Preliminary and incomplete. Please do not cite without permission.

Designing Advance Market Commitments for New Vaccines

Michael Kremer

Harvard University, Brookings Institution, Center for Global Development, and National Bureau of Economic Research. Address: Department of Economics, Harvard University, Littauer Center 207, Cambridge, MA 02138. Email: mkremer@fas.harvard.edu.

Jonathan Levin

Stanford University Graduate School of Business and National Bureau of Economic Research. Address: Stanford Graduate School of Business 655 Knight Way Stanford, CA 94305. Email: jdlevin@stanford.edu.

Christopher M. Snyder

Dartmouth College and National Bureau of Economic Research. Address: Department of Economics, Dartmouth College, Hanover, NH 03755. Email: chris.snyder@dartmouth.edu.

December 2019

Abstract: Advance market commitments (AMCs) have been proposed as mechanisms to stimulate investment by suppliers of products to low-income countries, where familiar mechanisms such as patents and prizes can fall short. In an AMC, donors commit to a fund from which a specified subsidy is paid per unit purchased by low-income countries until the fund is exhausted, strengthening suppliers' incentives to invest in research, development, and capacity. A $1.5 billion pilot AMC was undertaken to speed the roll out of a pneumococcus vaccine to the developing world covering additional strains prevalent there.

This paper undertakes the first formal analysis of AMCs. We construct a model in which an altruistic donor bargains with a supplier on behalf of a low-income country over vaccine price and quantity ex post, after the supplier has sunk ex ante investments. We use this model to explain the broad logic of an AMC—as a solution to a hold-up problem—as well as to analyze specific features of the pilot's design that we argue enhance its efficiency. We study a variety of design features including capacity forcing, supply commitments, price ceilings, and accrued interest, and consider a variety of economic environments including competing suppliers, competing demand from middle-income countries outside the program. We show that optimal AMC design differs markedly depending on where the product is in its development cycle.

JEL codes: O31, D02, O19, I18

Acknowledgments: The authors served as members of the AMC Economics Expert Group, which participated in designing the pilot program for the pneumococcus vaccine. The views expressed here are their own. The authors are grateful for helpful comments provided by Eric Edmonds, Richard Gilbert, Daniel Hosken, Scott Kominers, Ruth Levine, Gerard Roland, and seminar participants at the Center for Advanced Study in the Behavioral Sciences at Stanford University, Dartmouth College, M.I.T., University of Washington, the Conference on Development Economics and Market Design at Harvard Business School, the Conference in Honor of Eric Maskin at Harvard University, the IO Theory Conference at M.I.T., the International Industrial Organization Conference session on "Innovation and Technology Adoption in Healthcare," and the Market Shaping Conference at Stanford University. The authors are grateful to Egor Abramov and Kevin Xie for excellent research assistance. Levin thanks the Center for Advanced Study in the Behavioral Sciences for its hospitality and Snyder the Center for Global Development for generous funding.

1. Introduction

Familiar mechanisms such as patents and prizes, which work well to stimulate research and development (R&D) for products sold in high-income markets, may fall short for products sold in low-income markets. Patents generate deadweight loss along with monopoly rents, and monopoly rents may be limited in countries with mostly poor consumers, particularly if the country or aid agency acting on its behalf ignores these patents or uses bargaining power or public pressure to push down prices. Prizes may lead to the development of products that, while meeting the letter of the competition's technical specifications, do not meet the consumers' true needs.

The difficulty in meeting the needs of poor countries is particularly apparent in the market for vaccines. Vaccines are a highly cost-effective tool to improve global public health. Yet the provision of vaccines in poor countries already in widespread use in rich countries and the development of vaccines targeting diseases of poor countries have both been disappointingly slow. Adoption of vaccines developed in the 1980s, such as Hepatitis B and Hib, on childhood immunization schedules of most rich countries, has only begun in poor countries in the last few years. Highly effective vaccines for diseases that are more prevalent in poor countries such as malaria, tuberculosis, yellow fever, and HIV, await development. This situation has sparked a host of recent initiatives to catalyze developing country vaccine markets. Prominent among these was the piloting of an Advance Market Commitment (AMC) for pneumococcus vaccine and the possibility of future AMCs for other vaccines.

The idea of using AMCs for vaccines was proposed by Kremer and Glennerster (2004) and Kremer, Barder, and Levine (2005). In advance of firms' investments in R&D, donors pledge a fund to subsidize initial purchases of a newly introduced vaccine above and beyond unit production costs. This feature helps overcome the concern on the part of firms that they will not be able to recoup the large investments required to develop a new vaccines for poor countries, faced with consumers who lack sufficient income to constitute significant demand on their own and aid agencies liable to use their bargaining power to "hold up" the firms' investments, pushing the price toward unit production cost in ex post negotiations. The "market" component of the AMC is that the subsidy is only triggered when the low-income country makes a purchase decision, requiring it to make a small copayment. This feature helps avoid an outcome in which the firm produces a product meeting the technical specifications but not delivering much consumer utility for reasons that are hard to specify in a contract.

Since donors' budget rules typically do not allow indefinite commitments, and since health needs and opportunities are subject to change over time, in practice donors will not agree to subsidize vaccine purchases indefinitely, leaving open the question of market design in the so-called "tail period," after the AMC fund is exhausted. The original AMC design called for firms to offer the vaccine at a lower, sustainable price in the tail period as *quid pro quo* for the subsidy received. This feature helps overcome the concern on the part of poor countries and aid agencies that they adopt a vaccine that in the long run becomes prohibitively priced. In principle, an AMC allows companies to realize an early economic return on their investment while avoiding deadweight loss from monopoly pricing in adopting countries over a longer horizon.

While the original idea of an AMC focused on technologically distant projects, with the goal of stimulating the substantial additional R&D required for a marketable product, donors were interested in achieving a quicker "win" and chose to conduct a pilot AMCs focused on a technologically closer product. In 2007, five countries (Italy, Canada, Russia, Norway and the United Kingdom) in conjunction with the Gates Foundation pledged $1.5 billion toward a pilot AMC for a pneumococcus vaccine. Pneumococcus is the most common cause of severe pneumonia worldwide, and also a cause of meningitis, septicemia and ear infections. In rich countries, it is treated with antibiotics and typically causes little mortality. Worldwide, however, it kills over 700,000 children under five each year (World Health Organization 2007). A first-generation pneumococcus vaccine, Prevnar, covering seven strains of the disease was in widespread use in rich countries. Several firms had a second-generation vaccine covering additional strains in late-stage clinical trials. The second-generation is marginally more effective than the first- generation in rich countries but substantially so in poor countries because the additional strains are more prevalent there (Levine 2006, Sinha *et al*. 2007). The pilot AMC targeted this second-generation vaccine. The goals of the AMC included not just stimulating the remaining product development needed, but also improving access by speeding the installation of the capacity needed to serve the low-income market (as much as 210 million additional doses each year if universal vaccination were to be achieved), and to The three authors of this paper worked on the pilot AMC as part of the Economics Expert Group. The Group was presented with an AMC designed according to the specifications of the Framework Document (World Bank and GAVI 2006) and asked to specify parameters such as the subsidy rate, price cap during the tail period, inflation indexing, and so forth. The Economics Expert Group ended up suggesting modifications to the framework AMC, which were incorporated into the pilot program.

This paper provides the first formal analysis of AMCs, complementing the earlier conceptual analyses cited above. Besides solidifying the broad logic behind AMCs, a formal model facilitates the evaluation of alternative program designs and characterization of optimal parameters for those policies. Although we are chiefly interested in making general points about AMCs, we will spend some time focusing on the design of the pneumococcus pilot, examining whether the modifications to the framework design suggested by the

Economics Expert Group have a theoretical justification, and whether further modifications might result in further improvements.

A broad theme emerging from the analysis is that the optimal AMC design can differ dramatically depending on how far along the product is in its development cycle. For a technologically close product for which most of the required R&D investment has already been sunk, the target moves to incentivizing investment in the dedicated capacity needed to serve low-income countries. We show that subtle changes in AMC design can have large effects on capacity investment. The framework design, which subsidizes initial purchases using a fixed fund, which may do a good job stimulating R&D investment, does not stimulate much additional capacity investment unless there is intense supplier competition. Indeed, the AMC has no effect on a monopoly's incentives if the fund accrues interest. In monopoly and other less-competitive environments, better incentives are provided if the framework AMC is designed not to accrue interest, better still by a so-called supply-commitment design that reduces the size of the AMC fund in proportion to unfilled demand requirements. This result provides theoretical support for the Economics Expert Group's suggestion to move to a supply commitment for the pilot AMC, as they predicted that only one or two suppliers would initially be viable in that market. We show that capacity incentives can be further improved by moving to a capacity-forcing agreement that takes away the firm's discretion over capacity, possibly attaining the first best even for a monopoly supplier.

Our analysis of AMCs for technologically distant products emphasizes three ways in which they differ from AMCs for technologically close products. First, the AMC needs to induce R&D in addition to capacity investment, calling for larger subsidies. Second, the firm may have acquired less private information about production costs early in the product's development. This second effect helps with the efficiency AMC design because, from a mechanism-design perspective, less private information means a smaller information rent that has to be paid to the firm qua agent. Third, a complete contract over the product's technical specifications is difficult to write far in advance of product development, reinforcing the role of a country copayment as a sort of "kill switch" for products meeting the specifications but providing little consumer utility for numerous reasons that are hard to anticipate.

We start with a simple model with three players: a firm, a poor country, and a donor which makes vaccine purchases on behalf of the country. The timing is as follows. First, the donor specifies the terms of the AMC or other commitment contract. Next, the firm (initially a monopoly for simplicity) makes investments, which can include investments in R&D in the case of a technologically distant product or just in capacity in the case of a technologically close product. This leads to an ex post period during which the donor and firm bargain over vaccine supply each instant of continuous time. Our experience working on the pilot AMC strongly suggested that bargaining is an important, realistic element to include in the model. The absence of an AMC or other ex ante policy does not rule out the supply of the vaccine ex post, perhaps at some diminished capacity. Nor does the existence of an AMC preclude either the firm or the donor from making further demands or concessions later on. This leaves open the possibility that the AMC completely unravels, with the donor offering a subsidy ex ante with its left hand only to claw it back ex post with its right. We will indeed see that certain AMC designs provide no investment incentives for this reason.

Regarding the related literature, the previously cited work that proposed the AMC idea was largely conceptual. Berndt *et al.* (2007) provide empirical forecasts of costs and effectiveness of AMCs for neglected diseases. Snyder, Begor, and Berndt (2011) discusses the pneumococcus pilot. Although it does not provide formal propositions, it provides a table of calibrated results (in fact based on a model from an uncirculated draft of this paper).

The plan of the paper is a follows. The next section presents the model. Section 3 begins the analysis with the basic case of a monopoly firm producing a technologically close product. The section examines a series of increasingly efficient policies, from a framework AMC to a supply commitment to a forcing contract. The section ends by cataloging a variety of refinements that might be added to these basic policies such as price caps, purchase guarantees, escrowed interest, procurement through an agent, and country copayments. While analysis of these refinements does contribute to a conceptual understanding of AMCs, the main contribution is to raise considerations of practical policy interest; so for space considerations we relegate the details of the analysis to the appendix. Section 4 extends the analysis to allow for asymmetric information about the firm's costs. Section 5 extends the analysis from a monopoly to allow for an arbitrary number of firms. Section 6 moves the focus from a technologically close product for which R&D costs have not yet been sunk. The last section concludes with a summary of the main theoretical points and a discussion of policy implications, both for the pros and cons of the design of the pilot AMC for pneumococcus as well as for future AMCs. Appendix A provides proofs omitted from the text. Appendix B provides a full analysis of the refinements of the basic AMC cataloged at the end of Section 3.

2. Model

The model has two periods, ex ante and ex post, the dividing line being the point at which the firm sinks any required investments. The ex post period involves continuous time indexed by t. Let $r > 0$ be the market interest rate as well as all players' internal discount rates.

The model has three players: a firm, country, and donor. A profit-maximizing firm has the opportunity to develop a new product that it can supply to a representative low-income country. The analysis will initially

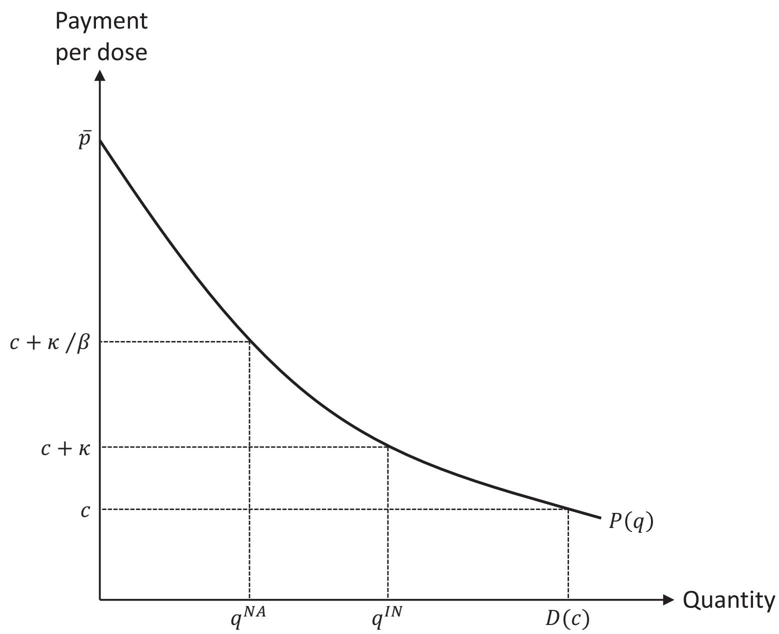

Figure 1: Donor inverse demand

focus on a monopoly firm; Section 5 extends the model to imperfect competition. For concreteness we will take the product to be a vaccine, though more generally it could represent any new good or service. The country obtains health benefit $B(q)$ from the vaccination of q of its citizens. This flow benefit can be realized each t because a new cohort of unvaccinated newborns and immigrants flows into the country each instant. Normalize $B(0) = 0$. Assume $B(q)$ is twice continuously differentiable, increasing, and strictly concave for all $q \geq 0$. To capture the idea that the health benefit eventually levels off as most of the population becomes served, assume $\lim_{q\to\infty} B'(q) = 0$.

