlin, 1989; Roberts, 1989, 1990; Roberts et al., 1994; Schulman, 1993),
* the US Federal Aviation Administration’s Air Traffic Control system (Roberts, 1989; Klein et al., 1995),
* several nuclear power plants (Klein et al., 1995; LaPorte and Thomas, 1995; Bourrier, 1996),
* electricity providers (Roberts, 1989; Schulman et al., 2004), and
* a large California fire department (Bigley and Roberts, 2001).

HRO researchers have gone on to produce theories and recommendations for organizations in general to achieve high reliability (LaPorte, 1996; Rochlin, 1999; Roberts and Bea, 2001a,b; Ericksen and Dyer, 2005). Perhaps similar theories could someday be formulated quantitatively as principles for multi/single and multi/multi AI delegation in powerful AI systems.

Summary. Collaborative governance of AI systems is attractive from the perspective of issues that concern everyone, such as existential safety. However, collaborative governance is not automatically more effective than independent governance. The objective of this subsection, facilitating collaborative governance, means finding collaborative AI governance techniques that are beneficial from many perspectives (including existential safety), and that avoid pitfalls of collaborative governance. How exactly to achieve this is a complex social question that is beyond the scope of this report to answer, but is something the authors are beginning to explored somewhat at a technical level.

7.2 Avoiding races by sharing control

If powerful AI technology is developed in a manner that makes it difficult for multiple stakeholders to share control of a single system, there is some degree of pressure competing stakeholders to race in AI development so as to secure some degree of control over the how the technology is first used. Conversely, the pressure to race can be alleviated somewhat by developing AI technology in a manner that makes it easier for multiple stakeholders to control a single system, such as by designing the system to receive inputs representing beliefs and values from multiple users. Hence, we adopt the following objective:

Objective 7.2 (avoiding races by sharing control) is to make collaborative oversight of AI systems by companies and governments sufficiently easy and appealing as to significantly reduce pressures for AI development teams to race for first-mover advantages in the deployment of powerful AI systems, thereby reducing Type 2a risk (unsafe development races). The nature of the collaboration between the overseeing stakeholders could involve exchange of information, exchange of commitments, or both.

This objective may be challenging to pursue while respecting the letter and spirit of antitrust laws. Thus, some degree of progress on Objective 7.1 (facilitating collaborative governance) may be needed to ensure that control-sharing between companies cannot lead to collusion or other unfair business advantages that would harm society.

## 1AC---Plan

#### The United States federal government should substantially increase its prohibitions on anticompetitive business practices by the private sector that preclude litigants from effectively vindicating their statutory causes of action in antitrust suits.

## 1AC---Solvency

#### Arbitration assessment is biased towards enforcement---requiring effective vindication solves, balances competing law, and restores statutory relief under antitrust.

Rubinoff ’20 [Matt; May 1; J.D. Candidate at Pennsylvania State University, Managing Editor; Arbitration Law Review, “Too Big to Arbitrate? Class Action Waivers, Adhesive Arbitration, and Their Effects on Antitrust Litigation,” vol. 12]

(d) The Post-Epic Systems World

The Epic Systems decision reignited an already tense debate on the status of class action waivers and mandatory arbitration clauses, and their possible detrimental effects on the working public.56 In the almost two years since the 2018 ruling, lower court decisions have just started to show the lasting impact of the Court’s unwavering support for enforcing adhesive arbitration clauses.57 Notably, the First, Fifth, Sixth, Seventh, Ninth, and Eleventh Circuit Court of Appeals have already addressed the remnants of the Court’s decision in Epic Systems. 58 As expected, early case decisions show strong, continued judicial support for the enforcement of class action waivers, even when class action proceedings result in a valid award for plaintiffs.59 Though decidedly settled in the current judicial system, removal of access to class action remains a contentious issue in the debate around arbitration clauses, especially as it relates to existing federal antitrust laws.

III. The Antitrust Landscape in an Adhesive Arbitration World

(a) Land of the Free, Home of the Lawsuit

Much of the conversation around mandatory arbitration agreements and the inclusion of class action waivers has to do with the perceived fight between the strong and the weak.60 David versus Goliath; the little consumer or employee versus the large corporation. As the Supreme Court continues to support the enforceability of adhesive arbitration agreements, concerns as to the rights of workers and consumers come front and center in the debate.61

Historically, Americans have always loved lawsuits.62 Any one citizen’s access to the court system provides for a sense of individual power, freedom, and the innate ability to right a wrong. Recently, however, the frequency of class action lawsuits has trended downward.63 On the heels of Epic Systems, according to publicly available data, “2019 marked the first year in more than a decade that there were fewer federal class action lawsuits alleging unpaid wages, job discrimination, and mishandled retirement benefits.”64 Further, the Economic Policy Institute estimates that by 2024, around eighty-three percent of private sector workers, nearly ninety-five million people, will be forced to resolve disputes through adhesive arbitration agreements.65

Those in favor of adhesive arbitration agreements suggest that the pre-determined adjudication procedure brings “efficiency and economy to the marketplace for all its participants.”66 While it is true that arbitration allows for quicker processing and award determinations compared to lagging judicial procedures, the theory does not consider the elimination of access to the marketplace for some of those participants.67 Similar to discouraged workers no longer factoring into a society’s estimated unemployment rate, sometimes the elimination of previous market actors prevents us from seeing the forest for the trees. 68 Accordingly, this analogy is an apt comparison to the state of adhesive arbitration and its effects on antitrust laws and litigation.

(b) Mitsubishi and the Effective Vindication Doctrine

In 1985, the Supreme Court attempted to peacefully mediate the differences between the enforcement of arbitration laws and antitrust violations. In Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth Inc. (“Mitsubishi”), the Court held that claims under American antitrust laws, as they relate to international agreements, may be submitted to arbitration so long as the agreement includes a valid arbitration clause.69 The Court reasoned that “so long as the prospective litigant effectively may vindicate its statutory cause of action in the arbitral forum, the statute will continue to serve both its remedial and deterrent function.”70 However, the Court also noted a plaintiff has the opportunity to show that the arbitral forum does not fully vindicate his or her rights, and upon that showing, a court may hold the arbitration agreement unenforceable and permit the plaintiff’s pursuit of the claim in federal court.71

Following Mitsubishi, this “effective vindication doctrine” provided an out for antitrust claimants who believed their cases were not suited for an arbitral forum, whether or not they signed a mandatory and adhesive agreement.72 Following this decision, for example, courts used the doctrine to invalidate class action waivers, as well as provisions that prevented an arbitrator from awarding the appropriate treble damages under an antitrust claim.73 For nearly three decades, courts struck a harmonious balance between enforcing arbitration agreements under the FAA and allowing valid antitrust claims to proceed in federal court when necessary or otherwise “vindicated.”74 This balance, however, ultimately tipped well in favor of arbitration enforceability in Italian Colors.

As described above, Italian Colors held in favor of enforceability of arbitration agreements under the FAA, and that the federal laws do not support the invalidation of a class arbitration waiver solely because the costs involved in the process exceed the total potential recovery. 75 The plaintiffs’ arguments included citing the effective vindication doctrine, claiming that the costs from mandatory individual arbitration proceedings, such as finding and preparing expert witness testimony, significantly prevented their rights to proper antitrust remedy in the form of class action adjudication.76

In dismissing these claims, the Court dismantled the credibility of the effective vindication doctrine, describing the doctrine as dictum. 77 Without this practical safeguard, scholars suggest the post-Italian Colors world leaves “potential antitrust defendants . . . more likely to use arbitration clauses to substantially reduce the probability of antitrust liability and the amount of damages recovered by successful antitrust plaintiffs.”78 Enforcing class action waivers in mandatory arbitration clauses prevents plaintiffs from accessing traditional antitrust litigation, and confers significant disadvantages to defendants. 79

Footnote 77:

77 See id. at 235 (“The ‘effective vindication’ exception to which respondents allude originated as dictum . . . where we expressed a willingness to invalidate . . . arbitration agreements that ‘operat[e] . . . as a prospective waiver of a party’s right to pursue statutory remedies’”)

End of footnote 77.

#### It threads the needle---course correction must be federal and empower private litigation while avoiding overdeterrence.

Loureiro ’19 [Raul; 2019; J.D. upon the year of publication from the University of Pennsylvania, writing under the guidance of Professor Stephen Burbank; University of Pennsylvania Journal of Business Law, “Ineffective Vindication of Antitrust Rights,” vol. 21]

Introduction

The policy of antitrust law in the United States is to increase consumer welfare, and this policy is undermined by the Supreme Court’s most recent interpretation of the principle of effective vindication. In this paper, I argue that a dynamic interpretation of the principle of effective vindication advances the policy goal of antitrust. Without a robust principle of effective vindication, it becomes far too easy for potential antitrust defendants to use arbitration agreements to shield themselves from antitrust liability. This is particularly problematic in the antitrust context given the large amount of enforcement that occurs through private suits.

The efficacy of the principle of effective vindication depends upon the interpretation of the principle as a dynamic concept. If a court is able to invalidate an arbitration agreement only under very narrow circumstances, then the principle’s purpose is undermined. The vitality of this doctrine is conditioned on the ability of a court to consider whether someone is functionally precluded from vindicating his federal statutory right regardless of whether an agreement precludes vindication on its face. The majority decision in the case of American Express Company v. Italian Colors Restaurant undermines the doctrine of effective vindication by conceptualizing the doctrine as static.1 This decision undermines the antitrust policy of maximizing consumer welfare by obstructing the private enforcement of antitrust rights.

I. The Principle of Effective Vindication

The Federal Arbitration Act 2 (FAA) made arbitration agreements “valid, irrevocable, and enforceable, save upon such grounds as exist at law or in equity for the revocation of any contract.” 3 This language has been interpreted by the Court as creating a strong policy in favor of arbitration agreements.4 The savings clause of this provision indicates that an arbitration agreement can be invalidated on contract common law grounds.5 However, in AT&T Mobility LLC v. Concepcion, the Court held that a state policy against class waivers in arbitration agreements on the basis of unconscionability was invalid because “[s]tates cannot require a procedure that is inconsistent with the FAA, even if it is desirable for unrelated reasons.”6 This decision preempting state contract policy weakened the savings clause insofar as unconscionability is a reason, rooted in common law, for invalidating a contract.

Assuming that Concepcion reduces the ability of courts to invalidate arbitration agreements, the decision makes the federal policy in favor of arbitration much stronger. This decision left some uncertainty regarding the situations under which a court could invalidate an arbitration agreement that included a class action waiver.7 One possible answer is using the effective vindication principle to invalidate such agreements, which “might prove more like the eye of a needle through which claimants must pass to gain refuge from class action waivers” rather than an open floodgate for arbitration agreement invalidation. 8

The Court first articulated the effective vindication principle in Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc., saying that the federal policy in favor of arbitration agreements applies “so long as the prospective litigant effectively may vindicate its statutory cause of action in the arbitral forum.”9 The Court was explicit: “By agreeing to arbitrate a statutory claim, a party does not forgo the substantive rights afforded by the statute; it only submits to their resolution in an arbitral, rather than a judicial, forum.”10 It stands to reason that under the principle of effective vindication an arbitration agreement would be invalidated if it resulted in a party forgoing its substantive rights.

II. The Italian Colors Case

The effective vindication principle was examined and all but rejected in Italian Colors. 11 In this case, the Court held that an arbitration clause that made a suit prohibitively expensive was valid under the effective vindication principle because, “the fact that it is not worth the expense involved in proving a statutory remedy does not constitute the elimination of the right to pursue that remedy.”12 This decision undermines the principle of effective vindication and limits its ability to be of any use by making a meaningless distinction.

The dispute in Italian Colors arose from an agreement between American Express and various merchants, which accept American Express cards as payment from their customers.13 The agreement in question contained an arbitration clause in which the merchants waived their right to bring suit as a class.14 Importantly, the agreement also foreclosed potential plaintiffs from “joinder or consolidation of claims, informal coordination among individual claimants, or amelioration of arbitral expenses.”15 The merchants claimed that American Express violated §1 of the Sherman Act16 by a tying arrangement between their charge cards and credit cards.17

Proving a violation of the antitrust laws is expensive because of the economic experts that have to be contracted to provide analysis of the market and the anticompetitive effects of the challenged behavior.18 In this case, “the cost of an expert analysis necessary to prove the antitrust claims would be at least several hundred thousand dollars, and might exceed $1 million, while the maximum recovery for an individual plaintiff would be $12,850, or $38,549 when trebled.”19 The key is that this assessment is in regards to an individual plaintiff. However, bringing the suit through some sort of joinder device would overcome the difficulty of prohibitive costs because the various plaintiffs would be able to share the cost of the proceedings.20

A. Narrowing Effective Vindication

The majority opinion in Italian Colors limits the principle of effective vindication through narrowing the circumstances under which the principle applies. The Court confirmed the existence of the principle, and notes that the principle “would certainly cover a provision in an arbitration agreement forbidding the assertion of certain statutory rights. And it would perhaps cover filing and administrative fees attached to arbitration that are so high as to make access to the forum impracticable.”21 The Court does not say in which situations—other than an agreement that forbids the right to pursue a statutory right—the principle will apply. 22 It is notable that the majority also qualifies the application of the principle in the situation of prohibitive court fees by saying that the principle “perhaps” applies in this situation. Although the Court indicates otherwise, it seems like it is only affirming the principle insofar as the arbitration clause explicitly forbids a potential plaintiff from making a claim for a given statutory right. It is difficult to see under what other circumstances the principle would apply when even prohibitive administrative fees are only “perhaps” covered.

The decision to narrow the effective vindication principle is defended in two ways, the first is through an appeal to pre-class action jurisprudence and the second is through a formalistic view of the principle. First, the opinion argues that antitrust suits existed before class action proceedings, and “the individual suit that was considered adequate to assure ‘effective vindication’ of a federal right before adoption of class-action procedures did not suddenly become ‘ineffective vindication’ upon their adoption.”23 But, this case was not only about the class action waiver, instead it had to do with all of the provisions in the agreement that prevented a claim from being economically feasible. 24 The class action waiver by itself does not render the vindication of this right ineffective, rather the result of ineffective vindication stems from the agreement as a whole.

It is because the Court looks narrowly at the agreement solely in terms of the class action waiver that it can make the argument based on the history of class action procedures. The dissent indicates the root of the majority’s argument: “To a hammer, everything looks like a nail. And to a Court bent on diminishing the usefulness of Rule 23, everything looks like a class action, ready to be dismantled.”25 In fact, the majority opinion does not address whether the provisions in the agreement that prevent fee shifting or sharing of the expert report have any impact on the matter.26 These provisions must be of great import to the case because had either of them not existed, the plaintiffs would have been able to proceed and vindicate their rights.27

The majority opinion fails to consider the bearing that the other provisions foreclosing economical adjudication have on the matter. Instead of seriously addressing these, the Court decided to assert that the other provisions were not important because they were not part of the holding in the court below.28 However, the second circuit decided the case based upon the notion that there was no way that the suit could be brought in a way that was economically rational. This rationale necessitates taking into account the provisions in the agreement that foreclosed other means of cost sharing.29 Therefore, the first defense for narrowing the effective vindication principle rests on a failure to take into account the entire agreement.

Moreover, the appeal to the pre-class action era ignores the fact that the cost of proving an antitrust claim has increased dramatically ever since the introduction of sophisticated economic analysis into antitrust doctrine. Not only is the cost of experts higher, but also the increased costs of discovery in the digital age militate toward a reconceptualization of the scope of the effective vindication principle.30 The cost of litigation and expert testimony in a modern antitrust case is another factor that the majority fails to consider in its first argument defending its position.

#### Removing barriers to private, class action litigation is crucial to overall deterrence of anticompetitive conduct---particularly for cartels---empirical data proves.

AAI ’21 [American Antitrust Institute; August 4; Washington, D.C.-based non-profit education, research, and advocacy organization, citing research conducted by Professor John Davis at the University of San Francisco School of Law; American Antitrust Institute, “The Critical Role of Private Antitrust Enforcement in the United States,” <https://www.antitrustinstitute.org/wp-content/uploads/2021/08/Huntington-Report-FINAL-1.pdf>]

I. Introduction

With all of the attention currently focused on public enforcement and legislative reform of the antitrust laws, less attention is being paid to private enforcement.1 But Congress considered private antitrust enforcement indispensable for promoting competition. The judiciary has so recognized time and time again. In California v. American Stores Co., for example, the Supreme Court proclaimed, “Private enforcement of the [Clayton] Act was in no sense an afterthought; it was an integral part of the congressional plan for protecting competition.”2

Private enforcement is not a substitute for vigorous public enforcement. Both are necessary to foster competition. But private enforcement plays an important part, one that becomes more significant when public enforcement recedes.3 And, unlike public enforcers, private enforcers can obtain significant damages on behalf of the victims of antitrust violations.4 This serves as a crucial source of deterrence for illegal anticompetitive conduct and the primary means of compensating victims for harms suffered at the hands of cartelists and dominant firms.5 The importance of the antitrust class action, a major private enforcement device, is clear. The recently released 2020 Antitrust Annual Report: Class Action Filings in Federal Court (“2020 Report”)6 by Huntington National Bank and the University of San Francisco School of Law (“USF Law”) reflects that the cumulative total settlement amount recovered for victims in antitrust class actions from 2009-2020 was over $27 billion.7

Antitrust class actions recover damages from companies engaged in harmful, illegal conduct, such as price fixing and attempted monopolization, in markets for important and essential products and services. The most active defendants during the period, for example, included companies providing financial services, pharmaceuticals, automobile parts, and electronics parts. 8 In light of the vital role played by private antitrust enforcement, and the antitrust class action in particular, continued empirical analysis of trends in activity is essential. This analysis aids in understanding and evaluating proposals for reforming the antitrust laws in the U.S. and such proposals’ impact on private enforcement, the public-private partnership, and ultimately on competition and consumers.