Suppose initially the country cannot pay for the vaccine itself, either because it is too poor to afford any payment or because its leaders, who may or may not be benevolent, have other priorities. (Section 6.1 and Appendix B extend the model to allow for country contributions.) A donor, who could represent one or more non-governmental organizations and/or high-income countries with development-aid budgets, has regard for the low-income country's welfare. Its flow utility is $\alpha B(q) - X$, where $\alpha \in (0,1)$ represents the degree of donor altruism and $X \geq 0$ is its expenditure on vaccines each instant t. Let $P(q) = \alpha B'(q)$ denote the donor's marginal willingness to pay for the vaccine, in effect its inverse demand curve, illustrated in Figure 1. (Some features of the figure will not be introduced until later sections.)

Some pedantry concerning the demand curve up front will help streamline the analysis in later sections. The strict concavity of $B(q)$ implies $P'(q) = \alpha B''(q) < 0$ for all $q \geq 0$. Hence inverse demand slopes down. Let $\bar{p} = P(0)$ denote the choke price. That $B(q)$ is twice continuously differentiable at $q = 0$ implies

$P'(0) = \alpha B''(0)$ exists, implying $P(0)$ is finite. Hence $\bar{p}$ is finite. Let $\tilde{D}(p) = P^{-1}(p)$ denote the inverse of $P(q)$. The following lemma establishes useful properties of this inverse.

Lemma 1. *For all $p \in (0, \bar{p})$, $\tilde{D}(p)$ is a finite, positive number, and $\tilde{D}'(p) < 0$.*

The proof, relying on the inverse function theorem, is provided in the appendix. Regarding the endpoints of the interval, it is easy to see that $\lim_{p \to 0} D(p) = \infty$ and $\tilde{D}(\bar{p}) = 0$. Define the demand curve $D(p)$ to be the following extension of $\tilde{D}(p)$ to all positive prices, i.e., $D(p) = \tilde{D}(p)$ for $p \in (0, \bar{p})$ and $D(p) = 0$ for $p \geq \bar{p}$. The properties of $\tilde{D}(p)$ established in the lemma are inherited by $D(p)$ for all $p \in (0, \bar{p})$.

The firm has three sources of cost. Let R be the research and development cost that must be sunk at time 0 for the product to exist. Let KQ be the cost that must be sunk at time 0 to install capacity Q, where $K > 0$. Notice that R is a fixed cost, independent of capacity or output, but KQ is increasing in capacity (indeed linear for simplicity). It will be sometimes convenient to express these sunk costs, which are stocks, as flows to facilitate comparison to other flows in the model. To this end, let $\rho = rR$ and $\kappa = rK$ be the per-period payments to finance the research and development and unit capacity investments, respectively. Once these sunk investments are in place, the cost of producing at rate $q \leq Q$ each period is cq, where $c > 0$ is the marginal cost of producing a unit in a period.

We will analyze a series of different AMC designs, but to fix ideas here we will describe the first one we will analyze, the framework AMC. In the framework AMC, the donor sets two terms ex ante, a fund size F and a per-unit subsidy s. Imparting commitment power to the AMC, the fund is assumed to be locked in an escrow that cannot be used for any other purpose.[1] As a baseline, we will specify that interest earned by the escrow accrues there and, as with the principal in the escrow account, cannot be used for any other purpose. (We will also analyze the variant in which interest earned by the escrow flows back to the donor.) The firm receives the subsidy s for each unit purchased by the country until the fund is exhausted. Let T denote the time at which this happens. Time interval $(0, T)$ will be called the AMC period and (T, ∞) the tail period. The framework AMC can be viewed as contract that is linear in output. Nonlinear contracts (e.g., forcing contracts) are also possible, as are contracts that are functions of variables other than output (e.g, capacity). Our approach will be to analyze various contractual forms and discuss the assumptions on contractual completeness and/or commitment power required for the form to be feasible in the relevant

[1] Looking inside the black box of the escrow, the assumption that the escrow can be committed to a single purpose can be viewed as prohibiting the escrow administrator from bilaterally renegotiating with the donor ex post over the use of the funds. Allowing the firm to bring a legal action against such a reallocation would restore the commitment value of the escrow. If the escrow funds are not committed to a single purpose and instead are fungible for the donor, one can show that the framework design of the AMC provides no investment incentives. As discussed in Appendix B, having an escrow in the model under some circumstances is equivalent to having a model in which the donor is split into two players, a principal who donates the funds and designs the AMC and an agent who carries out the ex post bargaining. The latter model resembles the structure of the pilot AMC for pneumococcus.

passages below.

There are a variety of ways to model the process of price formation and purchasing ex post. We assume that once the firm sinks its investments, it engages in Nash bargaining with the donor over the sequence of future vaccine purchases. The hold-up problem arises from the fact that bargaining takes place after investments have been sunk. Using a bargaining framework to model vaccine purchases has several virtues. Besides hold up, it builds in elements of bilateral monopoly, both important and realistic features of markets in which AMCs might be used. The absence of an AMC does not preclude the possibility of trade, nor does its presence preclude the possibility of bargaining under its shadow. Thus, while the AMC will have scope to affect equilibrium outcomes, it will not be through exogenous assumptions on the effect of the AMC on the structure of price formation. Let $\beta \in (0,1)$ be the firm's bargaining weight and $1-\beta$ the donor's.

A complete formulation of Nash bargaining requires several additional details to be delineated. We assume that parties immediately engage in a once-for-all Nash bargain at the start of the ex post period, at $t = 0$, covering the entire sequence of vaccine prices and quantities each instant t of continuous time, to which they can commit. Even though bargaining entails commitment, this does not automatically solve the hold-up problem because bargaining happens after investment. Dynamic-programming arguments suggest that—in a setting covering many cases analyzed below, involving efficient bargaining, freely transferable surplus, and perfect certainty—the present discounted value of each player's surplus resulting from the once-for-all Nash bargain at $t = 0$ is identical to that resulting from the subgame perfect equilibrium of the continuation game with individual Nash bargains each instant t. We adopt the perspective of once-for-all rather than instant-by-instant bargaining throughout the analysis because of its mathematically tractability. Another detail requiring further delineation concerns parties' threat points. Parties' threat points depend on the policy in force at $t = 0$. In the absence of an AMC or other commitment, we assume trade requires the active assent of both parties, implying that parties obtain 0 surplus in the continuation game following a breakdown of bargaining. The threat point for even as narrow a policy as a framework AMC is not well-defined without further specification. We assume as a baseline that AMCs (and other policies we will study) require the active assent of both bargaining parties for trade to occur, just as in the absence of an AMC. A later section will be devoted to the analysis of the alternative in which the AMC empowers the firm to trade at the specified terms without the assent of the donor, labeled a purchase guarantee.

The timing of the model can be summarized as follows. The ex ante period begins with the donor setting the terms of the AMC. Observing this, the firm sinks any required investment in research, development, and/or capacity. This leads to the ex post period, starting at time 0. The donor and firm engage in Nash bargaining over the supply of q units of vaccine to the country up to capacity Q. The once-for-all bargain

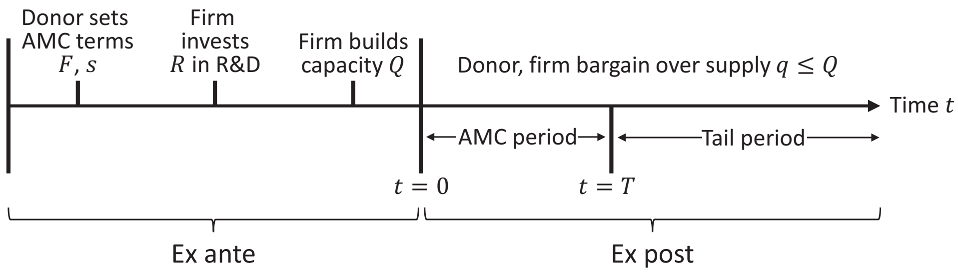

Figure 2: Timeline for model

at time $t = 0$ covers the sequence of vaccine prices and quantities for all t thereafter. Figure 2 provides a timeline for the events in the model.

To rule out trivial cases, we maintain the following condition throughout the analysis.

Assumption 1. *$c+\kappa < \bar{p}$.*

We will see in the analysis of the integrated benchmark, in which the donor is allowed to control the operations of the firm directly, that optimal capacity and output is $D(c+\kappa)$. Assumption 1 combined with Lemma 1 implies $D(c+\kappa) > 0$. Hence the integrated donor would install some capacity. If the assumption did not hold, then not only the integrated benchmark but all equilibria in which the donor bargains with a separate firm would involve the trivial outcome of no capacity or supply.

3. Technologically Close Products

We begin the analysis with the simple case of a near-term product, specifically a vaccine that is far enough in research and development before the AMC is implemented that the only substantial sunk investment required is the expenditure KQ for capacity, not R&D expenditure R. Assume $K > 0$, implying $\kappa = K/r > 0$. The product is also far enough along that its attributes can be specified fairly completely. For simplicity, we will take the extreme case in which these attributes can be specified perfectly, implying that ex ante commitments can specify ex post payments conditional on q (in short, q is contractible).

To facilitate the analysis, before analyzing different AMC designs, we study two benchmarks at the extremes of the range of possibilities. We first analyze the extreme with the greatest incentive for capacity investment: a counterfactual integrated outcome in which the donor directly controls the firm's operations. We then analyze the extreme with the least incentive for capacity investment: the outcome in the absence of integration and in the absence of an AMC. The outcome with an AMC will fall somewhere between these two benchmarks. We start our analysis of AMCs in Section 3.4 with the framework design, and then analyze possible improvements moving to supply commitments and forcing contracts.

3.1. Integrated Benchmark

In this subsection we analyze the integrated benchmark in which the donor can directly control the operations of the firm. Let Q^{IN} be the capacity and q^{IN} the flow quantity each instant that the donor chooses in this benchmark. These variables are set to maximize the present discounted value of benefits minus costs

$$\int_0^\infty [\alpha B(q) - cq] e^{-rt} dt - KQ = \frac{1}{r}[\alpha B(q) - cq - \kappa Q] \tag{1}$$

subject to

$$q \leq Q. \tag{2}$$

The constraint obviously binds. Substituting for q into (1) treating (2) as an equality yields

$$\frac{1}{r}[\alpha B(Q) - (c+\kappa)Q] \equiv w(Q,\theta). \tag{3}$$

Because this expression for joint welfare reappears in the analysis, we have introduced the shorthand $w(Q,\theta)$ for it, where $\theta = (r,\alpha,c,\kappa,\beta)$ is the vector of model parameters.

Taking the first-order condition of (3) with respect to Q yields

$$P(Q) - (c+\kappa) = 0, \tag{4}$$

implying $Q^{IN} = q^{IN} = D(c+\kappa)$. The following proposition summarizes this analysis.

Proposition 1. *Capacity and output for a monopoly firm in the integrated benchmark are given by* $Q^{IN} = q^{IN} = D(c+\kappa) > 0$.

Proof. The equalities were established in the text leading up to the proposition. The last inequality follows from Assumption 1 and Lemma 1. See Section 2 for the argument. *Q.E.D.*

3.2. Benchmark Without AMC

We next analyze the benchmark in the absence of integration and in the absence of an AMC. We solve for the equilibrium using backward induction starting from the ex post period.

Consider the Nash bargain between donor and firm at $t = 0$ over vaccine supply for all future t. The assumption that trade does not occur unless both parties assent means, as discussed in Section 2, that both parties' threat points are 0. Gains from trade equal the present discounted value of joint surplus over the continuation game

$$\int_0^\infty [\alpha B(q) - cq]e^{-rt}dt = \frac{1}{r}[\alpha B(q) - cq]. \tag{5}$$

Because Nash is efficient bargaining, parties settle on the q maximizing (5) subject again to the constraint (2) that output not exceed capacity. If constraint (2) binds, then $q^{NA} = Q$. If not, q^{NA} satisfies the first-order condition of (5) with respect to q: $\alpha B'(q) = c$, or rearranging $q = D(c)$.[2] Combining the cases in which (2) does or does not bind yields

$$q^{NA} = \min[Q, D(c)]. \tag{6}$$

Given both parties' threat points are 0, the firm's continuation payoff from Nash bargaining is its share β of gains (5) and the donor's is its share $1-\beta$.

Fold the game back to the firm's ex ante choice of capacity. The firm's equilibrium capacity choice Q^{NA} maximizes the present discounted value of profits

$$\frac{\beta}{r}[\alpha B(q^{NA}) - cq^{NA}] - KQ = \frac{\beta}{r}\left\{\alpha B(\min[Q, D(c)]) - c\min[Q, D(c)] - \frac{\kappa}{\beta}Q\right\}, \tag{7}$$

where the equality follows from substituting from (6) and rearranging. Posit that $Q^{NA} < D(c)$; the proof of the next proposition will fill in this detail. Substituting Q for $\min[Q, D(c)]$ in (7) yields

$$\frac{\beta}{r}\left[\alpha B(Q) - \left(c + \frac{\kappa}{\beta}\right)Q\right] \equiv \pi(Q, \theta). \tag{8}$$

Because this expression for the firm's ex ante profit in the absence of an AMC appears repeatedly in the analysis, we have introduced the shorthand $\pi(Q, \theta)$ for it, where θ is the vector of model parameters. Differentiating (8) with respect to Q yields the first-order condition

$$P(Q) - \left(c + \frac{\kappa}{\beta}\right) = 0. \tag{9}$$

Equilibrium capacity Q^{NA} is the solution of (9) for Q, i.e., $Q^{NA} = D(c + \kappa/\beta)$.

Since demand is nondecreasing, $D(c+\kappa/\beta) \le D(c+\kappa)$. The proof of the next proposition shows that this inequality is in fact strict, implying $Q^{NA} = D(c+\kappa/\beta) < D(c+\kappa) = Q^{IN}$. That $Q^{NA} < Q^{IN}$ is a key result of the paper, embodying the hold-up problem which AMCs are designed to address. The hold-up problem arises because returns from the firm's sunk investment in capacity are appropriated by the donor in ex post negotiations unless the firm has 100% of the bargaining power. In the absence of an AMC, this leads the

[2] To see this is a well-defined quantity, note $0 < c < \bar{p}$, where the second inequality holds by Assumption 1. Lemma 1 then implies $D(c)$ is a finite, positive number.

firm to underinvest in capacity compared to the integrated benchmark. In the limit $\beta \to 0$ in which the firm has no bargaining power, the hold-up problem is so severe that the firm installs no capacity. In the limit $\beta \to 1$ in which the firm has all the bargaining power, there is no hold-up problem and the donor. The next proposition summarizes the results of this subsection.

Proposition 2. *Capacity and output for a monopoly firm in the benchmark with no AMC are given by* $Q^{NA} = q^{NA} = D(c+\kappa/\beta)$ *and the ex ante present discounted value of the firm's profits by* $\pi(Q^{NA},\theta)$*. Comparing capacity to that in the integrated benchmark, we have* $Q^{NA} < Q^{IN}$, $\lim_{\beta\to 0} Q^{NA} = 0$*, and* $\lim_{\beta\to 1} Q^{NA} = Q^{IN}$.

The proof in the appendix fills in some omitted details.

Figure 1 illustrates equilibrium quantity in the two benchmarks. Integrated quantity $D(c+\kappa)$ is that emerging from marginal-cost pricing—generalizing marginal cost to include both the unit cost of production c and that unit's share κ of the cost of capacity. Quantity in the no-AMC benchmark $D(c+\kappa/\beta)$ is that emerging from marking up the cost subject to hold up in inverse proportion to the firm's bargaining power. The production cost c is not marked up because that is an ex post cost that can be handled by Nash bargaining, which is ex post efficient.

3.3. Principle of Full Capacity Utilization

Neither benchmark analyzed in the previous two subsections involved excess capacity. The key to this result is that both are ex post efficient, a consequence of integration in one case and Nash bargaining in the other. The ex post efficient outcome is for all capacity Q being used as long as the marginal benefit $\alpha B'(Q)$ exceeds the marginal production cost c, or rearranging, as long as $Q \le D(c)$. But $D(c)$ is more capacity than even an integrated firm would install facing any positive capacity cost κ. Hence all capacity is used each instant.

This insight will apply more generally to all the equilibria we will go on to derive in the remainder of this section. A positive capacity cost will induce the firm to install less capacity than $D(c)$; efficient ex post bargaining will then lead to all capacity being utilized each instant. We will invoke this result in the remainder of the section using the shorthand *principle of full capacity utilization*. The principle has the consequence that any subsequent results for equilibrium capacity Q^* will equivalently apply to equilibrium flow quantity q^*.

Deriving the principle for the integrated and no-AMC benchmarks was easy because changes in the quantity traded in any instant had no external effects on the surplus to be shared in future instants. A caveat in applying the principle to the case of a AMC is that producing more in the current instant draws the fund down, reducing the surplus that can be shared later. We will be able to invoke the principle nonetheless, but care will be needed in making the argument.

3.4. Framework AMC

The first AMC design we will analyze is the framework design, involving a fund F placed in escrow ex ante from which a subsidy s is paid for each unit of the vaccine purchased ex post until the fund is exhausted. We start by positing that the duration of the AMC period T is finite. Then T can be determined by an accounting identity. Given our baseline specification for AMC design that interest on funds accrues to the escrow, the accounting identity is that the present discounted value of payments into the fund equals the present discounted value of payments out of the fund:

$$F = \int_0^T sqe^{-rt}dt = \frac{sq}{r}\left(1-e^{-rT}\right). \tag{10}$$

The left-hand side of (10) is the size of the fund itself, which since it is made in a lump-sum payment ex ante is the same as its present discounted value. The right-hand side is the present discounted value of the flow of subsidy payments sq made each instant during the AMC period assuming assuming q units are sold each instant.

We again solve for the equilibrium using backward induction starting from Nash bargaining at $t = 0$. The model requires both parties to assent to trade during the AMC period and tail period, implying that their threat points are again 0. Gains from trade equal the present discounted value of joint surplus over the continuation game

$$\int_0^\infty \alpha B(Q)e^{-rt}dt + \int_0^T (s-c)Qe^{-rt}dt - \int_T^\infty cQe^{-rt}dt = \frac{1}{r}[\alpha B(Q) - cQ + sQ(1-e^{-rT})]. \tag{11}$$

The first term on the left-hand side is the donor's gross surplus (not including bargaining payments) over the whole ex post period, the second term is the firm's gross surplus (again not including bargaining payments) during the AMC period, and the third is the firm's gross surplus during the tail period. The right-hand side of (11) follows from integrating and rearranging. Several additional notes about (11) are in order. Note that s appears in the second term. While one might think it should net out of joint surplus as a mere transfer, because it was sunk in a dedicated escrow, ex post subsidy payments are not subtracted from the donor's contribution to joint surplus. Note further that capacity Q has been substituted for output in (11) on the basis of the principle of full capacity utilization. This principle obviously holds in the tail period because reducing output below Q has no later benefits. Work is required to verify that the principle holds during the AMC period because a reduction in output in one instant leaves more funds in the escrow, providing the future benefit of extending the AMC period. We relegate this verification to the proof of Proposition 3 in the appendix. For now, we just note that it is intuitive that the principle holds during the AMC period. Deferring the draw down of AMC funds by reducing output has a neutral effect on joint surplus because the funds accrue interest at the same in or out of the esrow. The only remaining effect of an output reduction is the inefficiency from producing less than the capacity-constrained optimum, Q.

Fold the game back to the firm's ex ante choice of capacity. The firm's equilibrium capacity choice Q^{TS} maximizes the present discounted value of its profits. (The superscript denotes equilibrium under a temporary-subsidy policy under consideration.) Given its threat point is 0, the its continuation payoff from Nash bargaining is its share β of the gains in (11). Subtracting capacity costs yields the following expression for ex ante profits,

$$\frac{\beta}{r}[\alpha B(Q) - cQ + sQ(1 - e^{-rT})] - KQ, \tag{12}$$

or, upon substituting from (10) into (12) and rearranging,

$$\frac{\beta}{r}\left[\alpha B(Q) - \left(c + \frac{\kappa}{\beta}\right) Q\right] + \beta F = \pi(Q, \theta) + \beta F. \tag{13}$$

Differentiating (13) with respect to Q yields $Q^{TS} = D(c + \kappa/\beta) = Q^{NA}$. We have obtained a stark result: the donor receives nothing for the funds F it contributes to the AMC, capacity and output are the same with a framework AMC as without. Mathematically, the contribution of the AMC to the firm's surplus, which is embodied in equation (13) by the term βF, is not a function of Q. Intuitively, if the firm reduces its capacity, it draws down the AMC at a slower rate, but this leaves more funds in the escrow account accruing interest at the same rate as the firm's own bank account. The firm ends up being indifferent as how fast it draws down the AMC fund. The AMC thus contributes nothing to incentives to invest in capacity.

Proposition 3. *Consider any terms $s > 0$ and $F > 0$ for which the framework AMC has temporary subsidy. The framework AMC adds no incentives for a monopoly firm to invest in capacity: $Q^{TS} = q^{TS} = D(c + \kappa/\beta) = q^{NA} = Q^{NA}$.*

The proof provided in the appendix fills in the detail about the principle of full capacity utilization applying to the AMC period.

It goes without saying that the donor would never make a positive contribution F toward this useless program. The result that an AMC with a temporary subsidy is useless leaves open the possibility that an AMC with a perpetual subsidy may be useful. With a temporary subsidy, an increase in output in instant t merely shifts the timing of the subsidy payouts with no net effect on the total earned from the program and thus no effect on incentives. With a perpetual AMC, a current increase in output does not reduce the subsidy that can be earned in the future—promised to be s per unit regardless of how much was produced before.

Hence a perpetual subsidy may enhance investment incentives. In fact we will show next that it does.

A perpetual subsidy requires a sufficiently large endowment F given s and q, namely $rF \geq sq$. The donor prefers the smallest fund satisfying that condition, implying the donor's optimum involves

$$F = \frac{sq}{r}. \tag{14}$$

We are left to solve for the optimal s. We again use backward induction, starting from the ex post Nash bargain at $t = 0$ over vaccine supply, positing some arbitrary s set by the donor.

For Nash bargaining to lead to trade, both parties must assent, implying that their threat points are 0 as before. Gains from trade equal the present discounted value of joint surplus over the continuation game

$$\int_0^{\infty} [\alpha B(Q) - cQ + sQ] e^{-rt} dt = \frac{1}{r}[\alpha B(Q) - cQ + sQ]. \tag{15}$$

We have substituted capacity Q for quantity, invoking the principle of full capacity utilization, which obviously applies in this case. As was the case above with a temporary subsidy, here, too, s appears in the expression for continuation surplus because the subsidy is paid out of a dedicated escrow, the funds in which do not have an opportunity cost to the donor.

Folding the game back to the ex ante period, the firm's equilibrium capacity choice maximizes its share β of (15) less capacity costs, which upon rearranging becomes

$$\frac{\beta}{r}\left[\alpha B(Q) - \left(c + \frac{\kappa}{\beta}\right) Q + sQ\right] = \pi(Q, \theta) + \frac{\beta s Q}{r}. \tag{16}$$

The firm's investment decision is characterized by the first-order condition with respect to Q,

$$P(Q) - \left(c + \frac{\kappa}{\beta}\right) + s = 0. \tag{17}$$

Equation (17) can be readily compared to the first-order condition (9) from the benchmark with no AMC. The perpetual subsidy simply results in s being added to this first-order condition, enhancing incentives when $s > 0$.

Folding the game back, the donor chooses the AMC terms to maximize the present discounted value of its $1-\beta$ share of the joint surplus (15) from Nash bargaining less the fund F endowing the perpetual subsidy:

$$\frac{1-\beta}{r}[\alpha B(Q) - cQ + sQ] - F = \frac{1}{r}\left\{(1-\beta)[\alpha B(Q) - cQ] - \beta s Q\right\}. \tag{18}$$

The right-hand side of (18) follows from substituting for F from (14) and rearranging. The donor's problem is to choose s to maximize (18), where Q is implicitly a function of s via (17), which is in essence an incentive-compatibility constraint. The one-to-one correspondence between capacity and the subsidy implicit in (17) allows us to express the donor's problem equivalently in terms of choice variable Q rather than s, yielding considerable insight. Solving (17) for s in terms of Q, substituting into (18), and rearranging gives the following expression for the donor's objective function:

$$\frac{1}{r}[(1-\beta)\alpha B(Q)+\beta QP(Q)-(c+\kappa)Q]. \tag{19}$$

Differentiating (19) with respect to Q yields the first-order condition characterizing equilibrium capacity,

$$(1-\beta)P(Q)+\beta[P(Q)+QP'(Q)]-(c+\kappa)=0. \tag{20}$$

This differs from the first-order condition (4) from the integrated benchmark. Because the integrated entity appropriates all surplus, (4) traded off the marginal social benefit of an additional unit of capacity, $P(Q)$, against marginal cost $c+\kappa$. Equation (20) involves the same marginal cost but on the benefit side incorporates a weighted average of marginal social benefit and marginal private monopsony benefit, with the weights given by the parties' bargaining powers. The greater the donor's bargaining power, i.e., the closer is $1-\beta$ to 1, the more inclined is the donor to use the subsidy to incentivize socially efficient production because it can appropriate more of the generated surplus in ex post Nash bargaining. If the donor has little bargaining power, i.e., β is close to 1, it shifts its focus to the threat point because that becomes its only source of surplus. In the threat point, there are no further payments made beyond the subsidy, so the donor is in effect a linear pricing monopsonist, with the subsidy functioning as the linear price. For moderate bargaining powers, the donor considers a weighted average of the gains from bargaining and its threat point.