II. Overview of the Commentary

The American Antitrust Institute (AAI)9 and Professor Joshua P. Davis at USF Law10 evaluated the 2020 Report with the goal of identifying its major implications for private enforcement in the U.S. The 2020 Report builds and expands on the 2019 Antitrust Annual Report: Class Action Filings in Federal Court (“2019 Report”)11, which assessed private enforcement activity from 2009-2019. Like the 2019 Report, the 2020 Report relies largely on data for private U.S. antitrust class actions available through Lex Machina, as well as supplemental data analysis.12 The 2020 Report extends the dataset to the eleven-year period covering 2009-2020, thus allowing for a deeper analysis of private enforcement trends and their implications. The analysis provided in this Commentary highlights the importance of private antitrust enforcement in the U.S. system and the particularly important role played by the antitrust class action.

III. Observations and Implications for Private Enforcement

The 2020 Report provides further evidence of a divergence between public and private enforcement trends. As public enforcement has waned, private filings have waxed, undermining the notion that class actions simply ride the coattails of public enforcement. On the contrary, the data suggest that as lax public enforcement fosters higher market concentration and invites bad behavior, private filings may compensate for underenforcement in an effort to address the resulting antitrust violations.

Looking beyond the number of enforcement actions filed and focusing on the results of the actions, the rich data reveal more nuance to this narrative. Despite increased private actions in the face of decreased public enforcement, the amount of money recovered from violators by both public and private enforcers has diminished. For public enforcers, this diminution is to be expected, as fewer cases have been brought. For private enforcers, though, the explanation likely lies with other trends, most notably the increasing headwinds faced by private enforcers due to heightened pleading and class certification standards. If these explanations are correct, the clear implication is that for private enforcement to fulfill its increasingly vital role as a complement and a backstop to public enforcement, these trends must be reversed.

A theme we noted in the 2019 Report, and that continues to feature in the 2020 Report, is the tremendous variability in the data on some measures related to settlement size. Aggregate settlement amounts over the period vary widely from year to year. By disaggregating the settlements by size, however, we are able to observe that settlements at different levels trend somewhat independently. Very large settlements, which are few in number, drive most of the variability in the aggregate data. But trends and anomalies in very small settlements cannot be entirely discounted, as they are the force behind one of the highest recovery years in the period, 2018.

Finally, building on analysis from our 2019 commentary, we take a deeper dive on attorneys’ fees and how they correspond to settlement amounts. Our findings reinforce the tentative conclusion from last year’s analysis that the so-called “megafund doctrine”—a dramatic decrease in attorneys fee percentages on settlements above a threshold of about $100 million—does not operate in federal antitrust cases in a significant way. Rather, decreases in the fee award percentage in antitrust cases are not discrete and drastic, but rather gradual, much like marginal tax rates in the United States. In what follows, we discuss each of the above observations in more detail and provide analysis of their implications for private enforcement, many of which suggest fertile areas for additional study.

A. The Relationship of Private to Public Enforcement: Riding on Coattails or Stepping into the Breach?

A long-running antitrust policy debate centers on the value that private enforcement adds to the antitrust enterprise as a whole. Critics of private actions maintain that private plaintiffs often follow an easy trail blazed by government enforcers, and that private enforcement therefore does not supplement worthwhile public actions as much as it should. Proponents of private actions have sought to debunk this claim with empirical evidence suggesting that many large and successful private antitrust cases often precede, or else expand the scope of relief sought in, any overlapping government actions.13

Data from the 2020 Report, together with future reports, may warrant a reexamination of this debate through a different lens. Apart from the question of the extent to which private enforcement serves as a useful complement to public enforcement, it may also be worth inquiring into the extent to which private enforcement serves as a substitute for public enforcement, particularly during periods of government forbearance.

As AAI noted in an independent report, “The State of Antitrust and Competition Policy in the U.S.,” several indicators had suggested a decline in government cartel enforcement at the midway point of the Trump Administration.14 Perhaps most notably, the average number of cartel investigations opened during the period from 2017-2018 was 80% lower than the annual average number of investigations opened over the previous three administrations (1993-2016). In addition, the average number of corporations fined by the Trump agencies in 2017 and 2018 fell by about 45% relative to the Obama Administration.

Notably, the 2020 Report shows that private federal consolidated antitrust filings rose significantly during each year of the Trump Administration. First, they increased from 74 in 2017 to 136 in 2018. But in 2019, they rose dramatically to a ten-year-high of 211. And, in 2020, that number was eclipsed, with 220 filings. In other words, a dramatic increase in private filings appears to have occurred immediately subsequent to a substantial decline in the opening of government cartel investigations and the number of corporations fined.

# 2AC

## Adv---Class Action

### Biz Con DA---2AC

#### Cartelization is driving massive economic slowdown…

#### No link---private suits are tailored to avoid disrupting business.

Graham ’20 [Victoria and Jake; October 7; Reporters; Bloomberg Law, “Slew of Private Antitrust Suits Awaits Big Tech Under House Plan,” <https://news.bloomberglaw.com/antitrust/slew-of-private-antitrust-suits-awaits-big-tech-under-house-plan>]

The increased potential for private lawsuits isn’t likely to have the same effect as government-led litigation, however.

“When we’re talking about Big Tech and breaking up these companies, that’s not the type of stuff that’s going to be addressed in class actions,” according to Stephen Calkins, a law professor at Wayne State University who studies antitrust.

Consumer antitrust litigation is typically a follow-on to federal lawsuits and part of the system of deterrence, Calkins said. It can be useful in addressing potentially anticompetitive behavior such as mandatory bundles, hidden charges, and automatic renewals.

But these typically aren’t the types of cases that change how these companies do business, he said.

Bipartisan Backing

The arbitration proposal is one of more than two dozen contained in a far-reaching report on big technology companies’ power, [released](https://news.bloomberglaw.com/mergers-and-acquisitions/house-panel-calls-for-sweeping-antitrust-reforms-for-big-tech) Tuesday by the House Judiciary Subcommittee on Antitrust, Commercial, and Administrative Law. The report [outlines](https://judiciary.house.gov/uploadedfiles/investigation_of_competition_in_digital_markets_majority_staff_report_and_recommendations.pdf) how Congress can better tailor antitrust laws to rein in the power of Amazon, Alphabet Inc.'s Google, Facebook Inc. and Apple.

The report also proposes removing barriers that have made it significantly harder for consumers and employees to sue powerful corporations through class actions.

Republicans have stated their opposition to the report itself, but recent [bipartisan efforts](https://news.bloomberglaw.com/daily-labor-report/house-passes-ban-on-forced-arbitration-class-action-limits) to prohibit mandatory arbitration agreements in employment, consumer, and other contracts suggest that particular aspect of the report stands on firmer ground.

Doing away with forced arbitration is one proposal “that is less out there or radical,” Sussman said.

#### Link is zero---threshold is empirical data.

Davis ’17 [Joshua and Robert Lande; 2017; Professor and Director of Center for Law and Ethics at the University of San Francisco; Venerable Professor of Law at the University of Baltimore, M.P.P. and J.D. from Harvard University; Scholar Works, “Restoring the Legitimacy of Private Antitrust Enforcement,” Ch. 6]

II. The Current Level of Private Enforcement Is Not Excessive

Notwithstanding the benefits of vigorous private enforcement, critics maintain that private antitrust enforcement in the United States is excessive, that it leads to overdeterrence, and that it promotes widespread frivolous antitrust litigation.31 These are myths.

A. The Number of Antitrust Cases is Modest

The number of new private federal antitrust cases has declined significantly during the last 30 years.32 The number of cases filed peaked in 1977 at 1611, dropped steadily in the 1980s to a low of 452 in 1990, averaged 600 cases per year in the 1990s, and has increased modestly to an average of 760 cases per year since 2008. The number of private federal antitrust actions filed as a percentage of the total number of civil cases filed in the federal district courts has fallen by approximately 75% from 1.2 % in 1977 to 0.26% in 2014.

[Table 1 omitted].

B. There is No Evidence of Overcompensation or Duplicative Recoveries

Some claim that private enforcement could result in overcompensation of victims, or duplicative recoveries, when direct purchasers sue under the Sherman Act and indirect purchasers sue under state laws.33 Some claim that this combination could result in a total of sixfold or more damages for antitrust violations.34 Yet, no evidence of even a single example where victims have received more than treble damages has ever been presented.35

A recent AAI study analyzed the overcompensation/duplication issue empirically by assembling a sample of every completed private U.S. cartel case since 1990 for which the authors could find the necessary information.36 For each of these 71 cases, the study collected neutral scholarly estimates of affected commerce and overcharges. It compared these to the damages secured in the private cases filed against these cartels.

#### Antitrust drives instantaneous economic recovery---empirics prove it massively boosts confidence

Elhauge 16 Petrie Professor of Law at Harvard Law School and Founding Director of the Petrie-Flom Center for Health Law Policy, Biotechnology and Bioethics.  He served as Chairman of the Antitrust Advisory Committee to the Obama Campaign. (Einer, <https://harvardlawreview.org/wp-content/uploads/2016/03/1267-1317-Online.pdf>, emuse)

Another big economic puzzle in recent years has been why, at a time when corporate profits have been at record highs, corporations have been so reluctant to invest those profits on expanding output. Ordinarily, high profits induce corporations to invest in expansion to try to get a greater share of those high profits, and that expansion in turns leads to high levels of economic growth and employment. Recently, after-tax corporate profits have risen to record levels of nearly $2 trillion per year, four times the corporate profits in the late 1990s and higher as a percentage of GDP than at any time in the last sixty years.67 Despite those record profits, U.S. corporate investments in expansion and capital projects have fallen; indeed, as a percentage of GDP, corporate investments were over 10% higher in 2000 than they were in each quarter from 2012 to July 2015. 68 Nor have firms been making up for the relatively low level of investments by heavily using their existing capacity to increase output. As of October 2015, U.S. capacity utilization was 77.7%, which remains below the long-term average of 80.1%.69 Instead of spending to expand output, S&P 500 companies have retained between $3.5 trillion and $5 trillion in cash and spent other profits on stock buybacks, dividend payments, and high executive compensation.70 To be sure, the United States has managed to return to sluggish growth since the Great Recession. But only at the cost of massive deficit spending that has increased our national debt by $9 trillion since 2008 and enormous monetary stimulus that has not only set short-term interest rates at virtually zero percent for years, but has also involved a “quantitative easing” program that effectively involved printing money to purchase $3.5 trillion in long-term securities.71 In a nation of 321 million people,72 this $12.5 trillion stimulus amounts to $39,000 per person even if one puts aside the fact that short-term interest rates have been set at virtually zero percent. Despite this massive fiscal and monetary stimulus, the labor force participation rate has dropped from 65.8% in February 2009 to 62.4% in October 2015, which is the lowest it has been since 1977. 73 The unemployment rate has fallen, but that partly reflects the fact that fewer people in the labor force are looking for work,74 which is quite rational given that there are fewer good jobs available than there should be. Further, the labor share of income is now at historically low levels.75 For some reason, while all this stimulus has produced high corporate profits, it has not produced the expected level of business expansion that would seriously increase employment levels and wages. As Paul Krugman has observed, “this kind of divergence — in which high profits don’t signal high returns to investment — is what you’d expect if a lot of those profits reflect monopoly power rather than returns on capital.”76 But what would that unexplained exercise of monopoly power be? After all, the United States has antitrust laws to curb anticompetitive creations of market power, and those antitrust laws are actively enforced by government agencies and private actors. Perhaps the explanation is that horizontal shareholdings are now pervasive because more and more stock is in the hands of institutional investors, but so far there has been no antitrust enforcement against horizontal shareholdings because the anticompetitive problem had not been appreciated until now.77 With such horizontal shareholdings, firms acting in the interests of their shareholders have incentives to constrain output rather than expand. The high profits they reap are not a signal to competitively expand individual firm output. Rather, the high profits are a symptom of the fact that they have successfully constrained overall market output. This could help explain why high corporate profits have not led to expansion and higher economic growth and employment levels. To be sure, one might doubt that anticompetitive conduct could have such large macroeconomic effects. But the Azar, Schmalz, and Tecu study suggests that horizontal shareholdings have lowered output by 6% in at least one industry.78 If generalizable to other industries, which seems plausible given that institutional investors have an even greater share of large corporate stock in other industries, this finding suggests that eliminating horizontal shareholdings could increase economic output by 6%, which would have a huge effect on economic growth and employment levels. Moreover, there is precedent for anticompetitive conduct having these sorts of large macroeconomic effects and for antitrust enforcement to thus have strong macroeconomic benefits. Antitrust enforcement was a key part of what brought the United States out of the Great Depression. To be sure, conventional wisdom is that World War II was responsible for that recovery. But while wartime spending certainly led to expansion in the 1940s, the recovery actually began in 1938 and had cut unemployment in half by 1941, which clearly preceded the United States’s December 1941 entry into World War II.79 Nor can prewar military buildup explain the recovery because average defense spending in 1938 actually dropped 18.5% and continued to be 12% below 1937 levels in 1939 and 9.5% below 1937 levels in 1940. 80 Military stimulus thus cannot explain the recovery that began in 1938 because that recovery actually had to overcome military spending cuts. Others assume that what caused the recovery that began in 1938 was the fact that the United States adopted looser monetary and fiscal policies in 1938. But this theory has two problems. First, statistical analysis shows that, while monetary and fiscal stimulus helped, they cannot explain the full strength of the ensuing recovery.81 Thus, economists have concluded that some factor other than monetary and fiscal policy is needed to explain why the economy “rebounded so strongly” from 1938 to 1941. 82 Second, prices actually declined from 1938 to 1941, with only one short deviation in September 1939, when Hitler’s invasion of Poland led to speculative buying.83 But even then prices remained below 1938 levels and continued their decline after that 1939 spike.84 This downward price trend is precisely the opposite of the price inflation one would expect if monetary and fiscal stimuli were what drove the recovery.85 Increased antitrust enforcement provides a missing factor that can help explain why the 1938–41 recovery was not only so strong, but also lowered prices. Although the Sherman Act was enacted in 1890, until 1938 antitrust enforcement was rare and anticompetitive conduct was common.86 President Theodore Roosevelt made many political speeches about being a trustbuster, but he brought few antitrust cases. Indeed, his entire Antitrust Division had only five lawyers.87 By the time his cousin President Franklin Delano Roosevelt took office in 1933, the Antitrust Division had expanded slowly to 15 lawyers, but that was hardly enough for vigorous enforcement in a nation of over 130 million people.88 Worse, from 1933 to 1938, the Roosevelt Administration fell prey to the natural, but mistaken, tendency to confuse the symptoms of the Depression (low prices and profits) with the disease (low production and employment). To beef up prices and profits, the Administration not only relaxed antitrust enforcement, but in 1933 affirmatively allowed cartels via the National Industrial Recovery Act (NIRA).89 The effect was to significantly raise prices. For example, from April 1933 to June 1934, prices for bituminous coal (which was cartelized under the Act) rose 20%, while prices for anthracite coal (which was not) dropped 7%.90 The NIRA exacerbated the Depression because higher prices meant consumers bought less, which reduced production and thus reduced employment, which in turn reduced the ability of consumers to buy, further reducing production and employment. Economic analysis shows that NIRA cartels lowered investment by 60%, employment by 11%, and output by 13%, causing about 60% of the post-1933 depression in national output.91 Even after the NIRA was held unconstitutional in 1935, 92 the mistaken economic intuition that underlay it continued to produce limited antitrust enforcement.93 That abruptly changed in March 1938, when President Roosevelt appointed Yale Law Professor Thurman Arnold to head the Antitrust Division. Arnold explicitly rejected the notion that antitrust enforcement should be relaxed during an economic downturn.94 He vastly increased antitrust enforcement, expanding the antitrust division to 583 lawyers by 1942. 95 In his five years in office, he brought 44% of all the antitrust cases that had been brought in the first 53 years of the antitrust laws.96 Arnold also made antitrust enforcement far more systematic and focused. Prior enforcement (even before the New Deal) had been not only isolated but also mercurial in a way that often seemed to challenge big businesses just for being big.97 The combination meant little deterrence of anticompetitive conduct both because enforcement was unlikely and also because it was unclear just what firms were supposed to do to avoid enforcement. Arnold made clear that (unlike his predecessors) he had no problem with businesses being big as long as their conduct was efficient and lowered consumer prices.98 This gave firms a far clearer and more desirable signal about how to modify their behavior. Further, Arnold deliberately used antitrust enforcement as a form of economic policy. He targeted industries that he thought were inefficient in a way that hampered economic growth. He also used multiple simultaneous lawsuits in each selected industry to thoroughly restore free competition at each stage of the industrial process.99 His strategy was to “hit hard, hit everyone and hit them all at once.”100 He multiplied the effect of his expansion of prosecutorial resources by using prosecutions to obtain extensive consent decrees designed to go beyond the alleged antitrust violations to make markets as competitive as possible, as quickly as possible, in as many industries as possible.101 Arnold further pursued an aggressive campaign to change the law on cartels, which resulted in the landmark May 1940 decision United States v. Socony-Vacuum Oil Co., 102 argued before the Supreme Court by Thurman Arnold himself, which adopted a sweeping definition of pricefixing and made it per se illegal.103 Socony made antitrust compliance far less dependent on agency enforcement levels, both because the deterrent effect was so clear and because the change in antitrust law spawned a sharp rise in private antitrust litigation. Encouraged by this and other changes in antitrust law, the number of initiated private antitrust cases rose from 8 in 1937 to 110 and 111 in the fiscal years ending June 30, 1940, and 1941.104 Judgments for private antitrust plaintiffs rose in 1947–1951 to an annual rate that was 16-fold the level in 1914–1940. 105 Arnold’s antitrust enforcement successfully lowered prices in the targeted industries.106 Arnold himself stated that his goal was to have macroeconomic effects: lowering prices that were elevated by anticompetitive conduct so that consumers could buy more, which would cause firms to increase production and thus employment, which in turn would increase consumer purchasing power, further increasing production and employment.107 Industrial output dropped 32% from July 1937 to May 1938, but after that began to rise by an average of 22% a year.108 In order to produce more, firms needed to hire more workers. Unemployment, which had risen from 14% in 1937 to 19% in 1938, steadily declined in the years after Arnold’s March 1938 appointment, reaching 10% by 1941. 109 The production turnaround from the low point in May 1938 did start shortly after the April 14–16, 1938, announcements of increased federal spending and looser monetary policy.111 But the production turnaround also followed shortly after Arnold and President Roosevelt gave speeches on April 28–29, 1938, that signaled a sharp coming increase in antitrust enforcement and that would predictably have started to deter anticompetitive behavior.112 Arnold also met with leading industrialists in May 1938 to explain what the new antitrust enforcement regime would look like.113 These signals were quickly confirmed by action because Arnold initiated industrywide antitrust suits with remarkable speed. He was sworn in on March 21, 1938, and by May 18, 1938, Arnold had issued criminal indictments against 86 firms and individuals in the auto industry, prompting all but one of them to begin negotiating settlements within weeks.114 Other industrywide cases followed quickly, including July 1938 cases against the motion picture and dairy industries, an August 1938 case against the medical industry, and in the following months numerous other cases that overhauled the housing, construction, tire, newsprint, steel, potash, sulphur, retail, fertilizer, tobacco, shoe, and various agricultural industries.115 By June 1939, Arnold had “1375 complaints pending in 213 cases involving forty industries with 185 continuing investigations.”116 Arnold’s cases produced quick results not only because he used massive criminal indictments to secure quick industrywide consent decrees, but also because merely launching an antitrust investigation sufficed to drop prices by 18–33% in various industries.117 To be sure, the Roosevelt Administration also pursued various regulatory policies. But what made the new antitrust policy unique was not only its timing, but also the extent to which it was an unexpected about-face. Roosevelt had pursued an anti-antitrust policy through the NIRA until the Supreme Court stopped him in 1935, and in the years thereafter he and his appointees remained, consistent with the policy views that prompted the NIRA, unenthusiastic about antitrust enforcement.118 The main objection raised during Thurman Arnold’s confirmation hearing was that his academic writings indicated he did not believe in antitrust enforcement.119 It was thus a true surprise, contrary to all market expectations, when Roosevelt and Arnold came out so strongly for vigorous antitrust enforcement at the end of April 1938. There was no comparable surprising shift in Roosevelt’s regulatory policies during this period. The closest thing to it was the socalled “Switch in Time That Saved Nine,” when the Supreme Court became more willing to sustain New Deal regulations.120 But that switch occurred in March 1937, 121 and was followed by a sharp drop in industrial output over the next 14 months, which was reversed only after the new antitrust policy was announced. Moreover, Roosevelt’s regulatory policies were more likely to raise prices than lower them, and thus (unlike the shift in antitrust policy) cannot explain the pricereducing nature of the economic recovery that started in May 1938. In short, while increased antitrust enforcement was hardly the sole force bringing the United States out of the Great Depression, the combined evidence indicates that it did play a key role. First, prior economic analysis shows (as mentioned above) that fiscal and monetary stimuli cannot statistically explain the full strength of the 1938–1941 recovery. Second, such stimuli cannot explain at all why that recovery produced lower prices. Third, increased antitrust enforcement directly lowered prices in many industries and would predictably have a deterrence effect that decreased prices in other industries. Fourth, under standard economic principles, such price reductions would increase output. Fifth, Arnold’s antitrust enforcement was affirmatively designed to have macroeconomic effects. Sixth, prior economic analysis shows that the pre-1938 policy of allowing cartels had large macroeconomic effects on output, investment, and employment, which indicates that Arnold’s 1938 reversal of that policy would likely have a similarly strong macroeconomic effect in reverse. Seventh, a conclusion that increased antitrust enforcement was an important driver of macroeconomic growth from 1938–1941 can explain, unlike fiscal and monetary stimuli, why that growth coincided with an economy-wide reduction in prices. Eighth, the economic literature does not indicate any other change in 1938 that could have caused such a strong, price-reducing stimulatory effect. It is unclear whether the effects of antitrust enforcement against horizontal shareholdings would be similarly large today in improving our recovery from the Great Recession. On the one hand, the anticompetitive effects of horizontal shareholdings are generally weaker than the anticompetitive effects of the sorts of cartels that Arnold attacked. On the other hand, given that institutional investors now own 80% of all large corporations’ stock, horizontal shareholdings are more pervasive across our economy now than cartels were in 1938, 122 and while there was some antitrust enforcement against cartels before 1938, current enforcement against horizontal shareholdings by multiple institutional investors is nonexistent.123 Initiating antitrust enforcement against anticompetitive horizontal shareholdings could therefore have stronger or weaker effects than Arnold’s 1938 expansion of antitrust enforcement. Either way, the economic effects of attacking anticompetitive horizontal shareholdings certainly seem salutary and likely to be significant for the national economy.