Solving the first-order condition leads to the following implicit solution for equilibrium capacity,

$$Q^{PS}=D(c+\kappa-\beta Q^{PS}P'(Q^{PS})). \tag{21}$$

Because $-\beta P'(Q^{PS})Q^{PS}>0$ for $\beta>0$, $D(c+\kappa-\beta P'(Q^{PS})Q^{PS})<D(c+\kappa)=Q^{IN}$ then. Hence, if the firm has any bargaining power, the perpetual subsidy cannot achieve the integrated outcome. The firm puts some weight on the choice that would be made by an inefficient linear-pricing monopsonist. We have proved the following proposition.[3]

[3]Comparison of Propositions 3 to 4 reveals an underlying discontinuity in outcomes. In particular, holding the subsidy rate

Proposition 4. *The framework AMC that is optimal for the donor involves a perpetual subsidy,* $s^{PS} = c+\kappa/\beta - P(Q^{PS})$*, where* Q^{PS} *is the monopoly firm's equilibrium capacity under this policy that is given by the implicit solution to* (21)*. For all* $\beta > 0$*,* $Q^{PS} < Q^{IN}$*, but* $\lim_{\beta\to 0} Q^{PS} = Q^{IN}$*.*

Equilibrium capacity can be alternatively expressed in terms of the familiar Lerner index:

$$L^{PS} = \frac{P(Q^{PS}) - (c+\kappa)}{P(Q^{PS})} = -\beta \frac{Q^{PS}P'(Q^{PS})}{P(Q^{PS})} = \frac{\beta}{|\eta^{PS}|}. \tag{22}$$

The first equality reflects our definition of the Lerner index in this context, the second equality follows from rearranging (20), and the third defining η^{PS} to be the elasticity of the donor's vaccine demand at the equilibrium quantity. If $\beta = 0$, the equilibrium is equivalent to efficient, i.e., marginal-cost, pricing. As β grows, the donor increasingly behaves as a standard monopsonist, adhering to an inverse-elasticity rule. As will be seen momentarily, there is a limit to how close the donor comes to implementing an inverse-elasticity rule because a framework AMC does not exist for β sufficiently close to 1.

The subsection concludes with a more technical discussion of existence of a framework AMC. Call an AMC *nontrivial* if it strictly improves the donor's surplus relative to the equilibrium in the absence of an AMC. Call an AMC *feasible* if the firm would be willing to sign it or in other words it satisfies the firm's participation constraint. So far in this subsection we have ignored the firm's participation constraint but focus on it now. Because the donor cannot commit not to bargain if the firm rejects the AMC, the participation constraint is that the firm must earn at least the ex ante profit Π^{NA} as in the absence of an AMC. One can show that this participation constraint is equivalent to a non-negativity constraint on the subsidy: $s \geq 0$. The optimum found in Proposition 4 is nontrivial and feasible framework AMC (abbreviated NFFA) if and only if it satisfies non-negativity constraint with strict equality, i.e., if and only if $s^{PS} > 0$. If $s^{PS} = 0$, then the optimum from Proposition 4 is trivial—no different than not having an AMC. If $s^{PS} < 0$, then the optimal AMC for the donor would be infeasible. Using the expression for s^{PS} in Proposition 4, $s^{PS} > 0$ is equivalent to $Q^{PS} > D(c+\kappa/\beta) = Q^{NA}$.

Proposition 5. *There exists a nontrivial, feasible framework AMC (NFFA) if and only if* $Q^{PS} > Q^{NA}$*. To characterize the parameters satisfying this condition, there exists an NFFA for* β *sufficiently close to* 0 *and not for* β *sufficiently close to* 1 ceteris paribus. *No NFFA exists for* κ *sufficiently close to* 0 ceteris paribus.

Suppose donor objective function (19) *is concave in Q. Then there exists* $\hat{\beta} \in (0,1)$ *such that an NFFA exists for all* $\beta < \hat{\beta}$ *and not for all* $\beta \geq \hat{\beta}$ ceteris paribus. *Furthermore, there exists* $\hat{\kappa} > 0$ *such that an NFFA exists for for all* $\kappa \in (\hat{\kappa}, \bar{p} - c)$ *and not for for* $\kappa \leq \hat{\kappa}$ ceteris paribus.

constant at s^{PS}, donor surplus must be discontinuous from the left in the size of the fund. For F below the threshold value necessary to endow a perpetual subsidy, donor surplus is decreasing in F because the greater expense produces no incentive effects. A perpetual subsidy becomes feasible at the threshold value of F, causing donor surplus to jump because the perpetual subsidy provides the firm with discontinuously better investment incentives.

The text preceding the proposition sketched the proof of the first statement in the proposition. The formal proof in the appendix fills in some omitted details as well as establishing the remaining results characterizing the parameters under which an NFFA exists. The last statement in the proposition takes care to consider only values of κ satisfying Assumption 1. This statement leaves open the possibility that $\hat{\kappa} > \bar{p} - c$. If so, the set $(\hat{\kappa}, \bar{p} - c)$ is empty, implying that no NFFA exists for any admissible κ *ceteris paribus*.

The conditions guaranteeing no NFFA exists make intuitive sense. For β close to 1, the firm obtains most of the bargaining surplus so chooses a capacity close to the integrated solution Q^{IN} in the absence of an AMC. With little need to further enhance incentives with a positive subsidy, the donor would prefer to extract rent with a negative subsidy if this were feasible. Similarly, for κ close to 0, the scale of the hold-up problem is limited because of the small scale of ex ante investment. Again the donor would prefer to extract rent with a negative subsidy if this were feasible. If the donor's objective function is concave, this intuition generates threshold values of the parameters β and κ.

3.5. Supply Commitment

The previous subsection provided a theoretical explanation of how a framework AMC can be used to improve incentives. However, there may be some practical difficulties in achieving the theoretical optimum. To achieve the optimum with a framework AMC accuring interest requires the donor to be able to commit to a perpetuity. To approach the optimum with a framework AMC that does not accrue interest requires the donor to pledge an arbitrarily large fund. Budget and political constraints may present barriers to these extreme forms of commitment. Furthermore, while some AMC designs considered so far were able to improve incentives, none was able to achieve the integrated outcome.

In this section we analyze an alternative policy to the framework AMC. The Economics Expert Group, which worked on the design of the pilot AMC for a pneumococcus vaccine, recommended a move from the framework design to the design analyzed in this subsection, which it called a supply commitment. We will show that the supply commitment can achieve the same outcome as the optimal framework AMC without needing a perpetual commitment or an infinite fund. In addition to the contractual terms F and s, a supply commitment sets a target supply $\hat{q}$. Rather than being able to appropriate the entire fund F given enough time, now the firm is only eligible to appropriate at most a fraction equal to the fraction $q/\hat{q}$ that it supplies of the target.

With this design, returning to the assumption that interest accrues in the AMC escrow, the accounting identity determining the length of the AMC period becomes

$$\frac{\min(\hat{q}, q)}{\hat{q}} F = \frac{sq}{r}(1 - e^{-rT}). \tag{23}$$

The left-hand side reflects fund allowed to be used for the subsidy; the right-hand side is the present discounted value of payments out of the fund; the accounting identity is that payments out equal allowed funds. An implicit assumption behind the accounting identity is that the duration of the AMC period associated with the supply commitment is finite. We posit this because the presumed benefit of a supply commitment is that it allows the donor to avoid needing to commit to a perpetuity. The min operator in (23) reflects the fact that the firm cannot collect more than F in subsidy payments from the AMC by producing more than $\hat{q}$.

Rearranging (23),

$$F = \frac{s}{r}\max(\hat{q}, q)(1 - e^{-rT}). \tag{24}$$

One can show that if the donor sets $\hat{q}$ less than the equilibrium output, the supply commitment is identical to a framework AMC. The firm is already collecting the full F, so further increases in output do not have the beneficial incentive effect of expanding the fund. We are interested in improving on the efficiency of a framework AMC with a supply commitment. Thus we posit that the donor sets the target weakly above equilibrium output:

$$q \leq \hat{q}. \tag{25}$$

We will ultimately show that the optimal supply commitment does no worse than the optimal framework AMC, so assumption (25) is without loss of generality. Substituting (25) into (24) and rearranging to solve for the duration of the supply commitment,

$$T = \frac{1}{r}\ln\left(\frac{s\hat{q}}{s\hat{q} - rF}\right). \tag{26}$$

This equation makes clear why a supply commitment improves investment incentives relative to a framework AMC. With the framework design, an increase in supply increases the subsidy earned each instant during the AMC period, but this is offset by shortening the AMC's duration. With a supply commitment, as (26) shows, any increase in supply up to the target has no effect on the AMC's duration because the fund that the firm is able to draw upon grows proportionately with supply up to the target. This may lead to more powerful investment incentives with a supply commitment than a framework AMC.

Proceeding with the analysis, the derivation of the present discounted value of the firm's ex ante profit proceeds the same here as with the framework design, again giving equation (12). As before we invoke the principle of full capacity utilization to justify substituting capacity Q for output q. The only difference is what fills in for T. Substituting for T from (26) and rearranging yields firm profit

$$\frac{\beta}{r}\left[\alpha B(Q)-\left(c+\frac{\kappa}{\beta}\right)Q\right]+\frac{\beta QF}{\hat{q}}=\pi(Q,\theta)+\frac{\beta QF}{\hat{q}}. \tag{27}$$

Differentiating with respect to Q, taking $\hat{q}$ to be exogenous to the firm, and rearranging yields first-order condition

$$P(Q)-\left(c+\frac{\kappa}{\beta}\right)+\frac{rF}{\hat{q}}=0. \tag{28}$$

Folding the game back to the donor's design of the optimal supply commitment, the present discounted value of donor surplus equals its share $1-\beta$ of joint continuation surplus with a finite AMC period—the same as derived in (11)—less its initial contribution F to the escrow:

$$\frac{1-\beta}{r}[\alpha B(Q)-cQ+sQ(1-e^{-rT})]-F=\frac{1-\beta}{r}\left[\alpha B(Q)-cQ+\frac{rFQ}{\hat{q}}\right]-F, \tag{29}$$

where the right-hand side follows from substituting for T from (26) and rearranging. The optimal supply commitment for the donor is given by the values of Q, F, and $\hat{q}$ maximizing (29) subject to incentive-compatibility constraint (28). We also continue to impose the constraint (25) posited above on target output, which after invoking the principle of full capacity utilization becomes

$$Q\leq\hat{q}. \tag{30}$$

The maximization problem can be simplified by solving (28) for s and substituting the resulting expression for s in (18). This turns a maximization problem with three choice variables and two constraints into one with two choice variables and one constraint (30). After rearranging, the objective function from the transformed problem is

$$\frac{1}{r}\left\{(1-\beta)\alpha B(Q)+\beta QP(Q)-(c+\kappa)Q+(Q-\hat{q})\left[c+\frac{\kappa}{\beta}-P(Q)\right]\right\}. \tag{31}$$

It is obvious that (30) binds. Substituting (30) treated as an equality into (31) causes the last term (involving the factor $Q-\hat{q}$) to disappear. Differentiating the resulting objective function with respect to Q yields a first-order condition for the equilibrium capacity identical to (20) from the optimal framework AMC with a perpetual subsidy, leading immediately to the following proposition.

Proposition 6. *The donor's surplus from the optimal supply commitment is the same as from the optimal* *framework AMC with a perpetual subsidy. The required funds are the same across the two policies ($F^{SC} = F^{PS}$) as are equilibrium capacities and outputs ($Q^{SC} = q^{SC} = q^{PS} = Q^{PS}$).*

The proof in the appendix fills in the remaining detail about equality between fund sizes.

We have so far put aside issues related to the existence of a feasible, nontrivial supply commitment. A feasible supply commitment satisfies the firm's participation constraint, which here can be shown to be equivalent to a non-negative subsidy rate $s \geq 0$, or equivalently a non-negative fund $F \geq 0$. Given that a supply commitment attains the outcome under a framework AMC with a perpetual subsidy, not surprisingly the conditions under which a feasible, nontrivial supply commitment exists are identical to those provided in Proposition 5 for a feasible, nontrivial framework AMC.

Proposition 6 is silent about the subsidy rate s^{SC} in the optimal supply commitment. That is because this rate is not pinned down in an optimum. A low rate paid over a long AMC period or a high rate paid over a short period can both be optimal. Specifically, any pairs (s, T) satisfying $s > 0$, $T > 0$, and

$$\frac{s}{r}(1 - e^{-rT}) = \frac{F^{SC}}{Q^{SC}} \tag{32}$$

can be part of an optimal supply commitment. All that matters for incentives is the present discounted value of subsidy payments, not the particular path by which this value is reached.

The limiting case in which an arbitrarily large s is paid over an arbitrarily short AMC period is of particular interest. The limit is equivalent to an alternative policy that has the donor commit to paying the firm a linear price F^{SC}/Q^{SC} per unit of installed capacity. Such a contract is feasible if capacity is contractible. If not, the donor can approximate the same outcome by paying a huge per-unit subsidy over a very short interval, coming arbitrarily close to making a lump-sum payment in the first instant of ex post time.

This way of viewing a supply commitment helps explain why it cannot attain the efficiency of the integrated outcome. A supply commitment in essence has the donor pay a linear price for capacity. Just as with a perpetual subsidy, the donor behaves as a hybrid of a social planner and a linear-pricing monopsonist, depending on its bargaining power. Indeed, the same Lerner index formula from (22) also applies under a supply commitment. To avoid this monopsony distortion and thereby improve efficiency would require some sort of nonlinear contract. We will analyze one sort of nonlinear contract, a forcing contract, next.

3.6. Forcing Contract

A forcing contract pays the subsidy to the firm if and only if it produces a specified level of output. This subsection shows that a forcing contract can be used to improve upon all the policies previously analyzed.

Indeed, a suitably designed forcing contract is optimal among the broad class of feasible commitments (feasible in that they satisfy individual rationality for the firm), allowing the donor to attain the integrated outcome.