#### Bizcon low---multiple indicators

Marcos ’9/28 [Coral; September 28, 2021; New York Times, “Stocks tumble in worst day since May, as tech shares slide and bond yields climb,” https://www.nytimes.com/2021/09/28/business/stock-market-today.html]

The prospect of the Federal Reserve not reaching as deep into its bottomless pockets is starting to hit home for investors.

The S&P 500 tumbled 2 percent on Tuesday — the worst one-day slide for the benchmark U.S. index since May — as investors faced the expected wind-down of the enormous bond purchases the central bank has made since the start of the pandemic.

“The deep sell-off highlights the extent of the nerves in the markets surrounding the moves of the Fed,” said Fiona Cincotta, senior financial markets analyst at Forex.com.

The coming slowdown of bond purchases is a sign of the Fed’s confidence that the economy is recovering from the upheaval of the pandemic. But, Ms. Cincotta noted, other factors are still making Wall Street wary.

“There’s also a combination of rising energy prices, concerns that inflation could be more entrenched in these elevated levels and the fact that consumer confidence is slowing,” she said.

The tumble extended into the Asian trading day on Wednesday, though investors signaled that confidence might be returning.

Stocks in Japan were down more than 2.6 percent midday. But losses in other Asian markets, like Hong Kong and mainland China, were more moderate. Futures markets were signaling that Wall Street would open modestly higher.

The trigger for Tuesday’s tumble, which cut across sectors, was a rise in the yield on the benchmark 10-year Treasury note. With the Fed preparing to slow its purchases as soon as November, investors have been selling off bonds before demand ebbs. On Tuesday, that pushed the 10-year’s yield up to 1.54 percent, its highest level since June.

Even though the Fed has said it doesn’t plan to increase interest rates for months or years, government bond yields are the basis for borrowing costs across the economy. When bond prices fall, yields rise — a move that can hinder the stock market’s performance because it makes owning bonds more attractive and can discourage riskier investments.

Higher rates would make borrowing more expensive for smaller companies, and the jump in yields was a blow to shares of several high-flying stocks. Etsy, the online craft marketplace, dropped 6 percent, and Shopify fell more than 5 percent. Both companies have soared during the pandemic.

“With tech stocks, you’re betting for a company to have a breakthrough years from now,” said Beth Ann Bovino, the chief U.S. economist at S&P Global. “If interest rates go up today, that value that you receive years from now is discounted.”

The biggest technology stocks — particularly Amazon, Apple, Microsoft, Google and Facebook — have a vast pull on the broader market and helped drag down the S&P 500. Apple fell 2.4 percent and was the best performer of the tech giants. Amazon dropped 2.6 percent while Microsoft, Facebook and Google were down by more than 3.5 percent.

But the declines cut across many sectors. Energy stocks were the exception, rallying after oil prices climbed early in the day. Schlumberger, ConocoPhillips, Halliburton and Exxon Mobil were among the best-performing shares in the S&P 500, though some of their gains faded as oil futures turned lower in the afternoon.

The Delta variant of the virus remains a concern for investors, while persistent supply-chain bottlenecks have affected everything from auto production to school lunches. In Washington, lawmakers remain deeply divided over spending on infrastructure and expanding social programs.

## CP---Core

### 2AC---CP---Core

#### Limiting forced arbitration necessarily ‘expands scope’ of ‘core antitrust laws’ because arbitration restricts their scope.

Rubinoff ’20 [Matt; May 1; J.D. Candidate at Pennsylvania State University, Managing Editor; Arbitration Law Review, “Too Big to Arbitrate? Class Action Waivers, Adhesive Arbitration, and Their Effects on Antitrust Litigation,” vol. 12]

Following Concepcion, the Supreme Court continued to support the exclusion of class action relief through adhesive class action waivers.31 This time, the controversy pitted federal arbitration laws against federal antitrust laws.32 In American Express Co. v. Italian Colors Restaurant (“Italian Colors”), the plaintiff merchants entered into an agreement with a subsidiary of American Express. 33 The contract included a class action waiver.34

Nonetheless, following a dispute, the plaintiffs filed a class action claiming defendant American Express violated Section 1 of the Sherman Antitrust Act and portions of the Clayton Act. 35 As part of their antitrust claim, the plaintiffs sought the applicable treble damages.36 American Express moved to compel arbitration.37 In response, the plaintiffs argued that the costs of individual arbitration greatly exceeded their total potential recoveries. 38

After the district court granted the motion to compel arbitration and dismissed the lawsuits, the Second Circuit reversed, holding the class action waiver unenforceable because of the unbalanced costs directed onto the plaintiffs.39 However, once again led by a Justice Scalia majority, the Supreme Court reversed.40 The Court held that the FAA does not permit courts to invalidate a class arbitration waiver solely because of the costs involved in the process exceed the total potential recovery.41 Moreover, the Court reasoned that the existing antitrust laws do provide any intention to preclude class action waivers such as the one in the plaintiffs’ case.42 For better or for worse, in the eyes of the Supreme Court, the FAA now not only ruled over state law and state courts, but also pre-empted otherwise legitimate recovery under federal antitrust laws.

#### Removing barriers to private antitrust litigation is crucial to overall deterrence of anticompetitive conduct---particularly for cartels---empirical data proves.

AAI ’21 [American Antitrust Institute; August 4; Washington, D.C.-based non-profit education, research, and advocacy organization, citing research conducted by Professor John Davis at the University of San Francisco School of Law; American Antitrust Institute, “The Critical Role of Private Antitrust Enforcement in the United States,” <https://www.antitrustinstitute.org/wp-content/uploads/2021/08/Huntington-Report-FINAL-1.pdf>]

I. Introduction

With all of the attention currently focused on public enforcement and legislative reform of the antitrust laws, less attention is being paid to private enforcement.1 But Congress considered private antitrust enforcement indispensable for promoting competition. The judiciary has so recognized time and time again. In California v. American Stores Co., for example, the Supreme Court proclaimed, “Private enforcement of the [Clayton] Act was in no sense an afterthought; it was an integral part of the congressional plan for protecting competition.”2

Private enforcement is not a substitute for vigorous public enforcement. Both are necessary to foster competition. But private enforcement plays an important part, one that becomes more significant when public enforcement recedes.3 And, unlike public enforcers, private enforcers can obtain significant damages on behalf of the victims of antitrust violations.4 This serves as a crucial source of deterrence for illegal anticompetitive conduct and the primary means of compensating victims for harms suffered at the hands of cartelists and dominant firms.5 The importance of the antitrust class action, a major private enforcement device, is clear. The recently released 2020 Antitrust Annual Report: Class Action Filings in Federal Court (“2020 Report”)6 by Huntington National Bank and the University of San Francisco School of Law (“USF Law”) reflects that the cumulative total settlement amount recovered for victims in antitrust class actions from 2009-2020 was over $27 billion.7

Antitrust class actions recover damages from companies engaged in harmful, illegal conduct, such as price fixing and attempted monopolization, in markets for important and essential products and services. The most active defendants during the period, for example, included companies providing financial services, pharmaceuticals, automobile parts, and electronics parts. 8 In light of the vital role played by private antitrust enforcement, and the antitrust class action in particular, continued empirical analysis of trends in activity is essential. This analysis aids in understanding and evaluating proposals for reforming the antitrust laws in the U.S. and such proposals’ impact on private enforcement, the public-private partnership, and ultimately on competition and consumers.

II. Overview of the Commentary

The American Antitrust Institute (AAI)9 and Professor Joshua P. Davis at USF Law10 evaluated the 2020 Report with the goal of identifying its major implications for private enforcement in the U.S. The 2020 Report builds and expands on the 2019 Antitrust Annual Report: Class Action Filings in Federal Court (“2019 Report”)11, which assessed private enforcement activity from 2009-2019. Like the 2019 Report, the 2020 Report relies largely on data for private U.S. antitrust class actions available through Lex Machina, as well as supplemental data analysis.12 The 2020 Report extends the dataset to the eleven-year period covering 2009-2020, thus allowing for a deeper analysis of private enforcement trends and their implications. The analysis provided in this Commentary highlights the importance of private antitrust enforcement in the U.S. system and the particularly important role played by the antitrust class action.

III. Observations and Implications for Private Enforcement

The 2020 Report provides further evidence of a divergence between public and private enforcement trends. As public enforcement has waned, private filings have waxed, undermining the notion that class actions simply ride the coattails of public enforcement. On the contrary, the data suggest that as lax public enforcement fosters higher market concentration and invites bad behavior, private filings may compensate for underenforcement in an effort to address the resulting antitrust violations.

Looking beyond the number of enforcement actions filed and focusing on the results of the actions, the rich data reveal more nuance to this narrative. Despite increased private actions in the face of decreased public enforcement, the amount of money recovered from violators by both public and private enforcers has diminished. For public enforcers, this diminution is to be expected, as fewer cases have been brought. For private enforcers, though, the explanation likely lies with other trends, most notably the increasing headwinds faced by private enforcers due to heightened pleading and class certification standards. If these explanations are correct, the clear implication is that for private enforcement to fulfill its increasingly vital role as a complement and a backstop to public enforcement, these trends must be reversed.

A theme we noted in the 2019 Report, and that continues to feature in the 2020 Report, is the tremendous variability in the data on some measures related to settlement size. Aggregate settlement amounts over the period vary widely from year to year. By disaggregating the settlements by size, however, we are able to observe that settlements at different levels trend somewhat independently. Very large settlements, which are few in number, drive most of the variability in the aggregate data. But trends and anomalies in very small settlements cannot be entirely discounted, as they are the force behind one of the highest recovery years in the period, 2018.

Finally, building on analysis from our 2019 commentary, we take a deeper dive on attorneys’ fees and how they correspond to settlement amounts. Our findings reinforce the tentative conclusion from last year’s analysis that the so-called “megafund doctrine”—a dramatic decrease in attorneys fee percentages on settlements above a threshold of about $100 million—does not operate in federal antitrust cases in a significant way. Rather, decreases in the fee award percentage in antitrust cases are not discrete and drastic, but rather gradual, much like marginal tax rates in the United States. In what follows, we discuss each of the above observations in more detail and provide analysis of their implications for private enforcement, many of which suggest fertile areas for additional study.

A. The Relationship of Private to Public Enforcement: Riding on Coattails or Stepping into the Breach?

A long-running antitrust policy debate centers on the value that private enforcement adds to the antitrust enterprise as a whole. Critics of private actions maintain that private plaintiffs often follow an easy trail blazed by government enforcers, and that private enforcement therefore does not supplement worthwhile public actions as much as it should. Proponents of private actions have sought to debunk this claim with empirical evidence suggesting that many large and successful private antitrust cases often precede, or else expand the scope of relief sought in, any overlapping government actions.13

Data from the 2020 Report, together with future reports, may warrant a reexamination of this debate through a different lens. Apart from the question of the extent to which private enforcement serves as a useful complement to public enforcement, it may also be worth inquiring into the extent to which private enforcement serves as a substitute for public enforcement, particularly during periods of government forbearance.

As AAI noted in an independent report, “The State of Antitrust and Competition Policy in the U.S.,” several indicators had suggested a decline in government cartel enforcement at the midway point of the Trump Administration.14 Perhaps most notably, the average number of cartel investigations opened during the period from 2017-2018 was 80% lower than the annual average number of investigations opened over the previous three administrations (1993-2016). In addition, the average number of corporations fined by the Trump agencies in 2017 and 2018 fell by about 45% relative to the Obama Administration.

Notably, the 2020 Report shows that private federal consolidated antitrust filings rose significantly during each year of the Trump Administration. First, they increased from 74 in 2017 to 136 in 2018. But in 2019, they rose dramatically to a ten-year-high of 211. And, in 2020, that number was eclipsed, with 220 filings. In other words, a dramatic increase in private filings appears to have occurred immediately subsequent to a substantial decline in the opening of government cartel investigations and the number of corporations fined.

## CP---Sunbursting

### 2AC---CP---Sunbursting

#### They view it as constitutionally grounded and will tear apart legislation---arbitration law is an exclusive judicial domain.

Carbonneau ’20 [Thomas; November 18; Samuel P. Orlando Distinguished Professor of Law at Pennsylvania State University, Faculty Director of the Arbitration Institute; Arbitration Law in a Nutshell, “The Central Themes of American Arbitration Law,” Ch. 2]

These criticisms are simultaneously powerful and perplexing. They make a strong case against the current usage and status of arbitration. The strength of the criticism resides primarily in the link it establishes to the traditional: what is familiar and what has always been done. In rejecting arbitration, the tribunals in coastal liberal enclaves ironically seek to conserve the existing order and defend established practices and mechanisms. The fear that private adjudication induces and which leads liberal courts to embrace conservative values is, in fact, an accurate reaction to and assessment of contemporary arbitration. The fear is entirely justified because the ‘emphatic federal policy’ betokens a revolution. It is a revolution in which the Court is both an instigator and a perpetrator. The judicial endorsement of arbitration represents a fundamental change in the values of legal civilization and the role these values play in American society.

The Court was the first U.S. public institution to understand the need for change in civil litigation and how it was to be achieved. The understanding took place over several decades and numerous rulings. At the outset, the policy objective was timid, careful, and hesitant. As the value of the decisional law grew, the Court’s doctrinal leadership became more confident and focused. Toward the end of the persistent evolution, new concepts and revamped paradigms emerged which challenged the fundamental givens in the previous order. The Court saw and promoted arbitration as an effective and lawful process for the adjudication of disputes that could remedy the legal system’s inability to offer workable civil litigation. The redefinition of arbitration revealed the dysfunctionality and exposed the contradictions of adversarial litigation, along with the consequence of the disabling applications of excessive procedural due process. In attempting to offer rigorously constitutional civil litigation, American law imposed a counterproductive, unconstitutional burden on American society. It denied most citizens their right to the redress of their civil grievances. For the Court, arbitration resolved this problem without placing greater demands on the public fisc or the court system. It is entirely accurate to see (and impossible to deny) that the U.S. Supreme Court’s decisional law on arbitration constitutes judicial legislation. The Court’s opinions have substantially altered and added to the content of the enacted statute. There appears to be a tacit understanding between the Court and the U.S. Congress that arbitration falls within the Court’s exclusive bailiwick.

## DA---Sua Sponte

### 2AC---DA---Sua Sponte

#### Inevitable and empirically false

Bhagwat ‘2k[Ashutosh; October; Professor of law at the University of California; Buffalo Law Review, 80 B.U.L. Rev. 967, lexis]

Even after it has decided that it should resolve an issue, the Court may choose to pass up cases presenting the issue until a case with suitable facts is presented. When the Court finally grants certiorari in a case and then decides it, the Court's opinions on the merits tend to focus almost exclusively on abstract doctrinal and policy issues, with essentially no discussion of the facts and equities of the particular case in front if it. Indeed, even after it has crafted a new legal standard, the modern Court has displayed an increasing tendency to remand the case to a lower court to apply that standard, rather than to fully resolve the dispute before it as most courts would do as a matter of course. n154 Finally, Erwin Chemerinsky has noted a recent trend on the Court of reaching out to decide issues which were not presented in the petition for certiorari, and were not briefed or argued by the parties. n155 All of these are well-accepted and today largely uncontroversial aspects of the Court's decision-making process. But taken together, they emphasize just how far the functions of the modern Supreme Court have drifted from the traditional judicial role of adversarial dispute resolution, as well as the general acceptance of this fact. n156

#### There’s always a test case

Pacelle ‘2 [Richard; 2002; Professor of Political Science at Missouri-St.-Louis; *The Role of the Supreme Court in American Politics: The Least Dangerous Branch?,* p. 122]

The passivity of the Supreme Court is an overrated limitation. Thousands of cases are on the annual docket, and virtually every issue the justices might want to consider is available. The justices have the ability to signal litigants to bring certain types of cases. There are a number of so-called repeat players, long-term actors in the judicial process, who bring cases, respond to the directives and cues in the Court’s decisions, and bring the next round of litigation (Galanter 1974). In addition, justices have been known to manipulate the issues in a case to make them more amenable to the type of issues they seek to address. Justices can “add” issues to a case or change the ones that are brought to them (Pacelle 1991, 32—33).

#### No judicial legitimacy impact

Feldman 20 – Distinguished professor of law and adjunct professor of political science at the University of Wyoming.

Stephen M. Feldman, “Court-Packing Time? Supreme Court Legitimacy and Positivity Theory,” *Buffalo Law Review*, vol. 68, no. 5, December 2020, pp. 1548-1552, <https://digitalcommons.law.buffalo.edu/cgi/viewcontent.cgi?article=4892&context=buffalolawreview>.

In fact, political science studies suggest that the public’s diffuse support for the Court is resilient, sustained by “a reservoir of favorable attitudes or good will.”87 A “positivity bias” helps the Court maintain this good will and institutional legitimacy. According to positivity (bias) theory, “anything that causes people to pay attention to courts— even controversies— winds up reinforcing institutional legitimacy through exposure to the legitimizing symbols associated with law and courts.”88 Even when the Court issues a decision contrary to an individual’s personal views, that individual is unlikely to lose faith in the Court. If anything, when news of Court activities draws an individual’s attention, then that attention (to the Court) will likely reinforce the individual’s positive views of the institution. In a sense, the more one knows about the Court, the more one is likely to find its decisions legitimate (the opposite is true for Congress).89

To be sure, the Court’s legitimacy is not bulletproof: It depends on a perception that the Court is not merely another political institution. For instance, a confirmation battle in the Senate is unlikely to damage the Court’s legitimacy, but if widely viewed advertisements (related to the confirmation battle) attack the Court as purely political, then diffuse support for the Court is likely to diminish.90 Thus, while a politically salient Supreme Court decision might offend some Americans based on political ideology,91 a lack of specific support for that decision does not translate into a meaningful reduction of diffuse support. Only those Americans who already reject the Court as an institution—those individuals who have not developed a favorable attitude and good will toward the Court—are likely to denigrate it because of a small number of specific decisions. For the most part, the Court is able to maintain its institutional legitimacy despite “the ideological and partisan cross-currents that so wrack contemporary American politics.”92 Even so, sustained disappointment with the Court’s decisions over the long term, especially in politically salient cases, can weaken diffuse support for the Court. To take one example, diffuse support for the Court diminished among black Americans during the post-Warren Court years (consider the Burger, Rehnquist, and Roberts Courts’ consistent hostility toward race-based affirmative action).93

Significantly, the people’s diffuse support for and loyalty to the Court does not depend on the myth of pure law—that is, the myth of the law-politics dichotomy. To the contrary, many Americans seem to understand that Supreme Court decision making entails a combination of law and politics— the law-politics dynamic. As Gibson and Caldeira conclude: “[T]he American people seem to accept that judicial decisionmaking (sic) can be discretionary and grounded in ideologies, but also principled and sincere. Judges differ from ordinary politicians in acting sincerely . . . .”94 This insight into the Court’s institutional legitimacy has enormous implications for Democratic court-packing. Although a court-packing controversy would undoubtedly entail debates over the Court’s politically-charged decisions, the Court’s overall diffuse support would probably remain relatively stable. Most likely, in these hyper-polarized times, individuals’ political ideologies—leaning Republican or Democratic— would influence reactions to a Democratic court-packing plan. Republicans of course would oppose it, but many Democrats would likely support it, especially if Democratic politicians emphasized that they sought to return the Court to sincere and principled decision making.95 To the extent that individual views of the Court’s legitimacy might change in response to a court-packing plan, partisan shifts would likely cancel each other out. In the end, despite divergent views of the court-packing plan, the overall legitimacy of the Court itself would likely be sustained (or even grow) whether because of a positivity bias favoring the Court or a widespread Democratic (policy) opposition to the Roberts Court’s conservatism (as well as Democratic abhorrence toward recent Republican Senate maneuvers, including the rushed confirmation of Barrett, which resulted in an ironclad six-justice conservative bloc).96

## DA---PTX

### 2AC---DA---PTX

#### Courts shield.

Mazzone ’18 [Jason; August 9; Professor of Law at the University of Illinois at Urbana-Champaign; Chicago-Kent Law Review, “Above Politics: Congress and the Supreme Court in 2017,” [vol.](https://scholarship.kentlaw.iit.edu/cgi/viewcontent.cgi?article=4207&context=cklawreview) 93]

Absent, too, in the modern Congress is any real sense that the Supreme Court can be brought to heel: say, by constitutional amendment, by stripping the Court of funding, by hauling in members of the Court to justify their rulings before congressional investigatory committees, by appointing special counsels to review and report back on what the Court does, by impeaching the Justices (or locking them up), or by simply ignoring or defying judicial rulings. Perhaps the Court does not rule in ways that offend enough members of Congress (or their constituents) for them to invest the energy—and political capital—required to generate these sorts of measures. Perhaps, instead, members of Congress do not consider such measures appropriate in our constitutional system. In either case, modesty on the part of Congress is the result, even in an era when a single party controls both the Congress and the White House. The lesson for the Court is that so long as it continues doing—more or less—what is has done in recent years, it has very little to fear from the Congress.

Conclusion

After President Trump nominated Neil Gorsuch to fill the vacancy on the Supreme Court left by the death of Justice Scalia, fifteen House Republicans sponsored a Resolution that “the House firmly supports the nomination of Neil Gorsuch to the Supreme Court” and “the Senate should hold a swift confirmation of this nomination.”229 The proposed resolution died, without further action, in the Committee on the Judiciary. While Gorsuch was, of course, confirmed, the failure of the Republican-controlled House to pass a simple resolution supporting the nomination is telling. After an election season in which the Supreme Court figured very prominently, aside from the Senate’s confirmation of a new Justice, Congress in 2017 accomplished nothing with respect to the Supreme Court. Various bills and resolutions—some sponsored by Republicans, others by Democrats, and some garnering bipartisan support—targeted statutory and constitutional rulings by the Court and sought also to impose new regulations upon the Court’s activities. Even the most modest of these proposals failed to advance through the legislative process and become law. We like to think that the Supreme Court, guided solely by the rule of law, is above politics. The experience of 2017 suggests that the Court may also be above politics in the quite different sense that its rulings and activities are largely immune to political response and redress.

#### Plan bipartisan.

John ’21 [Basil; July 14; Writer; KLFY News, “US lawmakers announce bipartisan push to end forced arbitration in the workplace,” <https://www.klfy.com/national/us-lawmakers-announce-bipartisan-push-to-end-forced-arbitration-in-the-workplace/>]

United States lawmakers on both sides of the aisle say workers should not be forced to stay quiet about bad behavior.

“End the days of institutional protections for harassers,” Sen. Kirsten Gillibrand (D-N.Y.) said.

On Wednesday, Gillibrand announced bipartisan legislation to end forced arbitration over cases of sexual harassment and assault in the workplace.

“Removing those provisions would give survivors their day in court,” Gillibrand said.

The bill eliminates contract clauses that prevent survivors from seeking justice and public accountability.

“Forced arbitration is unfair because it creates an uneven playing field,” Sen. Lindsey Graham (R-S.C.) said.

Graham says employees shouldn’t have to leave their rights at the door.

“To go to work, you can’t sign away your dignity,” Graham said.

The legislation has already received bipartisan backing in both the Senate as well as the House.

#### Class action litigation solves warming

Juhn ’21 [Mina; 2021; J.D. Candidate at the Fordham University School of Law, B.A. in 2013 from Wellesley College; Fordham Law Review, “Taking a Stand: Climate Change Litigants and the Viability of Constitutional Claims,” vol. 89]

Introduction

As the impacts of global warming on human and natural systems increase, so too do the number of lawsuits brought to mitigate them.1 There is widespread scientific consensus that global warming is changing the global climate in a manner deleterious to both human and natural systems.2 Eighteen of the nineteen warmest years on record have occurred since 2001, and even the staunchest climate change skeptics recognize the dynamics of the carbon cycle.3 In 2018, the Intergovernmental Panel on Climate Change (IPCC) issued a report stating that if greenhouse gas emissions continued at their current rate, the resulting temperature increases could destroy coral reefs, exacerbate wildfires, jeopardize food supplies, and contribute to political instability in developing nations.4 Climate change is “one of the key challenges of our lifetimes and future generations,”5 and there is significant evidence linking extreme weather events and the onset of climate change to human activity.6

The velocity and permanency of global warming have become topics of increasing urgency in recent decades7 but have been met in the United States by legislative stagnation and agency deregulation.8 Congress’s failure in the last thirty years to enact any major legislation to regulate greenhouse gas emissions has coincided with a sharp increase in climate lawsuits.9 This perceived inaction from the political branches has compelled some plaintiffs to turn to the judicial system to combat anthropogenic climate change via a growing class of environmental lawsuits known as “climate change litigation.”10 These litigants do not purport to supplant the political process, and there remains a general consensus that the political branches should address climate change via legislative or executive action.11 Yet, in the absence of a large-scale or effective response to the growing climate crisis, plaintiffs have attempted to compel regulatory and political action through the courts.12

Climate change litigation encompasses many types of lawsuits that seek to hold certain actors accountable for their contributions to or failures to act on climate change.13 Plaintiffs may bring litigation against federal and state governments, city administrations, and corporations; the most common claimants are environmental organizations, industry trade groups, local governments, and citizen groups.14

In the United States, climate change litigation has traditionally involved statutory claims or claims sounding in common-law tort doctrines.15 In contrast to administrative or common-law claims, rights-based climate change lawsuits against governments and public authorities have recently gained traction.16 These public law actions seek to compel federal, state, and local governments to escalate their efforts to address climate change by raising human rights and constitutional arguments.17

To date, the most prominent climate change lawsuit18 of this type— noteworthy both for the scope of its constitutional claims and the youth of its plaintiffs—is Juliana v. United States.19 The twenty-one young plaintiffs allege that, despite the federal government’s obligation to reduce carbon emissions, the government has actively facilitated the country’s increased carbon emissions by supporting the fossil fuel industry.20 The Juliana plaintiffs argue that the government’s actions have violated their “fundamental constitutional rights to life, liberty, and property.”21

In 2016, the case generated extensive media coverage when Judge Ann Aiken of the District Court of Oregon declared that “the right to a climate system capable of sustaining human life is fundamental to a free and ordered society.”22 In January 2020, a panel of the Ninth Circuit reversed the lower court’s decision for lack of standing.23 Although the Ninth Circuit acknowledged that federal policies contributed to sustaining high levels of carbon emissions, the court decided that carbon emissions policies presented political questions that required resolution by the political branches.24 As a result, the panel held that the plaintiffs did not satisfy the requirements for Article III standing and reversed the lower court’s decision.25

The Ninth Circuit’s decision is emblematic of the judiciary’s typical response to climate change litigation, particularly to the cases in which plaintiffs advance constitutional or rights-based claims.26 Often, the judiciary finds that policymaking discretion, class certification, or separation of powers concerns restrict their ability to grant relief and routinely dismisses suits for lack of standing.27

This Note examines federal climate change lawsuits in the United States alleging constitutional violations. Part I begins with a brief overview of statutory and common-law claims in climate change litigation and proceeds to a discussion of the legal rights and issues implicated by constitutional claims. Part I also introduces Juliana as a case study for evaluating the legal strategies and justiciability issues involved in rights-based arguments. Part II reviews the arguments for and against pursuing constitutional claims in climate change litigation, focusing on the procedural and substantive issues raised. Part III argues that constitutional claims in climate change litigation should be granted standing and that they merit judicial review.