Consider a forcing contract that pays a subsidy s per unit from a fund of size F if and only if the firm's output equals capacity in the integrated benchmark, Q^{IN}. Supposing for now that the AMC subsidy is temporary, and continuing with the baseline specification that interest accrues in the fund, following (10) the length of the AMC period is given by

$$F = \frac{sQ^{IN}}{r}\left(1-e^{-rT}\right). \tag{33}$$

If the firm rejects the contract, the continuation game resembles that without an AMC analyzed in Section 3.2. Proposition 2 showed the firm's ex ante profit is $\pi(Q^{NA},\theta)$ in that case. The firm can earn no more than $\pi(Q^{NA},\theta)$ if it accepts the contract but produces a different quantity than Q^{IN} during the AMC period. Thus it will be an equilibrium for the firm to accept the contract and produce Q^{IN} if we can show the firm's ex ante profit from the forcing contract is at least $\pi(Q^{NA},\theta)$.

Suppose the firm accepts the forcing contract, builds capacity Q^{IN}, and produces Q^{IN} during the AMC period. We will solve for the continuation equilibrium using backward induction, starting from the Nash bargain at $t = 0$. The donor's and firm's threat points are 0. Gains from trade equal the present discounted value of joint continuation surplus from the AMC and tail periods:

$$\int_0^T [\alpha B(Q^{IN}) - cQ^{IN} + sQ^{IN}]e^{-rt}dt + \int_T^\infty [\alpha B(Q^{IN}) - cQ^{IN}]e^{-rt}dt. \tag{34}$$

The second term implicitly assumes the firm produces Q^{IN} in the tail as well as the AMC period. This follows from the principle of full capacity utilization. The principle holds because the firm's installed capacity satisfies $Q^{IN} = D(c+\kappa)$ by Proposition 1, and $D(c+\kappa) < D(c)$. Integrating, substituting for T from (33), and rearranging yields the following expression for gains from trade:

$$\frac{1}{r}[\alpha B(Q^{IN}) - cQ^{IN}] + F. \tag{35}$$

Folding the game back, the present discounted value of the firm's ex ante profits equal its share β of (35) less capacity costs. Upon rearranging, this present discounted value is

$$\frac{\beta}{r}\left[\alpha B(Q^{IN}) - \left(c + \frac{\kappa}{\beta}\right) Q^{IN}\right] + \beta F = \pi(Q^{IN},\theta) + \beta F. \tag{36}$$

For the forcing contract to work, equation (36) must exceed the firm's ex ante profit $\pi(Q^{NA},\theta)$ in the absence of an AMC, which upon rearranging requires $F \geq [\pi(Q^{NA},\theta) - \pi(Q^{IN},\theta)]/\beta$. The optimal fund size F^{FC} is just large enough to satisfy this condition with equality. We have the following proposition.

Proposition 7. *The donor's optimal forcing contract involves a fund of size*

$$F^{FC} = \frac{1}{\beta}\left[\pi(Q^{NA},\theta) - \pi(Q^{IN},\theta)\right], \tag{37}$$

any subsidy satisfying $s \geq rF^{FC}/Q^{IN}$, *and a requirement to produce* Q^{IN} *during the AMC period. Equilibrium capacity and output under this contract equal* Q^{IN}. *This forcing contract is the optimal policy among any that satisfy individual rationality for the firm.*

Proof. By construction, fund size F^{FC} is sufficient to induce the firm to produce Q^{IN} during the AMC period under a forcing contract rather than any other output. The firm must build capacity at least Q^{IN} to produce this much. It is obvious the firm would not build more capacity than Q^{IN}. Thus the firm's equilibrium capacity and output are Q^{IN}. This is the same capacity and output as in the integrated benchmark. Therefore the present discounted value of parties' joint surpluses must be the same as in the integrated benchmark, which is the highest possible. By construction of F^{FC}, the firm's ex ante profit is $\pi(Q^{NA},\theta)$, the least it can earn from any individually rational policy. Thus the forcing contract is optimal for the donor among all policies that are individually rational for the firm. *Q.E.D.*

The optimal subsidy is not uniquely pinned down. An increase in s reduces the AMC length T, leaving the present discounted value of the subsidy stream the same at F^{FC}. The only requirement is that s is sufficiently high that the firm can extract the whole value of the fund rather than having it continue to grow without bound. If $s > rF^{FC}/Q^{IN}$, the subsidy associated with the forcing contract is temporary; if $s = rF^{FC}/Q^{IN}$, the subsidy is perpetual; but in either case the present discounted value of subsidy payments is F^{FC}.

If capacity is contractible, the donor can obtain the same outcome it did under the forcing contract just analyzed, which is a quantity-forcing contract, using a capacity-forcing contract. The donor could simply pay the firm F^{FC} to install capacity Q^{IN}. There is no more direct way of providing incentives to invest in capacity than "buying" that capacity outright. Of course capacity may be difficult to contract on in practice. The firm could sell the factory for another use after collecting the payment or could build a shabby one and abandon it. If capacity is not contractible, the quantity-forcing contract specified in Proposition 7 would still work and would fall inside the allowable set of contracts on technologically close products, which allow the commitment to be conditioned on output.

3.7. Refinements

A variety of additional refinements to the design of a basic AMC were considered in the pilot program. To conserve on space, we have omitted the analysis of these refinements from the text, instead devoting a separate appendix to it. Appendix B examines how the performance of the basic AMC studied so far changes when price caps, purchase guarantees, and country copayments are added. Appendix B also examines the effect of changing the party to which interest accrues from the AMC fund and of adding a separate player (independent of the donor) who acts as a procurement agent in bargaining, much as Gavi functioned in the pilot program.

4. Cost Uncertainty

We have so far studied a model of complete information. Among other things, this meant that the donor was able to design the AMC with symmetric information about the firm's cost parameters κ and c. In practice, the firm may have private information regarding these parameters, leading to contracting frictions. Our experience on the Economics Expert Group for the pneumococcus pilot bears this concern out: even after considerable expenditure of resources interviewing industry participants, consultants, and our own experts, the Group found it difficult to narrow the range of plausible estimates of c. With the benefit of time, information, and perhaps some invention, critics have contended that actual costs turned out to be much lower than the operating estimates, charging that the program gave away too much rent to firms (see Scudellari 2011 for a review) .

This section addresses the issue of asymmetric information on AMC design in a simple extension in which production costs can take on two values: $\underline{c}$ with probability $\nu \in (0,1)$ and $\bar{c}$ with probability $1-\nu$, with $\underline{c} < \bar{c}$. Ex ante, the firm knows the realization of c; the donor only knows the distribution. Assume that all players learn c ex post, so that in particular Nash bargaining occurs under symmetric information.

Although the pilot program specified a single option, to build understanding here we will examine the design of the optimal screening mechanism, a direct-revelation mechanism offering different forcing contracts depending on the firm's cost announcement. The forcing contract specifies output $\underline{Q}$ and fund size $\underline{F}$ if the firm announces cost $\underline{c}$ and output $\bar{Q}$ and fund size $\bar{F}$ if the firm announces $\bar{c}$. This is a fairly standard screening problem, so after introducing some notation, we will jump to the statement of a proposition characterizing the optimum, relegating the details of the analysis to the proof in the appendix. For the required notation, let $Q^{NA}(c)$ be the equilibrium output in the absence of an AMC, i.e., $Q^{NA}(c) = D(c+\kappa/\beta)$, and $Q^{IN}(c)$ be the integrated output, i.e., $Q^{IN}(c) = D(c+\kappa)$. Let $F^{FC}(c)$ be the solution to equation (37) for a particular value of c. All functions just defined depend on more parameters than c, but we use the argument to emphasize their dependence on the firm's type. Slightly abusing notation further, we will write the welfare function from (3) as $w(Q,c)$ and the profit function from (8) as $\pi(Q,c)$, the second argument stressing the dependence of both on the parameter c of interest rather than the whole vector of model parameters θ.

Proposition 8. *There exists $\hat{\beta} \in (0,1)$ such that the following direct revelation mechanism is the optimal AMC in the presence of cost uncertainty.*

- *For $\beta \geq \hat{\beta}$, each type receives the same forcing contract as under symmetric information: type $\underline{c}$ receives $\underline{Q}^* = Q^{IN}(\underline{c})$ and $\underline{F}^* = F^{FC}(\underline{c})$, and type $\bar{c}$ receives $\bar{Q}^* = Q^{IN}(\bar{c})$ and $\bar{F}^* = F^{FC}(\bar{c})$. Expected donor surplus is the same as under symmetric information.*
- *For $\beta < \hat{\beta}$, type $\underline{c}$'s quantity remains $\underline{Q}^* = Q^{IN}(\underline{c})$, but type $\bar{c}$'s is distorted downward to*

$$\bar{Q}^* = D\left(\bar{c}+\kappa+\frac{\nu}{1-\nu}(\bar{c}-\underline{c})\right). \tag{38}$$

  Expected donor surplus falls short of that under symmetric information by

$$\nu\left\{\frac{\beta}{r}(\bar{c}-\underline{c})\bar{Q}^* - \left[\pi(Q^{NA}(\underline{c}),\underline{c})-\pi(Q^{NA}(\bar{c}),\bar{c})\right]\right\}+(1-\nu)\left\{w(Q^{IN}(\bar{c}),\bar{c})-w(\bar{Q}^*,\bar{c})\right\}. \tag{39}$$

The proof in the appendix provides explicit expressions for $\hat{\beta}$ as well as the fund sizes for the case in which $\beta < \hat{\beta}$. The one wrinkle in the analysis is that the firm's outside option figuring into its participation constraint is type-specific. The firm's outside option is to participate in the no-AMC equilibrium characterized in Section 3.2. The efficient type's profit is higher in this equilibrium than the inefficient type's. For β sufficiently close to 0, the wrinkle indeed matters. The binding constraint for the efficient type is its participation constraint, not its truth-telling constraint. In this case the donor can obtain the integrated outcome in all states by designing the menu options for the two types independently. For β sufficiently close to 1, the wrinkle has no bearing on the problem. We obtain the usual results for screening mechanisms that there is "no distortion at the top" for the efficient type but that the inefficient type's forcing contract is distorted downward to extract information rent from the efficient type. As usual, the binding constraint for the efficient type is its truth-telling constraint.

The expression for expected donor surplus lost to asymmetric information on (39) can be understood as follows. The first term is the probability ν of a low-cost type times the information rent the donor loses to this type in curly braces. This is slightly more complicated than the usual information-rent term (which would be simply $(\beta/r)(\bar{c}-\underline{c})\bar{Q}^*$) because the rent is measured in excess of the type's participation constraint, which as noted in the previous paragraph is type specific. The efficient type's better outside option shows up as the reduction in its information rent given by the term in square brackets, $\pi(Q^{NA}(\underline{c}),\underline{c})-\pi(Q^{NA}(\bar{c}),\bar{c})$. The second term is the probability $1-\nu$ of a high-cost type times the social loss from distorting its equilibrium capacity downward in curly braces. Because this type's participation constraint binds under both symmetric and asymmetric information, the donor bears the whole distortion.

The analysis points out a political risk associated with the optimal mechanism design. For the case in which β is high, by design, the AMC provides the efficient type with an information rent beyond the already positive profit it would earn in the absence of an AMC. A critic of the AMC program with knowledge of true cost ex post could complain that this rent was an unnecessary giveaway to the firm. The only way to avoid this rent is to distort the inefficient type's quantity down further. This conservative approach to avoiding political risk could inflict substantial social distortions in high-cost states of the world.

5. Competing Suppliers

To this point we have analyzed the case of a monopoly firm. In this section, we extend the model to allow for competing firms. The results will mostly confirm the robustness of of previous results. For example, the Lerner index for the optimal supply commitment with N firms has the same form as (22) for a monopoly firm. Allowing for competition does uncover some new insights. For example, we show that AMC designs that were useless with a monopoly supplier, providing no investment incentives, function better under competition, strictly enhancing incentives for any positive AMC fund F. If competition is sufficiently intense, we show show that a framework AMC can be more efficient than a supply commitment.

To introduce competition in the model yet preserve the feature that ex post prices are set through bargaining, we assume in this section that ex post bargaining is characterized by the Shapley value. When $N = 1$, Shapley bargaining reduces to Nash bargaining with $\beta = 1/2$. Thus Shapley bargaining is a generalization of Nash bargaining when parties have equal bargaining power. The results of this section can be compared to previous sections' after substituting $\beta = 1/2$ into the previous results.

For pedagogical purposes, we provide details of the analysis in the simple case of duopoly firms in Section 5.1. Section 5.2 provides results for general number of firms $N \geq 2$, but most of the analysis there is shifted to the appendix proofs.

5.1. Duopoly Firms

Suppose there are $N = 2$ firms, indexed by $i = 1,2$. As a benchmark, we first solve for equilibrium with duopoly firms in the absence of an AMC. We use backward induction, starting with ex post Shapley bargaining.