I. Overview of Climate Change Litigation

Climate change litigation is unified in its focus on the harmful impacts of climate change, but plaintiffs have utilized an array of approaches, which has resulted in a range of litigation strategies.28 Climate change litigation extends beyond the traditional confines of environmental litigation—which focuses largely on air and water pollution, the preservation of endangered species, and environmental impact statements29—to cases that argue for a constitutional right to a stable climate and the rights of future generations.30

This part establishes several necessary legal foundations. Part I.A reviews federal statutory and common-law causes of action. Part I.B discusses the sources of rights for constitutional claims and the justiciability issues involved. Part I.C introduces Juliana—this Note’s central case study of a constitutional climate change lawsuit—and reviews its procedural posture and legal arguments.

A. Traditional Sources of Legal Rights

The legal bases for climate change lawsuits include statutory, common-law, and constitutional causes of action.31 Statutory claims are the most common, followed by claims sounding in common-law tort doctrines, in particular public nuisance, private nuisance, and negligence.32 Constitutional claims comprise a comparably smaller but growing subset of climate change litigation.33 This Note focuses primarily on the strength and generativity of constitutional claims34 brought by environmentalist plaintiffs but begins with an instructive overview of the other categories of climate litigation.

1. Statutory Authority

A significant number of climate change lawsuits allege violations of federal statutes and regulations, including the Administrative Procedure Act,35 the National Environmental Policy Act,36 the Clean Air Act of 196337 (CAA), the Clean Water Act of 1977,38 and the Endangered Species Act of 197339 Actions against private entities have also arisen under securities regulations or as requests under the Freedom of Information Act.40 According to an analysis of the Sabin Center for Climate Change Law’s climate change litigation database, which tracks lawsuits brought by environmental and citizen groups between 1990 and 2016, 65 percent of cases had a federal statutory cause of action.41

The most noteworthy climate lawsuits alleging statutory violations and challenging government efforts to regulate carbon emissions have been brought under the CAA. The U.S. Supreme Court’s seminal decision in Massachusetts v. EPA 42 in 2007 arguably instigated the current proliferation of climate change lawsuits and is the preeminent example of an environmental lawsuit seeking to compel government action under a statutory scheme.43 In Massachusetts, several U.S. states, cities, and environmental groups challenged the Environmental Protection Agency’s (EPA) denial of a rulemaking petition to regulate greenhouse gas emissions from motor vehicles.44 The plaintiffs alleged that the government’s inaction would result in specific climate change–induced harms—including serious adverse effects on human health—and sought to compel the government to regulate carbon dioxide emissions pursuant to the CAA.45 In response, the EPA argued that greenhouse gas emissions did not qualify as air pollutants and that the plaintiffs had not demonstrated a sufficient and particularized harm required to establish standing.46 The EPA further argued that even if it had authority to regulate greenhouse gases, it would be unwise to do so at that time.47

The Court found that greenhouse gas emissions qualified as air pollutants and consequently were subject to regulation by the EPA.48 The Court further held that the EPA’s refusal to regulate emissions constituted an actual and imminent harm and that the EPA was authorized and obligated under the CAA to regulate greenhouse gas emissions if the EPA determined that such emissions endanger public health and welfare.49

Writing in dissent, Chief Justice Roberts maintained that the claims constituted a nonjusticiable question and disputed the majority’s finding that the plaintiffs had demonstrated sufficient causation and redressability.50 Chief Justice Roberts also questioned whether the redress sought would alleviate a global problem like climate change, noting that approximately 80 percent of global emissions originate outside the United States.51 Justice Antonin Scalia, in a separate dissent, criticized the majority for depriving the EPA of Chevron deference in favor of its own policy determinations.52

In addition to causes of action brought under environmental statutes, plaintiffs have also brought claims under federal securities laws.53 These cases often allege that directors of oil and gas corporations violated their fiduciary duties to shareholders by misleading them as to the impact of the companies’ carbon output on global warming.54 For example, in Ramirez v. Exxon Mobil Corp.,55 a group of shareholders brought a class action suit under the Private Securities Litigation Reform Act of 1995,56 alleging that Exxon Mobil had defrauded investors by purposefully misrepresenting climate risks.57 In 2018, a Texas federal district court found that the shareholders had adequately pleaded the alleged misstatements and met the standard for a securities fraud claim.58 Current and former employees of Exxon Mobil advanced similar arguments in Fentress v. Exxon Mobil Corp.,59 in which the employees sued the company for violations of the Employee Retirement Income Security Act of 197460 (ERISA). The plaintiffs alleged that Exxon Mobil’s failure to disclose and consider the impact of environmental risks and its materially false statements about the health of the company constituted a breach of its fiduciary duties.61 While the court did not question the elements of climate change argued by the plaintiffs, citing Massachusetts, it found that the plaintiffs failed to meet the pleading standard for ERISA claims and dismissed the case.62 Plaintiffs in climate change litigation have had inconsistent degrees of procedural success when bringing claims under other federal statutes.

## K---Cap

### 2AC---Perm

#### Perm do both – regulated capitalism best. Plan challenges market concentration and consumer welfare standard. Blanket critique totalizes.

Smith, PhD, 19

(Noah, https://www.bloomberg.com/opinion/articles/2019-03-08/letting-16-year-olds-vote-is-a-good-idea)

Depending on who you ask, the term "neoliberal" can apply to anyone from Ronald Reagan to Barack Obama. Some on social media have turned the term into a running joke, holding ironic Twitter polls to see who is the “chief neoliberal shill” (the winner last year was none other than yours truly). But at least one economist has articulated a coherent vision of neoliberalism -- Brad DeLong, a professor at the University of California-Berkeley who worked at the Treasury Department during the Bill Clinton administration. In 1999, DeLong wrote that a combination of market liberalization in developing countries and trade opening by rich nations would allow the poor countries of the world to end centuries of poverty. The plan seems to have worked. Market liberalization in countries such as India and China seems to have precipitated a shift to faster growth, while trade and investment links with rich countries have helped these and other developing countries tremendously: These changes helped pull a billion people out of desperate poverty, and billions more are on the way to becoming middle class. But there was a big hole in DeLong’s neoliberal plan. While the developing world surged forward, the U.S. began to encounter a host of economic problems. Wage stagnation, reduced mobility and rising inequality eroded the foundations of the New Deal society that had sustained the U.S.'s middle class during the second half of the 20th century. The U.S. resisted nationalizing its health-care system, resulting in a cumbersome public-private hybrid arrangement that allowed costs to mushroom while letting some people go uninsured. And financial deregulation led to a crisis and a huge, long recession throughout much of the developed world. Now, DeLong is ready to throw in the towel. In a recent interview, he declared that left-leaning advocates of neoliberal policies in the U.S. were mistaken in thinking they would find a political partner on the center-right. The plan was always to cushion the blow of international trade and easing of regulations on business using government programs, such as universal health care and a robust social safety net, to make sure the working class wasn’t left behind. But, DeLong argues, Republicans rejected that compromise, insisting that any neoliberalism be of the free-market-fundamentalist variety: Barack Obama rolls into office with Mitt Romney’s health care policy, with John McCain’s climate policy, with Bill Clinton’s tax policy…[but] John Boehner, Paul Ryan, and Mitch McConnell [were] the leaders of the Republican Party, and…decided on scorched earth[.] As a result, DeLong declared that old-line neoliberals need to pass the baton to the political left. Others aren’t ready to let DeLong off so easily. In the Boston Review, a panel of economists writes that neoliberalism got the policy wrong as well as the politics. Their various suggestions for post-neoliberal policies include increasing labor’s power with greater unionization and wage boards, tighter regulation of the finance industry and restriction on trade in order to protect U.S. workers. Mike Konczal of the Roosevelt Institute echoes their assessment. Many of these are good ideas. But in rejecting neoliberalism as a concept, the critics go too far. First, progress in the developing world has been impressive -- something for which neoliberalism probably deserves a lot of credit -- but it is far from complete; most of South Asia is still very poor, and much of Africa is just beginning to industrialize. To curb the flows of trade and investment with these countries would be a grave abdication of the U.S.’s international and humanitarian responsibilities. Second, neoliberal policies might have led to faster productivity growth in the 1990s and early 2000s: Tech Boom or Something More? Total factor productivity\* Source: Federal Reserve Bank of St. Louis \* Index 2011 = 1 Contrary to popular belief, wages also increased during that period. The spurt of growth is commonly attributed to the information-technology boom, but that boom might not have been possible if the U.S. had more strictly regulated emerging industries in order to protect favored incumbents. It’s worth noting that West Europe and Japan, whose policies were somewhat less neoliberal than the U.S.’s, ended up producing relatively few big new tech companies, and have failed to catch up to U.S. levels of per capita income in the years since 1990. Finally, although economic blunders have come from the political right in the U.S. in recent decades, it’s also possible for the left to make big mistakes -- not just in poor countries, but in rich ones too. Germany suffered high unemployment in the 1980s and 1990s, thanks to its rigid labor market regulations; eventually, it eased those restrictions, which substantially lowered the unemployment rate. Sweden had a very progressive tax system, but scaled back redistribution in the 1990s in order to speed growth. France, too, has sometimes been forced to curb its ambitions for redistribution and regulation when these produced economic instability and slow growth. The U.S. needs a neoliberal contingent to help insure against missteps like these. So neoliberals’ ideas are still needed. A move toward social democracy should help correct much of the inequality that has arisen in the U.S., while fixing dysfunctional industries like health care and finance. But left-leaning neoliberals like DeLong will still be needed in order to restrain social democrats’ more ambitious impulses, to protect the U.S. economy’s entrepreneurial private sector, and to make sure that technological progress and international trade don’t get forgotten.

#### Pursuit of antimarket purity dooms alternative to irrelevancy – alienates potential allies and assumes non-market economics wouldn’t oppress.

Julie **NELSON** Global Development and Environment @ Tufts **‘6** *Economics for Humans* p. 37-40

Problems with the Market-Critic Prescriptions

At the end of the last chapter, I brought up evidence of poverty and corporate abuses that raise questions about the adequacy of the probusiness, free-market prescription for curing social ills. Do the prescriptions of the market critics for “small is beautiful/’“government to the rescue,” or “separate spheres” solutions give us grounds for more hope?

The “small is beautiful” prescription contains, of course, some truth. It is true that acting ethically is a more complicated process the larger and more complex the level of organization involved. Likewise, the “government to the rescue” advocates make some good points. It is easier for any one company to do the right thing if there is public pressure on all companies to do the right thing, and a government regulation can be a good tool for applying such pressure. On an even larger scale, international public agreements may be the only hope for addressing global climate change issues. These are far too big for any one nation, let alone one company, to take on. And there is some truth in the “separate spheres” view. There are some social welfare problems for which private, market solutions don’t work. Care for people who are poor and ill or otherwise needy cannot be provided on a purely market basis. The funds have to come from somewhere other than the “consumers” of the services. Public or private nonprofit allocations of money are necessary.

But while the values held in high regard by market critics are praiseworthy, and the prescriptions contain partial truths, I find the prescribed solutions lacking when held up to criteria of realism and effectiveness. Sometimes the proposed solutions could cause real damage.

A first problem is that these views tend to assume not only that the market sphere is driven exclusively by self-interest, but that self-interest is exclusive to the market sphere. They often seem to assume that if an organization is small, or nonprofit, or governmental, then non-self-interested motivations can be trusted to take over. We should consider the evidence on this.

Families, for example, are very small nonprofit organizations, presumably governed by interests of love and intimacy (as in the Victorian image).The newspaper reminds us daily, however, that families can also be characterized by domination and abuse, even violence. Sometimes being in a small-scale organization just means being under the thumb of a small-scale oppressor.

Community organizing is a great way to bring a group together to work on issues of social concern and to create opportunities for activism. Community organizing was very effective in South Boston in the 1970s, for instance, when big community demonstrations were organized to fight racial integration of the local public schools. Sometimes community groups carry out agendas of racism. And it is not uncommon for community activists motivated by not-in-my-backyard sentiments to try to push undesirable projects off on some other community. Communities, like individuals, can act in purely self-interested ways.

Nonprofit and religious organizations can bring people together to work for goals other than profit.The Boston diocese of the Catholic Church, for example, is legally not allowed to be motivated by profit. It was the maintenance of its own institutional hierarchies and reputation that motivated it to quietly move priests who sexually abused children from one parish to another, thereby supplying the abusers with fresh victims. Nonprofit institutions—even those ostensibly concerned with maintaining moral and spiritual values—are not immune to evil.

In an era of suspicious elections, campaign finance fiascos, and powerful lobbyists, one has to be naive in the extreme to believe that governments can be trusted to automatically or naturally work for the common good.

Appeals to small communities, nonprofits, or governments to take over economic activities “in the public interest” seem to me to bring in a deus ex machina solution.Yes, it would be nice if it worked. But how do we know that those selfish motivations critics assume drive the market are not also going to show up in families, community organizations, nonprofits, and the state?

A second problem with these views is that they largely pull the rug out from under their own noble drives. Because money and power are associated with greed and oppression, money and power are treated as inherently morally suspect. People who possess these, such as corporate executives who might be willing to engage in ethical discussion (if given the chance), are labeled as the evil “them,” separated by a large gulf from the moral “us.” Thus, potential allies and power bases are eliminated. This aversion to money and power has, I believe, been especially damaging to the sectors of the economy in which hands-on care is provided to children, the sick, and the elderly. Remember this poster: “It will be a great day when the schools have all the money they need and the air force has to hold a bake sale to buy a bomber”? How true. But the antimoney ideology reinforces exactly the bake-sale, nickel-and-dime mentality for human services that that poster decried. The damage this attitude has inflicted on caring work will be taken up further when I look at issues of money and motivations in chapter 4. A third problem is that, even if the prescriptions given by market critics were viable once put in place, there would still remain the problem of getting there. The massive promarket tide now flooding the United States and global institutions presents an intimidating reality check. The “small is beautiful” view tells us that we must have a massive economic restructuring— the thorough destruction of large corporations as a form of economic organization—before we can really be human in our economic lives. This would require a gargantuan change— larger, perhaps, than the Industrial Revolution and the rise and fall of Communism combined. If, on the other hand, we hope to be rescued by the rise of powerful, purely public-spirited interventionist governments, the current political climate makes it look like we may be waiting a very long time. Every step toward wresting control away from those with money and power will, market critics correctly perceive, be resisted by those with money and power.

Some people enjoy tilting at the economic machine—or at windmills, like Don Quixote in his hopeless crusades. In fact, I admire the spirit of people who keep to their praiseworthy, treasured values against all odds. But what if the futures envisioned by market critics, visions that tend to seesaw between the utopian and apocalyptic, are not the only options? What if the proposed solutions are unsatisfactory because the market critics have, unfortunately, combined good values with erroneous “facts” about what an economy is?

### 2AC---AT: Neoclassical Economics

#### Reading ‘neoclassical economics wrong’ to our stuff is an absurdist take. Economists no longer rely on it. You wouldn’t criticize a physics 101 textbook for excluding high-level friction equations.

Coyle ’18 [Diane Coyle; April 13; Economics Professor at the University of Manchester; Prospect Magazine, “In defence of the economists,” https://www.prospectmagazine.co.uk/economics-and-finance/dianecoyle]

SThese are typical. Most of modern economics is like this: empirical testing, often using new sources of data, addressing questions of immediate importance and relevance to policymakers, citizens and businesses.

Economists are not in thrall to neoclassical economics and they certainly do not neglect the role of technology. Economics does not require that people be rational, calculating automatons. It very often has people interacting with each other rather than acting as atomistic individuals, despite Reed’s charge. As a matter of course, it assumes there is a fog of uncertainty, and that different people know different things: that there are asymmetries of information.

Economists know that few markets are perfectly competitive and many—including me—specialise in studying markets where some firms have a lot of market power, customers may make choices that cost them more than necessary, and disequilibrium is the norm. There is nothing in economics that means we cannot take into account the ambiguous effects of changes such as an increase in pay or a higher tax rate. On the contrary, a lot of the empirical work we do involves trying to estimate the relative importance of things like this.

Among all the academic economists I know and read, many think the use and allocation of resources is best done by markets but understand that markets are shaped by government policy and that there is no such thing as the abstract “free” market. I literally can’t think of a single one for whom the “tenets of neoclassicism shape their day-to-day work,” as Reed puts it in his article. As for the charge that economics is guilty of a “neglect of technology,” this mystifies me as an economist—among many—currently researching technological change.

Yes, the subject is mathematical. It is not possible to do applied statistical estimation without writing things down as equations. Our models are written with mathematical notation as a shorthand and to enforce logical consistency, just as historians use models—”the causes of the First World War”—but write them out with lots of words.

It is both bizarre and frustrating to read over and over again the same absurd claims about economics. The caricature presumes that practitioners pass their careers in the discipline without ever being troubled by another thought after their first-year undergraduate course. Critiques of this kind are like looking at an introductory physics textbook and condemning physics for ignoring friction.