The Shapley value for firm i is a weighted average of its marginal contribution to surplus in coalitions consisting of i and all preceding players in every permutation of bargaining parties. In one of the six permutations, i's marginal contribution to the coalition is $\alpha B(Q_i) + cQ_i$ each instant because firm j is not involved. In two of the six, j is involved and i's marginal contribution is $\alpha B(Q_i + Q_j) - \alpha B(Q_j) - cQ_i$. In the remaining three, i makes no marginal contribution. Taking the weighted average of these marginal contributions, and then taking the present discounted value over the entire ex post period gives i's Shapley value. Subtracting i's capacity cost and rearranging gives the following expression for i's ex ante profit:

$$\frac{1}{r}\left\{\frac{1}{6}[\alpha B(Q_i) - cQ_i] + \frac{1}{3}[\alpha B(Q_i + Q_j) - \alpha B(Q_j) - cQ_i] - \kappa Q_i\right\}. \tag{40}$$

Note that capacity has been substituted for output in all surplus expressions in the previous paragraph, based on the principle of full capacity utilization. This principle, a forgone conclusion with a monopoly firm, is no longer so with competing firms, which can have good strategic reasons for holding excess capacity. Firm i's ex post continuation surplus depends not only on its marginal contribution to the grand coalition but also to the coalition in which it is the only operating firm. What turns out to be excess capacity in the grand coalition emerging in equilibrium can serve to increase a firm's marginal contribution in a smaller out-of-equilibrium coalition. The case in which firms hold excess capacity in the absence of an AMC is trivial for policy analysis because there would be no reason to incentivize more capacity with an AMC. To streamline the analysis in this subsection, we will posit that we are not in this trivial case, providing a sufficient condition for general N in the more technical subsection to follow.

Taking the first-order condition of (40) with respect to Q_i, and imposing symmetry (i.e., $Q_i = Q_j$) yields the following equilibrium condition on total capacity $Q = Q_i + Q_j$ after rearranging:

$$\frac{2}{3}P(Q) + \frac{1}{3}P(Q/2) = c + 2\kappa. \tag{41}$$

Let Q_2^{NA} denote the value of Q satisfying (41), thus equilibrium duopoly capacity in the absence of an AMC. It is straightforward to compare Q_2^{NA} to its monopoly analogue Q_1^{NA}, the solution to equation (9). Setting $\beta = 1/2$ in (9) gives $P(Q_1^{NA}) = c + 2\kappa = (2/3)P(Q_2^{NA}) + (1/3)P(Q_2^{NA}/2)$, implying $P(Q_2^{NA}) = P(Q_1^{NA}) - (1/3)[P(Q_2^{NA}/2) - P(Q_2^{NA})]$, implying $Q_2^{NA} \geq Q_1^{NA}$, with strict inequality as long as the relevant prices do not exceed $\bar{p}$.

These arguments have thus established $Q_2^{NA} > Q_1^{NA}$, i.e., industry capacity in the absence of an AMC is strictly higher with a duopoly than monopoly. Intuitively, competition between the firms ameliorates the hold-up problem. Firms expand capacity to improve their bargaining position in out-of-equilibrium coalitions without the other firm.

With the analysis of the benchmark without an AMC in hand, we proceed to the analysis of a framework AMC offered to competing firms. We will take a particular specification that proved useless with a monopoly firm, specifically one with a temporary subsidy with interest accruing to the escrow. Our main interest will be in showing that this AMC increases investment over and above the level without an AMC. We will thus maintain the assumption that equilibrium capacity with a framework AMC is strictly below $D(c)$. If it is weakly above, we are done because we are positing throughout the subsection that capacity is less than $D(c)$ in the absence of an AMC.

Given we are analyzing an AMC with a temporary subsidy, the following accounting identity determines the equilibrium duration of the AMC subsidy:

$$F = \frac{s(Q_i + Q_j)}{r}\left(1 - e^{-rT_2}\right). \tag{42}$$

We have substituted capacities for output because Shapley bargaining is efficient, so results in both firms' capacities being fully utilized in equilibrium. The subscript on AMC duration T_2 indicates that two firms (i and a rival) participate in the equilibrium coalition. The Shapley value requires computing surplus for the out-of-equilibrium coalition including just i and the donor. The accounting identity determining AMC duration T_1 for that coalition is

$$F = \frac{sQ_i}{r}\left(1 - e^{-rT_1}\right). \tag{43}$$

We can derive an expression for firm i's ex ante profit by computing the Shapley value surplus and subtracting off the capacity cost, yielding

$$\frac{1}{r}\left\{\frac{1}{6}[\alpha B(Q_i) - cQ_i + sQ_i(1 - e^{-rT_1})] + \frac{1}{3}[\alpha B(Q_i + Q_j) - \alpha B(Q_j) - cQ_i + sQ_i(1 - e^{-rT_2})] - \kappa Q_i\right\} \tag{44}$$

$$= \frac{1}{r}\left\{\frac{1}{6}[\alpha B(Q_i) - cQ_i + rF] + \frac{1}{3}\left[\alpha B(Q_i + Q_j) - \alpha B(Q_j) - cQ_i + rF\left(\frac{Q_i}{Q_i + Q_j}\right)\right] - \kappa Q_i\right\} \tag{45}$$

where (45) follows from substituting for T_1 and T_2 from (42) and (43). The term $FQ_i/(Q_i + Q_j)$ encapsulates the beneficial incentive effects of competition in this setting. Whereas the AMC program results in a lump-sum increase in firm profit in the monopoly case, this is no longer true under duopoly. An increase in capacity allows firm i to "steal" a greater share of the AMC for itself from the other firm.

Taking the first-order condition with respect to Q_i, and imposing symmetry (i.e., $Q_i = Q_j$) yields the following equilibrium condition on total capacity $Q = Q_i + Q_j$ after rearranging:

$$\frac{2}{3}P(Q) + \frac{1}{3}P(Q/2) = c + 2\kappa - \frac{rF}{3Q}. \tag{46}$$

Comparing (46) to (41), we see that they differ by the term $rF/3Q$. Standard comparative-statics techniques can be used to show $Q_2^{TS} > Q_2^{NA}$, i.e., framework AMC with temporary subsidy and interest accruing to the escrow strictly increases capacity with duopoly firms.

Intuitively, the duopoly analysis uncovered a benefit of a framework AMC hidden by the previous monopoly analysis. A framework AMC amplifies the incentive effect of competition. With a monopoly firm, there is no competition, so no competitive effect to amplify. With two firms, there is a competitive effect to amplify; and the AMC turns out to be beneficial. Another change from the monopoly case is that the subsidy rate s is irrelevant for incentives and indeed drops out of the AMC-design problem entirely with duopoly firms. Firms end up claiming a proportional share of the fund F, so they care about the size of F rather than the schedule over which it is paid out.

Extending the analysis of further issues to the case of competing firms—optimal framework AMC design, comparison of framework AMCs to supply commitments, etc.—turns out to be more conveniently done in the general case of $N \geq 2$ rather than the duopoly case, so we turn to the general analysis next.

5.2. Oligopoly Firms

A key definition that we will use in the analysis of the general case of $N \geq 2$ competing firms is the virtual inverse demand arising from Shapley bargaining,

$$\mathbf{P}(Q,N) = \sum_{n=1}^{N} [\tilde{w}(n,N)P(nQ/N)], \tag{47}$$

a weighted average of inverse demands for the capacity of an n-firm coalition, with the weights $\tilde{w}(n,N)$ given by conditional probabilities of coalition formation. More precisely, let $w(n,N)$ denote the unconditional probability that a coalition including i—randomly formed by permuting a grand coalition of N firms and the donor—happens to have exactly $n-1$ other firms (so n firms total) as well as the donor. Note that the donor must be included in the coalition for it to generate positive surplus. The proof of the next lemma shows

$$w(n,N) = \frac{n}{N(N+1)}. \tag{48}$$

Let $\tilde{w}(n,N)$ be the conditional probability derived from $w(n,N)$, i.e., conditional on the coalition's including the donor and thus being able to generate a positive surplus. Then

$$\tilde{w}(n,N) = \frac{w(n,N)}{\sum_{k=1}^{N} w(k,N)} = \frac{2n}{N(N+1)}, \tag{49}$$

as

$$\sum_{k=1}^{N} w(k,N) = \frac{1}{N(N+1)} \sum_{k=1}^{N} k = \frac{1}{2}. \tag{50}$$

By construction, $\sum_{n=1}^{N} \tilde{w}(n,N) = 1$. The reader can verify that $\mathbf{P}(Q,1) = P(Q)$ and that $\mathbf{P}(Q,2) = (2/3)P(Q) + (1/3)P(Q/2)$, the expression appearing on the left-hand side of (41) and (46).

We now have the requisite notation to state a sufficient condition under which the principle of full capacity utilization holds in the absence of an AMC. We posited this principle in the previous subsection's analysis of duopoly. The next lemma provides a sufficient condition for general $N \geq 1$.

Lemma 2. *For all $N \geq 1$, if*

$$\mathbf{P}(D(c),N) < c + 2\kappa, \tag{51}$$

then the principle of full capacity utilization holds in the absence of an AMC. Condition (51) *is non-vacuous, in particular satisfied* ceteris paribus *for sufficiently large $\kappa > 0$.*

The proof of the lemma in the appendix proceeds by finding an expression for equilibrium capacity in the absence of an AMC Q_N^{NA} under the assumption that $Q_N^{NA} < D(c)$ and then using the resulting expression for Q_N^{NA} to derive a sufficient condition for $Q_N^{NA} < D(c)$. The proof thus ends up providing a fairly comprehensive analysis of equilibrium in the absence of an AMC. As part of this analysis, the proof derives the following generalization of equilibrium condition (41) to $N \geq 1$ firms:

$$\mathbf{P}(Q,N) = c + 2\kappa. \tag{52}$$

Condition (51) can be easy to verify in examples. Suppose $N = 1$. Then (51) reduces to $\kappa > 0$, which is true in our model independent of the demand specification. This verifies the principle of full capacity utilization derived in Section 3.3. To take another example, when $N = 2$ and $P(Q) = 1 - Q$, (51) reduces to the simple condition $\kappa > (1-c)/12$.

The proof of the next proposition combines (52) with the facts that $\partial \mathbf{P}(Q,N)/\partial Q < 0$ and $\mathbf{P}(Q,N) > \mathbf{P}(Q,1)$ to show that competition leads to more capacity than monopoly in the absence of an AMC.

Proposition 9. *For all $N > 1$, $Q_N^{NA} > Q_1^{NA}$.*

The proof is provided in the appendix.

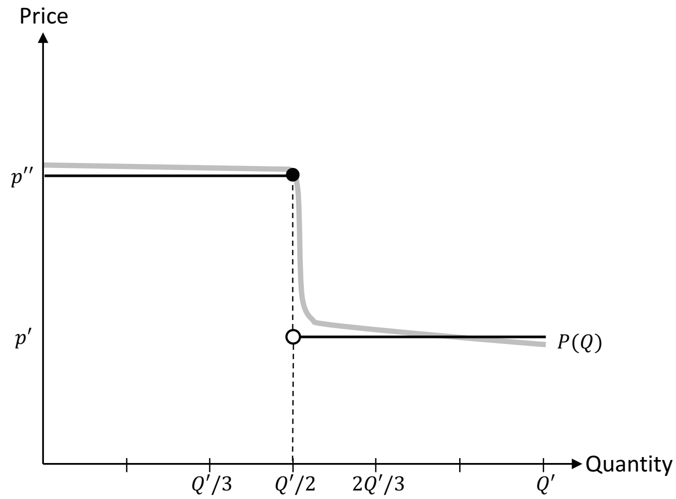

Figure 3: Counterexample to competition increasing capacity

The proposition says that starting from monopoly, an increase in N increases equilibrium capacity. More generally, one might conjecture that an increase in N starting from any number of firms would increase capacity. This conjecture turns out to be false, as shown by the counterexample in Figure 3. Consider the inverse demand given by the discontinuous, black curve, equaling p'' for quantities less than or equal to $Q'/2$ jumping down to p' for quantities above this threshold, where $p'' > p'$. We have $\mathbf{P}(Q',2) = (2/3)P(Q') + (1/3)P(Q'/2) = (2/3)p' + (1/3)p''$, while $\mathbf{P}(Q',3) = (1/2)P(Q') + (1/3)P(Q'/2) + (1/6)P(Q'/3) = (5/6)p' + (1/6)p''$, implying $\mathbf{P}(Q',2) > \mathbf{P}(Q',3)$. Given this inequality, costs c and κ can be constructed such that an increase from $N = 2$ to $N = 3$ firms reduces equilibrium capacity. Of course the black inverse demand does not satisfy the properties required of inverse demand such as continuous differentiability, but it is clear that an inverse demand such as the grey one in the figure which is close to the black one but does satisfy the reuired properties can be constructed. The next proposition, proved in the appendix, states that linearity of inverse demand is sufficient for an increase in competition to increase capacity starting from any number of firms.

Proposition 10. *If inverse demand $P(Q)$ is linear for all $Q < D(c)$, then $Q^{NA}_{N''} > Q^{NA}_{N'}$ for all natural numbers $N'' > N'$.*[4]

With an understanding of how competition affects equilibrium in the absence of an AMC, we turn to general results on the effect of competition on the performance of different policies. The previous section showed that a framework AMC lead to a strict increase in capacity with duopoly firms. The next proposition generalizes the result to any number $N \geq 2$ of competing firms.

[4]The proposition specifies that inverse demand is linear "in the relevant region" because inverse demand must eventually asymptote to the horizontal axis, so cannot be globally linear. A sufficient condition is that inverse demand be linear for all $Q < D(c)$.

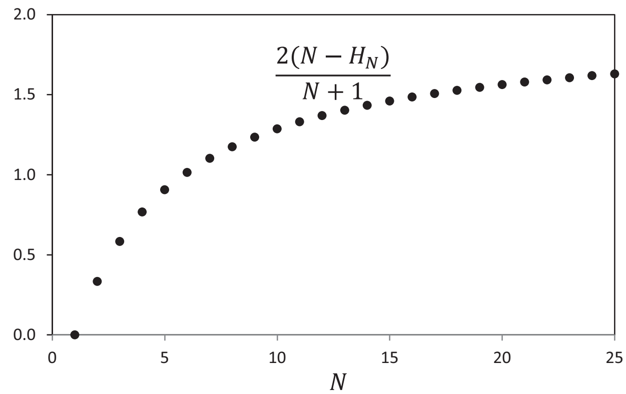

Figure 4: Scale factor used in analysis of effect of competition on framework AMC

Proposition 11. *Suppose there are $N \geq 2$ symmetric firms and that industry capacity in the absence of an AMC is less than $D(c)$. Introducing a framework AMC with temporary subsidy and interest accruing to the escrow strictly increases capacity.*

The key step in the proof, provided in the appendix, is deriving the following generalization of the equilibrium condition (46) to $N \geq 1$ firms:

$$\mathbf{P}(Q,N) = c + 2\kappa - \frac{rF}{Q}\left[\frac{2(N-H_N)}{N+1}\right], \tag{53}$$

where H_N denotes the Nth harmonic number: $H_N = \sum_{n=1}^{N} 1/n$. Figure 4 provides a graph of the factor in square brackets. Inspection of the figure suggests the factor in square brackets is positive and increasing in N for natural numbers $N \geq 2$ and approaches 2 in the limit as $N \to \infty$, properties which are proved formally in the appendix. The reader can verify that equation (53) reduces to (46) when $N = 2$, thus nesting the duopoly case. Equation (53) also nests the monopoly case. To see this, note that $N = H_N = 1$ when $N = 1$, in which case (53) reduces to $P(Q) = c + 2\kappa$, implying $Q = D(c + \kappa/(1/2))$, precisely the formula for equilibrium capacity found in Proposition 3 in the monopoly case when $\beta = 1/2$.

With equation (53) in hand, the proof of Proposition 11 follows easily. Equilibrium capacity in the absence of an AMC is the solution to (52), which is equivalently the solution to (53) setting $F = 0$. Equilibrium capacity in the presence of a framework AMC is the solution to (53) setting $F > 0$. Let $Q^*(F)$ be the solution to (53) for a given F. Then applying the implicit function rule to (53) yields

$$\frac{dQ^*(F)}{dF} = \frac{r}{\dfrac{rF}{Q^*(F)} - Q^*(F)\dfrac{\partial \mathbf{P}}{\partial Q}\Big/\left[\dfrac{2(N-H_N)}{N+1}\right]},$$

which is positive for $N \geq 2$ because the factor in square brackets is positive and $\partial\mathbf{P}/\partial Q < 0$ as argued earlier.

The previous proposition was about the incentive effects of a framework AMC, regardless of whether or not its design was optimal. We next turn to an analysis of optimal policies, characterizing equilibrium under an optimal framework AMC and comparing its performance to the optimal supply commitment. To facilitate comparisons, we will characterize equilibrium using the following generalization of the Lerner index:

$$\mathbf{L}_N^{TS} = \frac{\mathbf{P}(Q_N^{TS},N)-(c+\kappa)}{\mathbf{P}(Q_N^{TS},N)}. \tag{54}$$

Instead of the usual markup of inverse demand over cost, (54) involves the markup of Shapley virtual inverse demand. We have maintained the convention that bold letters distinguish the Shapley virtual analogue of a variable. Continuing with this convention, let $\boldsymbol{\eta}_N^{TS}$ denote the elasticity of Shapley virtual demand, which put in terms of the Shapley virtual inverse demand, satisfies

$$\frac{1}{\boldsymbol{\eta}_N^{TS}} = \frac{\partial \mathbf{P}(Q_N^{TS},N)}{\partial Q} \cdot \frac{Q_N^{TS}}{\mathbf{P}(Q_N^{TS},N)}. \tag{55}$$

To analyze equilibrium under a framework AMC, we can proceed as usual, deriving an expression for donor surplus, maximizing this expression subject to incentive-compatibility constraint (53), which after substitution can be transformed into an equivalent unconstrained capacity-choice problem. The resulting first-order condition for the donor can be rearranged into the following Lerner-index formula:

$$\mathbf{L}_N^{TS} = \left(\frac{1}{N+1-H_N}\right)\frac{1}{|\boldsymbol{\eta}_N^{TS}|} + \left(\frac{H_N+1-N}{N+1-H_N}\right)\kappa. \tag{56}$$

Applying similar steps to the analysis of the optimal supply commitment yields the following Lerner-index formula:

$$\mathbf{L}_N^{SC} = \left(\frac{1}{2}\right)\frac{1}{|\boldsymbol{\eta}_N^{SC}|}, \tag{57}$$

where $\mathbf{L}_N^{SC}$ is defined by analogy to (54) and $\boldsymbol{\eta}_N^{SC}$ by analogy to (55).

The similar structure of equations (56) and (57) facilitates comparison of the equilibrium Lerner indexes across the two policies. They have up to two terms: one comprised of an inverse elasticity and its associated coefficient and one comprised of κ and its associated coefficient. In equation (57), κ does not appear on the right-hand side, so effectively the coefficient associated with this term is 0. Thus to compare Lerner indexes, we need to compare coefficients on the inverse elasticities, i.e., $1/(N+1-H_N)$ versus $1/2$, and coefficients on κ, i.e., $(H_N+1-N)/(N+1-H_N)$ versus 0. After rearranging, one can see that these comparisons amount to the same thing, namely, both coefficients in (56) are greater than their analogues in (57) if and only if

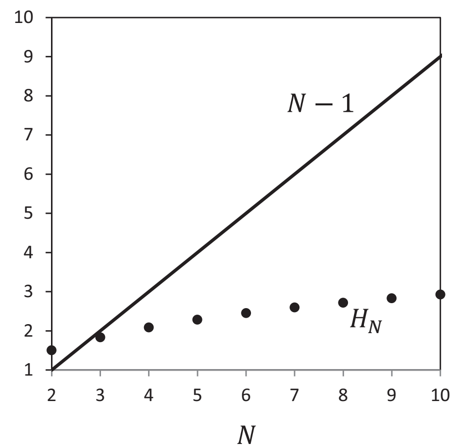

Figure 5: Toward a comparison of Lerner indexes across policies

$H_N > N-1$. Figure 5 graphs H_N (dots) against $N-1$ (solid line). We conclude from the figure that $H_N > N-1$ and thus $\mathbf{L}_N^{TS} > \mathbf{L}_N^{SC}$ when $N = 2$ and the reverse when $N \geq 3$. The implications for equilibrium capacities are the opposite of the Lerner indexes: a supply commitment generates more equilibrium capacity for $N = 2$ and a framework AMC more capacity for $N \geq 3$. The conclusions for capacity carry over to conclusions about the relative efficiency of the two policies, as the next proposition states.

Proposition 12. *Suppose there are $N \geq 2$ symmetric firms and that condition* (51) *is satisfied. The equilibrium capacity associated with the policy that is optimal for the donor satisfies* (56) *for a framework AMC and* (57) *for a supply commitment. The donor strictly prefers a supply commitment if $N = 2$ and a framework AMC if $N \geq 3$.*

Before discussing the broader intuition behind the proposition, several details require discussion. Concerning the proof, the proof in the appendix provides a formal derivation of Lerner-index conditions (56) and (57) and a formal comparison of the efficiency of the two policies. Concerning condition (51) posited in the statement of the proposition, this was the condition guaranteeing that the principle of full capacity utilization applies in the absence of an AMC. This same condition is also relevant to the design of optimal policies. Noting that the condition (52) for equilibrium capacity in the absence of an AMC is the limit of (53) as $F \to 0$, by continuity the principle must also hold for a framework AMC with a sufficiently small F. A larger F violating the principle must be suboptimal because the donor would not waste money incentivizing unused capacity. The same is true for the optimal supply commitment.

The general analysis of competition has uncovered a surprising and powerful result. The monopoly analysis suggested that supply commitments outperform framework AMCs (at least the design with a temporary subsidy). This continues to be the case with few competing firms. But with enough competition—the dividing line is three or more firms—surprisingly, the result reverses and framework AMCs start to outperform supply commitments. The intuition is that the two policies have their own pros and cons, the tradeoff between which changes with more competition. The framework AMC has the drawback that an increase in output shortens the AMC period, an effect that does not arise with a supply commitment because the size of the accessible fund scales with output, keeping duration constant. This is a factor in favor of a supply commitment, emphasized in the previous discussion in Section 3.5. On the other hand, a supply commitment has its own drawback having to do with the nature of Shapley bargaining. Out-of-equilibrium coalitions with fewer than all N firms lack the capacity to use up the whole fund. Unused funds end up being wasted, dulling investment incentives. Under Shapley bargaining, these out-of-equilibrium coalitional outcomes to factor into equilibrium surplus allocations and thus into equilibrium investment incentives. It is as if the industry were subsidized using a weighted average of of smaller funds rather than the whole F. The result is powerful because it only depends on the number of firms, not the demand specification or cost parameters.

Although no funds end up being wasted in equilibrium, the out-of-equilibrium waste of funds could be viewed as a defect of a supply commitment that one could consider remedying with suitable design modifications. One possibility would be to have the unused funds flow back to the donor. This would completely undermine the incentive effects of a supply commitment because the subsidy would just be a transfer, not a gain to the bargaining parties. For the supply commitment to work, unused funds either have to stay in the escrow or flow to a third party. Which is irrelevant in equilibrium. A third party would pay nothing for a claim on unused funds because it never happens that funds are unused in equilibrium. This is unlike the case of interest accrual studied in Section B.3 in Appendix B in which a third party would pay a positive up-front fee to receive the flow of interest payments from the AMC fund. Thus the version of the supply commitment we have studied cannot be improved by specifying a different disposition for unused funds.

In Section 3.5, with only the monopoly results for reference, we argued that the advantage of supply commitments over framework AMCs supported the recommendation of the Economics Expert Group to switch to a supply commitment for the pneumococcus pilot. Having generalized the results to an arbitrary number N of competing firms, we see that our earlier conclusion might have been too hasty but in fact was still correct. There were two viable suppliers of a second-generation pneumococcus vaccine when the Group was deliberating, Pfizer and GSK; and this is a value of N for which according to Proposition 12 supply commitments outperform framework AMCs. Indeed, even if a third viable supplier were to emerge, Figure 5 suggests that any advantage of a framework AMC over a supply commitment would likely be small.

6. Design for Technologically Distant Products

In this section we move from an analysis of a technologically close to a technologically distant product. One might naïvely think that the preceding analysis for a close product carries over fairly directly to a distant one with little change beyond the obvious need to scale up the fund F so that it covers R&D cost R required for a still-to-be-invented product in addition to the capacity cost KQ. In fact, the difference between the optimal design of an AMC for close and distant products runs deeper than that. This section will not provide an exhaustive survey of the differences but will focus on some key ones in each of three subsections.

Section 6.1 describes a featured role gained by the country copayment, as a "kill switch" for products that meet the technical specifications of the AMC contract but are not particularly valuable for consumers because the valuable features were difficult to anticipate or specify that far in advance. As we will see, the donor has difficulty committing to use the kill switch, so giving control of it to the country via the copayment can lead to more powerful incentives for quality provision.

Section 6.2 points out a sense in which policy design is less sensitive with distant than close products. An AMC for a close product was focused on incentivizing capacity investment, which as we saw was sensitive to the policy details beyond the funding level. An AMC for a distant product shifts at least some of the focus to incentivizing R&D. For this type of investment, what matters is the level of rents a policy delivers rather than the shape of the policy. The point is made with a surprising demonstration, taking the AMC design that was useless for a close product—a framework AMC with a temporary subsidy—and showing that it can be used to improve investment incentives for a distant product.

Section 6.3 shows that an AMC for a distant product is not necessarily more expensive to run than for a close product despite an extra investment R needing to be covered. Surprisingly, the donor may prefer the distant product. As a product comes closer to being developed, the firm learns more about the nature of its production process, exacerbating the information asymmetry between donor and firm. Thus there is a tradeoff between paying for R with a distant product versus paying the information rent with a close product. Indeed the balance is tipped somewhat in favor of a distant product because not all of R needs to be covered by the AMC for reasons we will discuss.

6.1. Copayment as "Kill Switch"

This subsection analyzes perhaps the most significant new issue with a technologically distant product. The country copayment, which plays only a supporting role with a close product as discussed further in Appendix B), gains a featured role with distant products. With a close product, there was little danger of wasting AMC funds on a product with limited consumer value because the donor could see the nature of the available products and only commission an AMC if suitable ones happen to exist, writing the product specifications so that only products of sufficient quality are eligible. With a distant product, contractual incompleteness becomes a much larger problem. It may be difficult to write product specifications preventing a low-value product from worming its way into the program. If the donor is allowed an overly strict interpretation of product specifications, it may be able to hold up firms' investments, undermining the commitment value of the AMC. This is where the country copayment comes in, and where the "market" feature of AMCs gains force. Confronting the country with a small copayment creates a kill switch for low-value products which, in the model introduced momentarily, incentivizes the firm to produce high-value products. Giving the kill switch to the donor does not work as well as to the country because the donor cannot commit not to renegotiate with the firm to split the AMC funds locked in the escrow.

Return to the basic model of Section 2; in particular return to the assumption of a monopoly firm and symmetric cost information (so there is again only one type of production cost, c). We will extend the basic model to a distant product as follows. The firm has the opportunity to invest $R > 0$ in R&D. Doing so results in a useful product providing health benefit $\alpha B(q)$ and requiring the positive capacity cost K and production cost c per unit assumed to this point. If the firm does not invest R, it produces a useless product providing no health benefit but also involving no associated capacity or production costs. Because of contractual incompleteness, the terms of the AMC written at the distant time before R needs to be invested cannot distinguish between the useful and useless product.