### 2AC---AT: Cap

#### Growth key to space col---extinction

Everett 16 (Sean, CEO of Prome Biological Intelligence, a global biotechnology company, editor of Medium’s news outlet dedicated to space colonialization titled “The Mission”, BS Mathematics & Actuarial Science, MBA from UChicago,“Humanity’s Extinction Event Is Coming” https://medium.com/the-mission/humanitys-extinction-event-is-coming-c0f84f1803f)

But the reality is that an asteroid impact, a change in our magnetic field, or the rising temperature of Earth’s climate are all events that we currently cannot escape. There is no back-up plan. We are, for better or worse, tied to the fate of this planet. As history has shown, that’s not a good fate to be tied to. In fact on September 7, 2016 a 30-foot asteroid flew between the Earth and the Moon. Our most powerful instruments only detected it with two days notice. Two days. If the asteroid was only 1000-foot wide, it would destroy all human life and we’d have no back-up to get out of it. Even the White House is worried about it. Five, yes five, major extinction events have occurred on our planet that we know about. We’re due for another. And when that happens, what’s our alternative? You can’t move to another house. You can’t buy survival, even with a billion dollars in the bank. The only way out, is up. We must find a way to become multi-planetary if we want to save humanity, your family, and yes, even yourself. Only this can restore the honor we seemed to have lost from the brave days of the 60s, while also ensuring our survival. It’s for the species, folks. And as a species, we have not allowed ourselves the opportunity to blast off for the stars. Only the space race in the 60s when we were afraid enough of a self-inflicted global extinction event (read: nuclear) that we put forth the funding required to launch into orbit and onto our moon. We didn’t have calculators back then, and now we have supercomputers in our pocket, but no one is allowed out of our atmosphere, save for a few communication and spy satellites. Doesn’t that make you mad? It’s not some oppressive government that tells us no. It’s us. We pay our taxes. We elect leaders. Those leaders choose Defense as the primary budget line item, but forget about defending against the forthcoming apocalypse. Funding for NASA in the United States has decreased from 4% of the national budget in the 60s to about 0.5% from 2010 onwards. That’s just the money side. But in order to move past this threshold from our home planet to space and then onto other planets, we need to do two things: Travel there. Survive. Luckily, we can simplify the problem of passing this barrier by sending machines in our place. Like TARS from Interstellar, they can go places humans cannot and explore the environment for habitability and resources, even in particularly hostile conditions. Maybe not black hole hostile, but definitely Mars hostile, as the Curiosity Rover has shown. Only now, with a few bold, private startups are we beginning to see a re-emergence of the space industry. We are about to pass a few very important tests that allow us to explore and visit the cosmos. The first is launching physical things into space. This is the catalyst that will jump start a new space race. Prices of sending cargo are falling dramatically, down to nearly $500 per pound of payload with SpaceX’s Falcon 9 heavy re-usable rocket. Note that the re-usable part is key. We can’t throw away our “space car” every time we Uber it. And once that becomes standard and cost-optimized we might be able to get that down to $10 per pound. Imagine what could happen when it costs the same amount to ship something across town as it does into space. The second, and this is just as important, is the wave of autonomous machines. Tesla has popularized the notion of self-driving cars. SpaceX lands their rocket onto a small barge in the ocean autonomously. Companies are buying startups in the space. Self-driving will be our gift, our talisman, on the quest to save the species by becoming multi-planetary. II. Shipping Ourselves to Space The graph below is from the Founders Fund manifesto, showing the decreasing cost of launching something into space. It begins with the 1960s US-versus-Russia space race and extends to the present day SpaceX-versus-Blue Origin reusable rocket race. The cheapest method we have today is SpaceX’s Falcon series rockets. With the Falcon 9 Heavy, it’s predicted launching cargo into space will be cheaper than ever before, at $750 per pound of payload delivered to low earth orbit (LOE)on an expendable rocket. You have to note here, however, that these statistics are as cheap as possible. It costs more to deliver payload on a non-reusable rocket, and on something that’s further out than LEO, like geosynchronous orbit, or to Mars. For example, based on SpaceX’s published pricing, it would be at least 4x more expensive to deliver far less cargo to Mars. So what happens when we reduce that cost to $10 per pound? Namely, an explosion of startups, much like iOS. Instead of pushing to production for your continuously deployed web and mobile app, we will see future developers push to production by deploying physical things into space. “STAGE” takes on an entirely new meaning for software developers when it means your automated regression tests fail, it could blow up a rocket and hurt people on board. That’s why SpaceX and Blue Origins exist. To make this continuous-deployment-to-space process as cheap and fast as possible. By Elon’s calculations, every 15 minutes. III. Self-Driving Space Explorers The most successful products for space, at least in the beginning, will make money by pushing this stuff into orbit. Things like science experiments and new 3D printers. A company called Made in Space creates a number of these products, including the empty box you see below used for sending things up with Blue Origin. The box shown in gray is a specialized 3D printer that works in zero gravity. Remember how most 3D printers work. It squeezes out a single layer of liquid ooze, and then another, over and over again until it builds up enough vertically that it creates an object. This can be simple plastic or more esoteroic metals. But when you’re “dripping” something, held down in place by gravity, the entire process has to be re-imagined for space. Things in zero-G would just float away. Enter these chaps. There’s also the very real need for oxygen, food, water, and shelter from the harsh elements. Funny how we will end up recreating Maslow’s Heirarchy in every new voyage or planetoid we want to colonize. And space mining is off to the races with the recent announcement of Deep Space Industry’s Prospector-1: Their vision is to extract water from asteroids and use the chemical components to hydrate us, but also as oxygen (breathing) and hydrogen (fuel). To do that, you have to identify candidate asteroids, physically get to them, land and attach, and then do surveying, prospecting, and extraction. In short, you’re going to need some level of self-driving capabilities to make this happen. And wouldn’t it be nice if it “just worked” right out of the box. Unfortunately, in space you don’t have fleets of these space craft, millions of miles of training data, maps, or an internet connection to the cloud so how the heck are deep learning algorithms going to work? I don’t think they will. And that’s what I believe we need a better approach.

#### Every facet of their position is wrong---growth is sustainable, transition is impossible

Piper '21 [Kelsey; 8/2/21; Writer, ran Stafford Effective Altruism, B.S. in Symbolic Systems; "Can we save the planet by shrinking the economy?" https://www.vox.com/future-perfect/22408556/save-planet-shrink-economy-degrowth/]

Most of the world is very poor. Billions of people go hungry, can’t afford a doctor when they get sick, don’t have adequate shelter and sanitation, and struggle to exercise the freedoms essential to a good life because of material deprivation.

But for all the immiseration around us, one thing is undeniable: For the past several centuries — and especially for the past 70 years, since the end of World War II — the world has been getting much richer.

That economic boom means a lot of things. It means cancer treatments and neonatal intensive care units and smallpox vaccines and insulin.

It means, in many parts of the world, houses have indoor plumbing and gas heating and electricity.

It means that infant mortality is down and life expectancies are longer.

But an increasingly wealthy world also means we eat more meat, mostly from factory-farmed animals. It means we emit lots more greenhouse gases. It means that consumers in developed countries buy a lot and throw away a lot.

In other words, it means a lot of good things and certainly some bad things as well.

Mainstream climate and environmental policy has developed over the years with a certain assumption — that we can get rid of the bad things while still preserving the good things. That is, it’s sought to figure out how to reduce carbon emissions, preserve ecosystems, and save endangered species while continuing to improve material living conditions for everyone in the world.

But to a vocal slice of climate activists, that approach seems increasingly doomed. The degrowth movement, as it’s called, argues that humanity can’t keep growing without driving humanity into climate catastrophe. The only solution, the argument goes, is an extreme transformation of our way of life — a transition away from treating economic growth as a policy priority to an acceptance of shrinking GDP as a prerequisite to saving the planet.

At the core of degrowth is the climate crisis. Degrowth’s proponents argue that to save Earth, humans need to shrink global economic activity, because at our current levels of consumption, the world won’t hit the IPCC target of stabilizing global temperatures at no more than 1.5 degrees of warming. The degrowth movement argues that climate change should prompt a radical rethinking of economic growth, and policymakers serious about climate change should try to build a livable world without economic growth fueling it.

It’s a bold, even romantic vision. But there are two problems with it: It doesn’t add up — and it would be nearly impossible to implement.

Addressing climate change will take genuinely radical changes to how our society works. Stirring as it might be to some, though, degrowth’s radicalism won’t fix the climate. Degrowth is most compelling as a personal ethos, a lens on your consumption habits, a way of life. What it’s not is a serious policy program to solve climate change, especially in a world where billions still live in poverty.

The basics of degrowth

Pinning down what degrowth means can be tricky because degrowthers often differ on details. But there are some common threads to their thought.

In general, degrowthers believe that in the modern world, economic growth has become unmoored from improvements in the human condition.

Jason Hickel, an anthropologist at the London School of Economics and the author of Less Is More: How Degrowth Will Save the World, has emerged as one of the leading spokespeople for the movement. To Hickel, the case for degrowth goes like this: The world is producing too much greenhouse gases. It is also overfishing, is overpolluting, is unsustainable in a dozen ways, from deforestation to plastic accumulating in the oceans.

Scientists have made impressive progress on technologies that, he argues, should have been sufficient to address the climate crisis — think solar panels, meat alternatives, eco-friendly houses. But because wealthy societies are so focused on growing the economy, those gains have been immediately plowed back into the economy, producing more stuff for the same ecological footprint, yes, but not actually shrinking the ecological footprint.

Hickel argues that this problem is unsolvable within our current framework. “In a growth-oriented economy,” he writes in Less Is More, “efficiency improvements that could help us reduce our impact are harnessed instead to advance the objectives of growth — to pull ever-larger swaths of nature into circuits of extraction and production. It’s not our technology that’s the problem. It’s growth.”

His solution? To abandon the lodestar of economic policy in nearly every country, which is to aim for economic growth over time, increasing wealth per person and expanding the ability of their citizens to purchase the things they want and need. Instead, Hickel argues, rich countries should focus on getting emissions to zero — even if the result is a much-contracted economy.

If that sounds unappealing, he devoted much of the book — and much of our interview — to arguing that it wouldn’t be. He points out that some countries, like the United States, are rich but get very little for their spending, in terms of national well-being; poorer countries like Spain have better health care systems. He argues that current levels of well-being could be maintained at a tenth of Finland’s current GDP — assuming that society also adopted wide-scale redistribution and socialist labor policies.

At the heart of Hickel’s argument is an idea that divides degrowthers and their critics: the concept of “decoupling” growth from environmental impact. Hickel and his fellow degrowthers are skeptical that economic growth as we know it can ever truly be achieved without accompanying growth in emissions.

But critics argue that not only is it possible — it’s already been happening. For the past decade, as many countries have transitioned to green energy, they have successfully seen their emissions shrink while their GDP has grown.

“There have been really big changes since 2005,” when people were debating whether decoupling was even possible, Zeke Hausfather, a climate scientist at the Breakthrough Institute, told me. “Green energy has gotten cheap. Solar power is the cheapest energy at the margins in every country today. Global coal use has peaked.” His research finds evidence of “absolute decoupling” — emissions shrinking while GDP grows — in 32 countries, including the United States, the United Kingdom, and Germany.

Degrowthers I spoke to don’t dispute that decoupling is possible. But they argue it won’t be enough to shrink emissions as rapidly as they need to. And there’s a compelling bit of evidence for that view: Even as some countries have decoupled, others have increased emissions, and overall atmospheric carbon is at its highest level ever recorded.

Where an optimist might see, in the decoupling of the past few decades, signs that growth and climate solutions can coexist, a pessimist might find the degrowth diagnosis more persuasive: that our growth-focused society clearly isn’t up to the task of solving climate change.

The pessimists have picked up momentum of late. It’s true, in one sense, that degrowth is a somewhat fringe idea: No politician has endorsed it, and no serious policy proposals based on it have been put forth. But degrowth has nonetheless drawn sympathy in some quarters — including among prominent climate thinkers.

Steven Chu, who served as secretary of energy under President Obama, has endorsed it, arguing, “You have to design an economy based on no growth or even shrinking growth.”

More than 11,000 scientists signed William Ripple’s 2019 letter “World Scientists’ Warning of a Climate Emergency,” which argues “our goals need to shift from GDP growth and the pursuit of affluence toward sustaining ecosystems and improving human well-being by prioritizing basic needs and reducing inequality.”

And a recent paper in Nature explored how a “degrowth” of 0.5 percent of GDP per year might interact with climate and emissions targets, arguing that while “substantial challenges remain regarding political feasibility,” such approaches should be “thoroughly considered.”

The tension at the heart of degrowth: Can we fix global poverty without economic growth?

One big problem with degrowth is this simple fact: In the coming decades, most carbon emissions won’t be coming from rich countries like the US — they’ll be happening in newly middle-income countries, like India, China, or Indonesia. Already, developing nations account for 63 percent of emissions, and they’re expected to account for even more as they develop further and as the rich world decarbonizes.

Even if emissions in rich countries go to zero very soon, climate change is set to worsen as poorer countries increase their own emissions.

That will, of course, have deeply negative climate impacts. But the alternative is a nonstarter — should the world really prioritize curbing emissions and economic growth if it meant suppressing the growth of those countries?

Degrowthers see no dilemma here. What Hickel envisions is global movement in two directions: Poor countries could develop up to a certain level of prosperity and then stop; rich countries could develop down to that level and then stop. Thus, climate catastrophe could be averted, all while making the world’s poor more prosperous.

“Rich countries urgently need to reduce their excess energy and resource use to sustainable levels so our sisters and brothers in the global South can live well too,” Hickel put it. “We live on an abundant planet and we can all flourish on it together, but to do so we have to share it more fairly, and build economies that are designed around meeting human needs rather than around perpetual growth.”

From a climate change perspective, though, there’s a problem. First, it means that degrowth would do nothing about the bulk of emissions, which are occurring in developing countries.

Second, the global economy is more interconnected than Hickel implies. When Covid-19 hit, poor countries were devastated not just by the virus but by the aftershocks of virus-induced slowdowns in consumption in rich countries.

There’s some genuine appeal to the idea of an end to “consumerism,” but the pandemic offered a taste of how a sudden drop in rich-world consumption would actually affect the developing world. Covid-19 dramatically curtailed Western imports and tourism for a time. The consequences in poor countries were devastating. Hunger rose, and child mortality followed.

Covid-19, of course, wreaked direct economic havoc at the same time, with lockdowns having an especially negative impact on some poor countries; the effects of the pandemic and international demand shock were combined, and in some cases they’re hard to separate. But the United Nations, the World Bank, and expert analyses point to the decline in global consumption as a significant part of the picture.

Degrowthers reject this concern on two fronts: First, they argue that a sustained, deliberate reduction in consumption wouldn’t be anything like a recession. Recessions, they agree, are really bad, but that’s because consumption falls in affected sectors, instead of being targeted at things that don’t improve well-being. Degrowth, they say, would be different.

Second, they contend that there is some path to economic growth in poor countries that doesn’t rely on trade with rich ones — certainly some countries managed economic growth when the whole world was poor, after all.

Hickel’s perspective is that most trade between rich and poor countries is extractive, not mutually beneficial — and that maybe when that dynamic ceases, poor countries will have the chance for the catch-up growth they merit. That’s one take. But it means that degrowth’s case for not crushing the poor world is predicated on a speculative take on how those countries can grow — one that democratically elected leaders in those countries largely don’t share.

What GDP doesn’t capture — and what it can tell us

In a way, the debate over degrowth is a debate over the meaning of one economic indicator: gross domestic product (GDP).

GDP measures the transactions within an economy — all the occasions when money changes hands in exchange for goods and services. It’s not wealth, but it’s one of the primary ways we measure wealth.

It certainly doesn’t capture everything of value. When parents spend a quiet weekend at home teaching their children to read, for example, nothing GDP-generating has happened — but value has certainly been created.

Degrowth articles burst with such examples. GDP, they love to point out, includes the production of things like nerve gas, even though that has no social value. And it doesn’t include storytelling, singing, gardening, and other simple human pleasures.

“If our washing machines, fridges, and phones lasted twice as long, we would consume half as many (thus the output of those industries would decline), but with zero reduction in our access to those goods,” Hickel told me. If everyone worked half the hours they currently do, and made half the income, they might mostly be better off — at least, assuming that their basic needs were still met.

“We propose policies like a living wage, a maximum income ratio, wealth taxes, etc. to accomplish this,” Hickel told me. “Given all of this, the language of poverty really gets it wrong: longer-lasting products, living wages, shorter working weeks, better access to public services and affordable housing — we are calling for the opposite of poverty. Yes, industries like SUVs and fast fashion would decline, but that doesn’t mean poverty. We can replace them with public transportation and longer-lasting fashion, thus meeting everyone’s needs.”

There’s a lot of speculation here, and a lot of what degrowth’s critics would call hand-waving. Degrowth is fundamentally premised on the claim that we can cease to focus on growth while getting better than ever at addressing human needs. If that’s true, then that would certainly be great news.

But in many ways, it’s a vision more wildly optimistic — disconnected from actual policy results — than any of the more standard “sustainable development” models degrowthers criticize for being out of touch.

First, in the world today, there’s an extremely strong association between growth and welfare outcomes of every kind. GDP, while imperfect, is a better predictor of a country’s welfare state, outcomes for poor citizens in that country, and well-being measures like leisure time and life expectancy than any other measure.

“GDP does leave out non-commercialized activities that are welfare-enhancing,” economist Branko Milanovic writes in a rebuttal of degrowth:

It is, like every other measure, imperfect and one-dimensional. But ... it is imperfect at the edges while fairly accurate overall. Richer countries are countries that are generally better-off in almost all metrics, from education, life expectancy, child mortality to women’s employment etc. Not only that: richer people are also on average healthier, better educated, and happier. Income indeed buys you health and happiness. (It does not guarantee that you are a better person; but that’s a different topic.) The metric of income or GDP is strongly associated with positive outcomes, whether we compare countries to each other, or people (within a country) to each other.

The things degrowthers care about — leisure time, health care, life expectancy — are strongly correlated with societal wealth. The generosity of a welfare state and the availability of transfers to a state’s poorest people are also strongly correlated with societal wealth. Innovation, discovery, invention, and medical technology improvements are also strongly correlated with societal wealth.

The strong correlation between child mortality and GDP per capita is apparent on the above graph. There are some outliers — some countries outperform or underperform their GDP somewhat, in terms of preventing child deaths — but in general, wealth strongly predicts child survival. No single, simple medical intervention causes the difference. Wealthier societies on average get better health outcomes across the board.

This graph looks at child mortality not just by comparing rich countries to poor ones but also by comparing countries over time, as they get richer: Getting richer improves outcomes for children.

Leisure time, too, has increased — and hours worked have declined — as the world has gotten wealthier.

It might be possible in principle to do better — to decouple, if you will, health and well-being from access to material resources, so that everyone is well-off with many fewer resources.

But the examples degrowthers point to remain speculative ones; if we ought to be skeptical, as degrowthers argue we should be, about the decoupling of wealth from ecological impact, we ought to be at least as skeptical about the prospects of decoupling wealth from living standards.

“In the end, economic growth is about the production of stuff that people need and then the consumption of those things by the people who need it,” Max Roser at Our World in Data, a research institute focused on finding, visualizing, and communicating historical economic and health data, told me. He added:

The money aspect, and the abstract concept of GDP, distract us and make it less obvious what it’s actually about. People want to have enough food, they need to go to the doctor, they need childcare, they want a good education. People need lots of stuff, and one thing that people care about are goods and services, and they need to be produced, and economic growth is about an increase in the quality and quantity of the goods and services that people need.

There’s also the knotty problem of who gets to decide which goods and services people choose to spend their money on. Many of the climate scientists I spoke to shared Hickel’s impatience for many specific carbon-intensive modern industries. “I’m not going to defend bitcoin,” the Breakthrough Institute’s Hausfather told me. (The cryptocurrency has attracted intense criticism for being astoundingly carbon-intensive.)

But there is a lot in between bitcoin and basic subsistence needs. And “enough for everyone who needs it” inherently requires value judgments about what people really need, and what things they value that are frivolous luxuries. That’s why so many anti-poverty programs have moved away from giving people “what they need” toward just giving them cash — that is, giving them wealth, which they can choose to spend however they please.

“Even poor people have so many needs for goods and services that you can’t possibly put them on a list and say, ‘Now we’re done here,’” Roser told me. “That’s the beauty of money, that you can just go out there and get what you need rather than what some researcher determines are your needs.”

Degrowth is unrealistic — and gaining traction

As a policy program, degrowth suffers from being both too radical and not radical enough.

There’s a lot of broad-brush policy prescriptions in the degrowth lit, but those details never really add up.

While it’s not a short book, Less Is More feels surprisingly sparse when it comes to envisioning how the changes it recommends could be brought about. The chapter on solutions recommends cutting the workweek and changing tax policy — two solid proposals — but then rounds that out by recommending ending technological obsolescence, advertising, food waste, and student debt.

I’m not particularly opposed to those policies. But they seem laughably inadequate for the magnitude of the task at hand: confronting the climate crisis. Degrowth successfully persuades that guiding humanity and our planet through the 21st century will be really, really hard — but not in a way degrowth particularly solves.

Where degrowth literature is relentlessly pessimistic about the prospect of our problems being solved under our current economic system, it turns oddly optimistic about the prospect that they’ll be solved once we embrace a different way of viewing wealth and progress. If cutting carbon emissions fast enough to matter requires shrinking the global economy by 0.5 percent a year indefinitely, starting right now, as the Nature paper estimates, that’ll take policy measures much larger and more ambitious than any proposed in Less Is More.

“If we are to avert catastrophic warming, we have to lower carbon emissions by a factor of two within the next 10 years. I find it highly implausible that capitalism/market economics will be abandoned by the world on that time frame,” Pennsylvania State University climatologist Michael Mann told me. “That means we have to act on the climate crisis within the framework of the current system.”

In that sense, there’s actually something anti-radical about any climate plan so radical that it can’t be concretely brought about in the next decade.

And yet, implausible as it is, degrowth is gaining a foothold in intellectual and policy circles. What accounts for its seemingly growing popularity? This was a question that puzzled me until I heard the same answer from one degrowth advocate and one opponent: that it’s not, really, exactly about climate.

“It started in the 1990s in France, picking up on radical European politics in the 1970s,” Giorgos Kallis, a researcher studying degrowth at the Universitat Autònoma de Barcelona, told me. “There was an in-between political space there — radical greens, putting much more emphasis on localized production, emphasis on conviviality and autonomy. This is a discourse that comes from them. It wasn’t just about avoiding a particular environmental problem. It was a holistic proposal.”

That was also the diagnosis of Zion Lights, a former spokesperson for Extinction Rebellion, who has become one of the climate movement’s internal critics, arguing that the movement focuses too much on environmentalist-friendly proposals that have nothing to do with climate.

“It has become difficult to talk about making energy policies for combating climate change, for example, without being told that such thinking is actually irrelevant because it doesn’t involve system change,” she recently argued. “We need cheap, clean energy at scale and we need it now.”

In that sense, a good analogy for degrowth might actually be locavorism — the movement that focuses on eating food grown locally. It’s popular with environmentalists, both those whose convictions are about climate change and those who long for a return to the land. Its actual climate impacts are limited or even negative — for some products, it’s better for them to be grown in their optimal environment even with carbon-intensive shipping — and it definitely does less for the climate than, for example, going vegan. But it retains its allure.

How to fight climate change while building good human societies

Degrowth’s radicalism isn’t where I part ways with it: The future will almost certainly require us to eat much less meat, dramatically change land use, and potentially invest a significant chunk of society’s resources in mitigation indefinitely.

But I don’t tend to see such efforts as fundamentally futile. Degrowthers do — even when there have been significant successes.

Climate scientists have spent a long time warning the world about climate change, but they nonetheless tend to sound a more optimistic note than degrowthers like Hickel. “It’s undoubtedly a monumental challenge,” Mann told me. “We have the technology to solve the problem — renewable energy, smart grid technology, and existing energy storage. We just need the political will to act.”

Take solar panels. Two decades ago, cheap solar panels were just a dream. Now they’re everywhere and have become a crucial tool in the fight against climate change.

Not only that, solar panels have democratized electricity. Just one small-scale instance: In rural Kenya, you can see donkeys saddled with solar panels so that farmers can charge their phones. And there are many such examples that count as a win for both human progress and our fight against climate change.

It should go without saying that since rich governments got us into this climate mess, they should be at the forefront of getting us out of it. We need massive investments in carbon capture, green energy, plant-based meat, mitigation, and straight-up cash transfers to poor countries disproportionately affected by the climate crisis.

Many of the researchers I spoke to were open to the idea that in the long run, humanity would need to rethink many of our cherished assumptions about how economies work, in order to build a civilization that can flourish for thousands or millions of years. They didn’t reject degrowth as a philosophical contribution to the question of what future human civilizations should care about.

But such articulations of different philosophies of human flourishing should not be mistaken for public policy.

We don’t have very long, and we need to decarbonize quickly. We have technologies that have made a big difference already, and they must be made available on an unprecedented scale. We have more speculative solutions, technological and societal, and we should be prepared to try those, too. The scale of the problem is such that we need to act now — and we need to be clear-eyed about which ideas truly move the needle.

#### No impact to the environment

Halstead ’19 [John; April 2019; Ph.D. from the University of Oxford, researcher at Founders Pledge, citing Dr. Peter Kareiva, a Ph.D. in ecology and evolutionary biology at Cornell University and Director of UCLA’s Institute of the Environment and Sustainability; Centre for the Study of Existential Risk, “Centre for the Study of Existential Risk Six Month Report: November 2018 - April 2019,” <https://forum.effectivealtruism.org/posts/zbZxisJRJBCdtYvh9/centre-for-the-study-of-existential-risk-six-month-report>]

There are independent reasons to think that the risk is negligible. Firstly, according to wikipedia, during the Eocene period ~65m years ago, there were thousands fewer genera than today. We have made ~1% of species extinct, and we would have to continue at current rates of species extinctions for at least 200 years to return to Eocene levels of biodiversity. And yet, even though significantly warmer than today, the Eocene marked the dawn of thousands of new species. So, why would we expect the world 200 years hence to be inhospitable to humans if it wasn't inhospitable for all of the species emerging in the Eocene, who are/were significantly less numerous than humans and significantly less capable of a rational response to problems?

Secondly, as far as I am aware, evidence for pressure-induced non-linear ecosystem shifts is very limited. This is true for a range of ecosystems. Linear ecosystem damage seems to be the norm. If so, this leaves more scope for learning about the costs of our damage to ecosystems and correcting any damage we have done.

Thirdly, ecosystem services are overwhelmingly a function of the relations within local ecosystems, rather than of global trends in biodiversity. Upon discovering Hawaii, the Polynesians eliminated so many species that global decadal extinction rates would have been exceptional. This has next to no bearing on ecosystem services outside Hawaii. Humanity is an intelligent species and will be able to see if other regions are suffering from biodiversity loss and make adjustments accordingly. Why would all regions be so stupid as to ignore lessons from elsewhere? Also, is biodiversity actually decreasing in the rich world? I know forest cover is increasing in many places. Population is set to decline in many rich countries in the near future, and environmental impact per person is declining on many metrics.

I also find it surprising that you cite the Kareiva and Carranza paper in support of your claims, for this paper in fact directly contradicts them:

"The interesting question is whether any of the planetary thresholds other than CO2 could also portend existential risks. Here the answer is not clear. One boundary often mentioned as a concern for the fate of global civilization is biodiversity (Ehrlich & Ehrlich, 2012), with the proposed safety threshold being a loss of greater than 0.001% per year (Rockström et al., 2009). There is little evidence that this particular 0.001% annual loss is a threshold—and it is hard to imagine any data that would allow one to identify where the threshold was (Brook, Ellis, Perring, Mackay, & Blomqvist, 2013; Lenton & Williams, 2013). A better question is whether one can imagine any scenario by which the loss of too many species leads to the collapse of societies and environmental disasters, even though one cannot know the absolute number of extinctions that would be required to create this dystopia.

While there are data that relate local reductions in species richness to altered ecosystem function, these results do not point to substantial existential risks. The data are small-scale experiments in which plant productivity, or nutrient retention is reduced as species numbers decline locally (Vellend, 2017), or are local observations of increased variability in fisheries yield when stock diversity is lost (Schindler et al., 2010). Those are not existential risks. To make the link even more tenuous, there is little evidence that biodiversity is even declining at local scales (Vellend et al., 2013, Vellend et al., 2017). Total planetary biodiversity may be in decline, but local and regional biodiversity is often staying the same because species from elsewhere replace local losses, albeit homogenizing the world in the process. Although the majority of conservation scientists are likely to flinch at this conclusion, there is growing skepticism regarding the strength of evidence linking trends in biodiversity loss to an existential risk for humans (Maier, 2012; Vellend, 2014). Obviously if all biodiversity disappeared civilization would end—but no one is forecasting the loss of all species. It seems plausible that the loss of 90% of the world’s species could also be apocalyptic, but not one is predicting that degree of biodiversity loss either. Tragic, but plausible is the possibility of our planet suffering a loss of as many as half of its species. If global biodiversity were halved, but at the same time locally the number of species stayed relatively stable, what would be the mechanism for an end-of-civilization or even end of human prosperity scenario? Extinctions and biodiversity loss are ethical and spiritual losses, but perhaps not an existential risk."

# 1AR

## CP---Sunbursting

### 1AR---Resolved

#### Resolved indicates a vote

**Webster’s** Revised Unabridged Dictionary, 1998 (dictionary.com)

Resolved: 5. To express, as an opinion or determination, by resolution and vote; to declare or decide by a formal vote; -- followed by a clause; as, the house resolved (or, it was resolved by the house) that no money should be apropriated (or, to appropriate no money).

### --1AR---AT: Colon

#### Means nothing

Encarta 7 (World Dictionary, “colon”, [http://encarta.msn.com/encnet/features/dictionary/DictionaryResults.aspx?refid=1861 598666](http://encarta.msn.com/encnet/features/dictionary/DictionaryResults.aspx?refid=1861598666))

co·lon (plural co·lons)

noun

Definition:punctuation mark: the punctuation mark (:) used to divide distinct but related sentence components such as clauses in which the second elaborates on the first, or to introduce a list, quotation, or speech. A colon is sometimes used in U.S. business letters after the salutation. Colons are also used between numbers in statements of proportion or time and Biblical or literary references.

### 1AR---Should

#### Should reflects hypothetical future action---they cut a straw man from the dissenting opinion---independent VI for evidence ethics

Summers 94 Justice – Oklahoma Supreme Court, “Kelsey v. Dollarsaver Food Warehouse of Durant”, 1994 OK 123, 11-8, <http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn13>, EO

**SUMMERS**, Justice.

[885 P.2d 1354]

¶1 Plaintiff (Appellee) was awarded a substantial judgment pursuant to a jury's verdict in Bryan County. Defendant (Appellant) filed a timely Motion for New Trial and Motion for Judgment Notwithstanding the Verdict. On May 18, 1993, the trial judge signed and caused to be filed a handwritten document entitled "Court Minute", which stated "the Court finds that the motions should be overruled."

¶2 On June 2, 1993, the judge signed and caused to be filed a typed Order as follows: "IT IS THEREFORE ORDERED, ADJUDGED & DECREED by the Court that Defendant's Motion for New Trial and Motion for Judgment Notwithstanding the verdict be, and they are hereby, overruled." Defendant commenced its appeal here on June 30, 1993. Plaintiff moves to dismiss for untimeliness, claiming the earlier writing commenced the appellate clock. We deny the motion and allow the appeal to proceed.

¶3 Our decision does not rest on whether judge-signed "court minutes" prior to the legislative change of October 1, 1993 are or are not appealable orders.[1](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker1fn1) That is because the writing of May 18 here under neither view contains language sufficient to make it an order of the court. A statement that the motions should be overruled is not the same as stating that the motions are overruled.

¶4 Webster describes "should" as a word used in an auxiliary function to express (1) condition (if he should leave his father his father would die); (2) propriety (this is as it should be); (3) futurity (he realized she should have to do most of her farm work before sunrise); or (4) what is probable or expected (they should be here before noon.) Merriam Webster's Collegiate Dictionary, 1085 (10th ed. 1993). "Should" **as used by the judge here** places his statement **in the subjunctive mood** rather than in either the indicative (or declarative) mood or the imperative mood. See II G. Curme, A Grammar of the English Language, 391 (1980) and volume I of Dr. Curme's work at 224.

¶5 The subjunctive mood is a verb form representing an act or state, not as fact, but as contingent or **possible**. Merriam Webster's Collegiate Dictionary, supra at P. 1172. In Jackson v. State of Indiana, 273 Ind. 49, 402 N.E.2d 947 (1980) the defendant on appeal argued that the trial court erred in overruling his motion for a mistrial. The appellate court pointed out that the defendant had not made such a motion:

Defendant now contends that the trial court erred in refusing to grant his motion for mistrial. However, the record **does not reflect** that defense counsel made such **a motion**; rather, he merely stated, "Your Honor, at this point, I should move for a mistrial." The use of the subjective, "should," reflects a mere contingent or **hypothetical action**.

Id. 402 N.E.2d at 951 (emphasis added). The trial judge in our case, just as the defense counsel in Jackson, chose to set his statement in the subjunctive mood, here a representation that the motions were going to be, but had not yet been, overruled.

¶6 We also note that the trial court in its May 18 entry stated that "the Court finds that the motions should be overruled." (emphasis ours) Our ruling today is consistent with Tillman v. Tillman, 199 Okla. 130, [184 P.2d 784](http://www.oscn.net/applications/oscn/deliverdocument.asp?citeid=1926) (1947). In that case the trial court had stated for the record certain findings which seemed indicative of the court's opinion as to the extent of his jurisdiction. This Court said:

A judgment . . . is distinct from findings of the court, [citation omitted] Findings and opinion of the court are never the judgment, but only expressions as to what the court considers its judgment should be. Id. 184 P.2d at 785. (emphasis added).

¶7 The language used by the court on May 18 falls short of amounting to an order of the [885 P.2d 1355] court. It announced that the motions should be overruled, but it did not overruled them. The later language used on June 2 declaring that the motions "be, and they are hereby, overruled", did not express futurity or probability, but was an indicative, declarative statement of what the judge did with the motions. It overruled them. It was an order of the court, and it was timely appealed by Defendant's petition in error filed June 30.

¶8 The motion to dismiss is denied.

¶9 HODGES, C.J., LAVENDER, V.C.J., and ALMA WILSON, KAUGER, WATT, JJ., concur.

¶10 SIMMS, HARGRAVE and OPALA, JJ., dissent.

Footnotes:

[1](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker0fn1) [12 O.S. 1993 Supp. § 696.2 [93-696.2](C)](http://www.oscn.net/applications/oscn/deliverdocument.asp?citeid=94065&date=11/8/94), effective October 1, 1993, provides:

"The following shall not constitute a judgment, decree or appealable order: a minute entry. . . ."

The issue was resolved in Manning v. State ex rel. Dept. of Public Safety, [876 P.2d 667](http://www.oscn.net/applications/oscn/deliverdocument.asp?citeid=20150) (Okl. 1994), in which we held that such orders could be appealable if they fully disposed of a case, but made the ruling prospective.

**OPALA**, Justice, with whom SIMMS and HARGRAVE, Justices, join, **dissenting**.

¶1 Concluding that the appellant's petition in error was timely brought, the court holds today that the June 2, 1993 record entry, and not the May 18, 1993 filed memorial,[1](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn1) of the trial court's ruling triggers appeal time. The court reasons the earlier memorial is not an order because (1) it is couched as a "finding" and (2) it stated that Dollarsave Food Warehouse of Durant's [defendant or Dollarsaver] motion for judgment n.o.v. "should be overruled" rather than using the customary phrase "should be and the same hereby is overruled".[2](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn2) I cannot accede either to the court's reasoning or to its conclusion.

I

THE SUBSTANTIVE CONTENT OF THE MAY 18 MEMORIAL DETERMINES WHETHER IT QUALIFIES AS AN ORDER

¶2 The meaning and effect of a legal instrument depends on its substantive content rather than on the form or title provided by its author.[3](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn3) Although the May 18, 1993 memorial bears the printed title "court minute", the paper clearly meets all the attributes of a recordable[4](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn4) order.[5](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn5) Its content directs that [885 P.2d 1356] the relief sought by Dollarsaver - i.e., judgment notwithstanding the verdict - be denied and the "direction" is signed by the judge. The instrument is, without a doubt, fit for entry upon the journal record and also constitutes an order that triggers appeal time because it fully complies with the § 696.3 standards.[6](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn6)

II

THE LEGAL EFFECT OF A RECORD ENTRY THAT USES THE WORD "FINDS" AS A SYNONYM FOR OR AS INTERCHANGEABLE WITH THE VERB "ORDERS" IS DETERMINED BY THE JUDGE'S INTENT

¶3 In concluding that the judge-signed direction is but a "finding",[7](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn7) the court disregards the very essence of the May 18 entry. It memorializes the ruling made upon the defendant's motion for judgment non obstante veredicto.[8](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn8) When in a case calling for the assessment of the parties' relative negligence a motion for judgment n.o.v. comes on for ruling, it is not within the nisi prius court's province to make any findings of fact.[9](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn9) While the use of the verb "finds" to overrule Dollarsaver's motion was perhaps grossly unartful,[10](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn10) the context in which the word ("finds") is used unmistakenly indicates that it was intended as a synonym for the verb "rules" or "orders".[11](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn11) The court's construction today renders the entry nugatory. Its result, which is neither mandated by nor consistent with the clear meaning and substance of the memorial's intended meaning, is both improvident and contrary to precedent.[12](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn12)

[885 P.2d 1357]

III

IN TERMS OF THE POSTJUDGMENT MEMORIAL HERE UNDER REVIEW THE MEANING TO BE GIVEN THE WORD "SHOULD" IS THE SAME AS THAT IN THE VERBS "ORDERS" OR "RULES", EVEN THOUGH "SHOULD" IS NOT FOLLOWED IMMEDIATELY BY THE TRITE; AND ANCIENT PHRASE "AND HEREBY IS"

**[their card starts here]**

¶4 The legal question to be resolved by the court is **whether** the word "**should**"[13](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn13) in the May 18 order connotes futurity or may be deemed a ruling in praesenti.[14](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn14) The answer to this query is not to be divined from rules of grammar;[15](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn15) it must be governed by the age-old practice culture of legal professionals and its immemorial language usage. To determine if the omission (from the critical May 18 entry) of the turgid phrase, "and the same hereby is", (1) makes it an in futuro ruling - i.e., an expression of what the judge will or would do at a later stage - or (2) constitutes an in in praesenti resolution of a disputed law issue, the trial judge's intent must be garnered from the four corners of the entire record.[16](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker3fn16)

[CONTINUES – TO FOOTNOTE]

[**13**](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker2fn13)"Should" not only is used as a "present indicative" **synonymous with ought** but also is the past tense of "shall" with various shades of meaning not always easy to analyze. See 57 C.J. Shall § 9, Judgments § 121 (1932). O. JESPERSEN, GROWTH AND STRUCTURE OF THE ENGLISH LANGUAGE (1984); St. Louis & S.F.R. Co. v. Brown, 45 Okl. 143, 144 P. 1075, 1080-81 (1914). For a more detailed explanation, see the Partridge quotation infra note 15.

Certain contexts mandate a construction of the term "should" as more than merely indicating preference or desirability. Brown, supra at 1080-81 (jury instructions stating that jurors "should" reduce the amount of damages in proportion to the amount of contributory negligence of the plaintiff was held to imply an obligation and to be more than advisory); Carrigan v. California Horse Racing Board, 60 Wash. App. 79, [802 P.2d 813](http://www.oscn.net/applications/oscn/deliverdocument.asp?box1=802&box2=P.2D&box3=813) (1990) (one of the Rules of Appellate Procedure requiring that a party "should devote a section of the brief to the request for the fee or expenses" was interpreted to mean that a party is under an obligation to include the requested segment); State v. Rack, 318 S.W.2d 211, 215 (Mo. 1958) ("should" would mean the same as "shall" or "must" when used in an instruction to the jury which tells the triers they "should disregard false testimony").

[**14**](http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=20287#marker2fn14) In praesenti means literally "at the present time." BLACK'S LAW DICTIONARY 792 (6th Ed. 1990). In legal parlance the phrase denotes that which in law is presently or immediately effective, as opposed to something that will or would become effective in the future [in futurol]. See Van Wyck v. Knevals, [106 U.S. 360](http://www.oscn.net/applications/oscn/deliverdocument.asp?box1=106&box2=U.S.&box3=360), 365, 1 S.Ct. 336, 337, 27 L.Ed. 201 (1882).

### 1AR---Deficit

#### the Court will never back down.

Carbonneau ’20 [Thomas; November 18; Samuel P. Orlando Distinguished Professor of Law at Pennsylvania State University, Faculty Director of the Arbitration Institute; Arbitration Law in a Nutshell, “The Central Themes of American Arbitration Law,” Ch. 2]

A small group of legislators have always opposed the Court’s pronouncements on adhesive arbitration, believing that they infringe upon the interests of their preferred constituents and the terrain of legislative authority. None of the proposed ‘anti-arbitration’ bills ever went beyond the committee-level of consideration. When the Court validated adhesive contracts and class action waivers in consumer arbitration, a group of congressional representatives and U.S. senators proposed a more restrictive legislative regulation of arbitration. The Fairness in Arbitration Act nullified arbitral clauses in adhesive consumer contracts entirely because the proponents believed that such agreements epitomized fundamental unfairness. The opponents of arbitration demonstrated an in-depth knowledge of the arbitral process and its propensity to self-regulate. To prevent arbitrators from resurrecting the prohibited contracts on the bases of their own authority, the legislation eliminated the arbitrators’ Kompetenz-Kompetenz powers. Under their power to rule on their own jurisdiction, the arbitrators could conclude that they retained decisional powers under the arbitral clause to disregard the provisions in the Fairness in Arbitration Act. By eliminating the arbitrators’ Kompetenz-Kompetenz powers, the legislators prevented them from undoing the prohibition they implanted in the enactment against adhesive contracts. Extinguishing Kompetenz-Kompetenz powers was a fail-safe measure. The legislators had decided that adhesive bargains for arbitration were malum in se and could not be tolerated under any circumstance or for any reason. The proposed legislation was filled with political convictions that contradicted the Court majority and its absolute support for arbitration. Like its predecessors, the bill never surfaced from the committee.

Undoubtedly, adhesive arbitration is the most controversial aspect of contemporary U.S. arbitration law. It pits significant political forces against one another. The intensity of their exchanges may undermine arbitration’s general reputation; the tendency to demonize the opponent will per force convey a distorted picture of arbitration and its social utility. Neither the Court nor interest groups are likely to back down from a confrontation. The Court’s validation of both adhesive consumer contracts and class action waivers ignited even more acrimonious political discord. Commercial parties abhor class litigation because it makes the conduct of business more difficult and expensive, converting it (to some extent) into a lawyer’s playground. Commercial parties especially resent having their organization invaded by ‘legal eagles’ who ‘loot’ its value for their own ends. A contario sensu, the advocates of government intervention consider class litigation essential to the fairness of society. Be that as it may, it seem that the Court continues to believe that exceptions to arbitrability will, overtime, lead to the collapse of the doctrinal edifice it has built to re-constitutionalize the American legal process.

## K---Cap

### 1AR---AT: Innovation

#### Central planning dooms innovation and their empirics are wrong

Vander Elst 18 [Philip Vander Elst is a freelance writer and lecturer who has spent nearly 30 years in politics and journalism and now works with Areopagus Ministries, “Soviet Communism Was Dependent on Western Technology,” 3/11/18, https://fee.org/articles/soviet-communism-was-dependent-on-western-technology/]kiihnl

The "Brain Drain" and the **Problem with Central Planning**

Far from Soviet Communism never having “had the chance to develop” because of interference from the West as Fiona Lali believes, the endemic economic failure and oppressive character of the Soviet Union flowed inevitably from its Marxist model of economic and social development. **A society in which the State owns and controls every sector of the economy**, and is the sole landlord, employer, doctor, educator, and welfare provider, **cannot fail to be destructive of** freedom, **personal incentives, creativity**, and entrepreneurship, while monopolistic government central planning, reflecting the limited knowledge and political priorities of the ruling bureaucracy, **inevitably stifles innovation** and technical progress. That is why the negative experience of Soviet Communism was repeated in every other Communist revolution and country during the last century.f

Given these truths, the idea that Western interference hindered the outworking and therefore the success of the Communist experiment in the Soviet Union is absurd. As will be shown below, the exact opposite was the case. In one form or another, **Western capital**, “know-how,” **and technology actually pulled Soviet Communism’s chestnuts out of the fire in nearly every decade** of the Soviet Union’s existence, principally by compensating it for its above-mentioned systemic inability to generate significant levels of indigenous technological innovation.

While there was nothing inherently lacking in the quality of Soviet scientific research, the **limitations of central planning** **and the absence of market mechanisms and incentives prevented the systematic testing of the fruits of research against competing alternatives.** Instead of allowing the dispersed knowledge, opinions, and talents of millions of individuals freely co-operating in the marketplace to determine the success or failure of new ideas and discoveries, nearly all economic activity in the Soviet Union was narrowly constrained within the developmental straitjacket imposed by its all-powerful Communist rulers; hence the need to import skilled personnel, know-how, and technology from the freer and more dynamic societies of Western Europe and North America.

### 1AR---Warming---T/

#### Yes decoupling---best and most recent studies AND leakage is wrong.

Zeke Hausfather 21, Director, Climate and Energy at The Breakthrough Institute, "Absolute Decoupling of Economic Growth and Emissions in 32 Countries," Breakthrough Institute, 04/06/2021, https://thebreakthrough.org/issues/energy/absolute-decoupling-of-economic-growth-and-emissions-in-32-countries.

The past 30 years have seen immense progress in improving the quality of life for much of humanity. Extreme poverty — the number of people living on less than $1.90 per day — has fallen by nearly two-thirds, from 1.9 billion to around 650 million. Life expectancy has risen in most of the world, along with literacy and access to education, while infant mortality has fallen. Despite perceptions to the contrary, the average person born today is likely to have access to more opportunities and have a better quality of life than at any other point in human history. Much of this increase in human wellbeing has been propelled by rapid economic growth driven largely by state-led industrial policy, particularly in poor-to-middle income countries.

However, this growth has come at a cost: between 1990 and 2019, global emissions of CO2 increased by 56%. Historically, economic growth has been closely linked to increased energy consumption — and increased CO2 emissions in particular — leading some to argue that a more prosperous world is one that necessarily has more impacts on our natural environment and climate. There is a lively academic debate about our ability to “absolutely decouple” emissions and growth — that is, the extent to which the adoption of clean energy technology can allow emissions to decline while economic growth continues.

Over the past 15 years, however, something has begun to change. Rather than a 21st century dominated by coal that energy modelers foresaw, global coal use peaked in 2013 and is now in structural decline. We have succeeded in making clean energy cheap, with solar power and battery storage costs falling 10-fold since 2009. The world produced more electricity from clean energy — solar, wind, hydro, and nuclear — than from coal over the past two years. And, according to some major oil companies, peak oil is upon us — not because we have run out of cheap oil to produce, but because demand is falling and companies expect further decline as consumers increasingly shift to electric vehicles.

The world has long been experiencing a relative decoupling between economic growth and CO2 emissions, with the emissions per unit of GDP falling for the past 60 years. This is the case even in countries like India and China that have been undergoing rapid economic growth. But relative decoupling alone is inadequate in a world where global CO2 emissions need to peak and decline in the next decade to give us any chance at limiting warming to well below 2℃, in line with Paris Agreement targets.

Thankfully, there is increasing evidence that the world is on track to absolutely decouple CO2 emissions and economic growth — with global CO2 emissions potentially having peaked in 2019 and unlikely to increase substantially in the coming decade. While an emissions peak is just the first and easiest step towards eventually reaching the net-zero emissions required to stop the world from continuing to warm, it demonstrates that linkages between emissions and economic activity are not an immutable law, but rather simply a result of our current means of energy production.

In recent years we have seen more and more examples of absolute decoupling — economic growth accompanied by falling CO2 emissions. Since 2005, 32 countries with a population of at least one million people have absolutely decoupled emissions from economic growth, both for terrestrial emissions (those within national borders) and consumption emissions (emissions embodied in the goods consumed in a country). This includes the United States, Japan, Mexico, Germany, United Kingdom, France, Spain, Poland, Romania, Netherlands, Belgium, Portugal, Sweden, Hungary, Belarus, Austria, Bulgaria, El Salvador, Singapore, Denmark, Finland, Slovakia, Norway, Ireland, New Zealand, Croatia, Jamaica, Lithuania, Slovenia, Latvia, Estonia, and Cyprus. Figure 1, below, shows the declines in territorial emissions (blue) and increases in GDP (red).

To qualify as having experienced absolute decoupling, we require countries included in this analysis to pass four separate filters: a population of at least one million (to focus the analysis on more representative cases), declining territorial emissions over the 2005-2019 period (based on a linear regression), declining consumption emissions, and increasing real GDP (on a purchasing power parity basis, using constant 2017 international $USD). We chose not to include 2020 in this analysis because it is not particularly representative of longer-term trends, and consumption and territorial emissions estimates are not yet available for many countries.

There is a wide range of rates of economic growth between 2005-2019 among countries experiencing absolute decoupling. Somewhat counterintuitively, there is no significant relationship between the rate of economic growth and the magnitude of emissions reductions within the group. While it is unlikely that there is not at least some linkage between the two factors, there are plenty of examples of countries (e.g., Singapore, Romania, and Ireland) experiencing both extremely rapid economic growth and large reductions in CO2 emissions.

One of the primary criticisms of some prior analyses of absolute decoupling is that they ignore leakage. Specifically, the offshoring of manufacturing from high-income countries over the past three decades to countries like China has led to “illusory” drops in emissions, where the emissions associated with high-income country consumption are simply shipped overseas and no longer show up in territorial emissions accounting. There is some truth in this critique, as there was a large increase in emissions embodied in imports from developing countries between 1990 and 2005. After 2005, however, structural changes in China and a growing domestic market led to a reversal of these trends; the amount of emissions “exported” from developed countries to developing countries has actually declined over the past 15 years.

This means that, for many countries, both territorial emissions and consumption emissions (which include any emissions “exported” to other countries) have jointly declined. In fact, on average, consumption emissions have been declining slightly faster than territorial emissions since 2005 in the 32 countries we identify as experiencing absolute decoupling. Figure 2, below, shows the change in consumption emissions (teal) and GDP (red) between 2005 and 2019.

There is a pretty wide variation in the extent to which these countries have reduced their territorial and consumption emissions since 2005. Some countries — such as the UK, Denmark, Finland, and Singapore – have seen territorial emissions fall faster than consumption emissions, while the US, Japan, Germany, and Spain (among others) have seen consumption emissions fall faster. Figure 3 shows reductions in consumption and territorial emissions for each country, with the size of the dot representing the size of the population in 2019.

[Chart omitted]

Absolute decoupling is possible. There is no physical law requiring economic growth — and broader increases in human wellbeing — to necessarily be linked to CO2 emissions. All of the services that we rely on today that emit fossil fuels — electricity, transportation, heating, food — can in principle be replaced by near-zero carbon alternatives, though these are more mature in some sectors (electricity, transportation, buildings) than in others (industrial processes, agriculture).