We seek to analyze the role of the country copayment as a "kill switch" in this setting. For comparison, we first analyze how an AMC without a country copayment performs in this setting. Take the design we found in Section 3 to be efficient: a forcing contract. We will show this design can be improved on for some parameters by adding a country copayment.

As before, the disposition of unused AMC funds has a crucial impact on incentives. As discussed in the initial setup of the model in Section 2, if the funds flow back to the donor, the AMC has no commitment value because the opportunity cost of the funds will net out in Nash bargaining. Thus, the AMC must specify the funds stay in the escrow. But then the donor gains from bargaining with the firm over trade of even the useless product because they can fulfill the terms of the forcing contract costlessly, dividing the otherwise locked funds F between them, βF for the firm and $(1-\beta)F$ for the donor. To donor cannot commit to pull the "kill switch" on the useless product.

The conclusion from this analysis is that the firm earns βF if it does not invest in R&D. If it invests, subtracting the additional R&D cost R from equation (36), we see that it earns $\pi(Q^{IN}, \theta) + \beta F - R$. Hence the firm is willing to invest in R&D if and only if $R \leq \pi(Q^{IN}, \theta)$. Otherwise, there is no investment, no useful product, and no social surplus. Thus a forcing contract without a country copayment can do nothing to promote R&D incentives.

We will show that efficiency can be improved for a large set of parameters if a small country copayment s_c is added to the forcing contract. To keep the analysis simple, we will take s_c to be infinitesimal. Efficiency could be further improved by taking some appropriately calibrated, moderate value of s_c. The logic of why this is so is discussed in detail in Appendix B: substituting some country for donor funds benefits the donor because it does not internalize the country's expenditures, just its health benefit. Thus, we do not need to complicate the analysis to accommodate that point here. As long as $s_c < \alpha_c B(Q^{IN})/Q^{IN}$, a condition of course satisfied by an infinitesimal copayment, the copayment will not lead the country to demand less than full capacity Q^{IN}. On the other hand, a positive copayment, no matter how small, leads the country to pull the "kill switch" on the useless product. Now the firm earns 0 if it does not invest compared to the $\pi(Q^{IN},\theta)+\beta F-R$ it earns if it does, implying that the firm is now willing to invest if and only if $R \leq \pi(Q^{IN},\theta)-\beta F$. The optimal fund size F^{DP} (where the superscript indicates the case of a "distant product") satisfies this condition with equality: $F^{DP}=[\pi(Q^{IN},\theta)-R]/\beta$. The proof (provided in the appendix) of the next proposition shows that the donor benefits from implementing an AMC of this design as long as R&D investment would be efficient for an integrated entity, which is equivalent to the condition $R < w(Q^{IN},\theta)$.

Proposition 13. *R&D investment would be efficient for an integrated entity yet cannot be induced by an equilibrium forcing contract without country copayment if and only if* $R \in \left(\pi(Q^{IN},\theta), w(Q^{IN},\theta)\right)$. *If this condition holds, adding an infinitesimal country copayment to the forcing contract can not only induce efficient R&D investment but also allow the donor to appropriate all the net surplus from the integrated outcome for itself.*

6.2. Incentives at the Extensive Versus Intensive Margin

As we saw in Section 3, the performance of an AMC for technologically close products depends on the shape of the scheme as much as the level of funding. For example, an AMC involving a forcing contract could induce efficient capacity Q^{IN} using a fund of size F^{FC}, while applying those same funds F^{FC} to a framework AMC with a temporary subsidy would add nothing to incentives for capacity investment (indeed no other funding level would do any better). The nuances of supply commitments, forcing contracts, price caps all mattered for capacity incentives. With a distant product, the AMC targets R&D as well as capacity investment. In the model, R&D is a fundamentally different type of investment than capacity.[5] Incentives

[5] That the R&D variable R is dichotomous while capacity variable Q is continuous is not the crucial difference. In an alternative model in which R is specified as a continuous variable, increases in which increase the probability that the product is useful, R&D

for R&D operate at the extensive margin, depending on the level not the slope of the policy. Incentives for capacity operate at the intensive margin, where slope and other design features become crucial.

We will illustrate these points in this subsection with a stark demonstration. We will take the AMC design found to be useless for a close product—a framework AMC with a temporary subsidy—and show it provides useful incentives for a distant product.

Consider a model close to the one from the previous subsection with a monopoly firm and one cost type c. Here we will abstract away from the moral hazard problem in the previous subsection, whereby the firm could save money by producing a useless product. Now assume that there is no choice but to develop a useful product, requiring positive investment R to do so.

By Proposition 3, a framework AMC with temporary subsidy is useless for capacity incentives. Thus the equilibrium capacity is Q^{NA} with or without such an AMC. The only possible benefit of such an AMC is to improve R&D incentives. If $R \leq \pi(Q^{NA}, \theta)$, then the firm is willing to invest in R&D even in the absence of an AMC. Thus for a framework AMC with temporary subsidy to have the possibility of enhancing R&D incentives relative to no AMC, we must have $R > \pi(Q^{NA}, \theta)$. Suppose this inequality holds.

If the firm does not invest in R&D, the firm's profit is 0 because nothing is produced. If it invests in R&D, in the presence of a framework AMC with temporary subsidy, its ex ante profit is $\pi(Q^{NA}, \theta) + \beta F - R$, which one can see by subtracting the additional R&D cost R from equation (13). Thus the firm is willing to invest in R&D investment if and only if $F \geq [R - \pi(Q^{NA}, \theta)]/\beta$. The efficient fund size satisfies this condition with equality: $F^{DP} = [R - \pi(Q^{NA}, \theta)]/\beta$. The proof (in the appendix) of the following proposition shows that the donor strictly benefits from offering a framework AMC with this fund size if and only if $R < w(Q^{NA}, \theta)$.

Proposition 14. *A framework AMC with temporary subsidy can strictly improve donor surplus relative to no AMC if and only if* $R \in \left(\pi(Q^{NA}, \theta), w(Q^{NA}, \theta)\right)$.

Some remarks about the proposition are in order. The condition on R in this proposition is identical to that in Propositions 13 except for the capacity used in the right-hand side interval, here Q^{NA} rather than Q^{IN}. Proposition 14 does not say the framework AMC is fully efficient. It provides efficient R&D incentives but still does not help with capacity incentives. To improve incentives for capacity investment would require a different design for the AMC, involving a permanent subsidy, supply commitment, or better yet a forcing contract.

incentives would still depend primarily on the fund size F. We acknowledge the possibility of constructing other alternative models in which the incentive margins are flipped with R&D incentives operating at the intensive while capacity incentives at the extensive margin. However, the model we focus on seems nature and in any event serves to illustrate possibile differences between R&D and capacity incentives not necessary differences.

6.3. Preferring Distant Products

We saw in the previous subsection that designing an AMC for a technologically distant product can be a conceptually less complex problem than for a close product. In this subsection we show that the problem can be not just conceptually simpler but in fact cheaper for the donor. This is initially surprising because an extra investment R must be covered for a distant product, an investment assumed already sunk in the case of a close product. The tradeoff is that the problem of asymmetric cost information may be more severe with a close product; with a distant product, the horizon may be long enough that the firm and donor may be symmetrically ill informed about what costs will turn out to be.

To explore this idea, we start with the model of cost uncertainty from Section 4. This model will apply to the case of a close product perfectly. No additional investment R needs to be considered because this is assumed to have been sunk before the start of the game. In the case of a distant product, we will assume that the firm must invest $R > 0$ for any product to be produced. As in the previous subsection, if this investment is made, the product is certain to provide a health benefit. Assume that the AMC is offered far enough in advance of production that the donor and firm share the same uncertainty about c, both knowing its distribution—$\underline{c}$ with probability $\nu \in (0,1)$ and $\bar{c}$ with probability $1-\nu$—but not the realization at that point. We will restrict attention to the AMC design found to be efficient in Section 3.6: forcing contracts. Defining the expectations operator as

$$\mathrm{E}\big[\pi(Q^{NA}(c),c)\big] = \nu\pi(Q^{NA}(\underline{c}),\underline{c}) + (1-\nu)\pi(Q^{NA}(\bar{c}),\bar{c}), \tag{58}$$

we have the following proposition.

Proposition 15. *Consider the models of a close and distant product, and suppose the donor offers the forcing contract that is efficient in the given model. Let $\hat{\beta} \in (0,1)$ be the threshold level of β from Proposition 8.*

- *If $\beta > \hat{\beta}$, the donor is indifferent between the close and distant product if*

$$R - E\big[\pi(Q^{NA}(c),c)\big] \tag{59}$$

  is negative and strictly prefers a close product if this expression is positive.

- *If $\beta < \hat{\beta}$, the donor prefers the close product if expression* (59) *exceeds* (39) *and the distant product if the reverse inequality holds.*

The appendix provides a formal proof. To derive some intuition, we will examine the cases based on β separately. First suppose $\beta > \hat{\beta}$. In that case, as Proposition 8 states, the existence of asymmetric cost information does not harm the donor with a close product. Because the binding constraint for the low-cost type is its participation not truth-telling constraint, the optimal mechanism involves separate first-best forcing contracts for each cost type. With no cost of asymmetric information to trade off against the potentially higher investment cost that needs to be covered, the donor can certainly do no better with a distant product in this case. It is easy to see that must do strictly worse with a distant product for sufficiently high values of R.

What is less clear is how a distant product can ever achieve a tie with a close product when $\beta > \hat{\beta}$, i.e., when the donor suffers no loss due to asymmetric information with a close product. To see how, note that the donor does not have to cover the full R&D cost R with a distant product but just (59). The firm's ex ante expected profit if it rejects the AMC and engages in Nash bargaining equals $\mathrm{E}\left[\pi(Q^{NA}(c),c)\right]$, which the donor must concede to the firm to ensure it does not reject the AMC contract even in the absence of R&D costs. For a distant product, it is only the excess of R above $\mathrm{E}\left[\pi(Q^{NA}(c),c)\right]$ that must be subsidized to ensure the firm invests in R&D given it is willing to sign the AMC contract.

Next suppose $\beta < \hat{\beta}$. Now, by Proposition 8, the presence of asymmetric cost information reduces the donor's surplus by (39). If this loss exceeds the cost (59) of inducing R&D investment with a distant product, then the donor prefers a distant product; if the reverse is true, the donor prefers the close product.

The proposition provides a full characterization of the donor's preferences between a close and distant product for all parameters, so its statement is somewhat involved. It is worth emphasizing the particular implication we set out to prove, that is, the existence of a non-emtpy set of parameters for which the donor prefers a distant to a close product. The proposition shows this parameter set is indeed non-empty, arising for sufficiently low values of β and R.

7. Conclusion

This paper provides a theoretical analysis of the effectiveness of AMCs in incentivizing investment in new products (such as a vaccine) for low-income countries. We analyze a number of alternative design features, some of which are found to enhance incentives, some of which detract.

Our key contribution may not be particular results about particular design features but rather the broader theoretical framework, which modles the AMC as being layered over existing policies that would have proceeded in its absence. In this framework, the incentive effects of an AMC becomes a subtler question. The infusion of even considerable AMC funds may merely replace other funds that the donor diverted to other purposes, muting the incentive effects of the AMC. Indeed, we saw that the AMC may have no effect in extreme cases. Incentive power is not a foregone conclusion but requires careful design.

Additional themes emerging from the analysis include that AMC design for a technologically close product (as with pneumococcus vaccine in the pilot) involves quite different considerations than the design for a technologically distant product that is not far along in research and development. Surprisingly, the design for a technologically close product may be more complex. Asymmetry of cost information entails the leakage of information rents to firms, which can be disciplined by possible supply restrictions at some deadweight loss. Manufacturer competition helps expand supply whether or not an AMC is introduced, but competition does interact with the AMC, promoting its effectiveness in a way that is less sensitive to fine design details.

Appendix A: Proofs

This appendix provides proofs omitted from the text of lemmas and propositions as well as the proof of an unnumbered claim regarding the properties of Figure 4.

Proof of Lemma 1

To apply the inverse function theorem to $P(q)$, we need to verify that it is continuously differentiable with nonzero derivative. The assumption that $B(q)$ is twice continuously differentiable for all $q \geq 0$ implies $P'(q) = \alpha B''(q)$ is continuous, in turn implying that $P(q)$ is continuously differentiable for all $q \geq 0$. The assumption that $B''(q) < 0$ for all $q \geq 0$ implies $P'(q) = \alpha B''(q) < 0$, in turn implying $P'(q) \neq 0$ for all $q \geq 0$. This establishes that $P(q)$ satisfies the conditions for the inverse function theorem for all $q \geq 0$. The theorem then implies that the inverse $\tilde{D}(P(q)) = P^{-1}(P(q))$ exists for all $q \geq 0$ and that

$$\tilde{D}'(P(q)) = \frac{1}{P'(q)}. \tag{A1}$$

To bound the codomain of $P(q)$, note $\lim_{q\to\infty} = \alpha \lim_{q\to\infty} B'(q) = 0$, where the first equality follows from the definition of $P(q)$ and the second by assumption on the health benefit. Further, $\lim_{q\to 0} P(q) = \bar{p}$ by definition. We showed that $P(q)$ is continuously differentiable for all $q > 0$, implying it is continuous for all $q > 0$. Thus for all $p \in (0, \bar{p})$, there exists $q > 0$ such that $P(q) = p$. Hence for all $p \in (0, \bar{p})$, $\tilde{D}(p) = \tilde{D}(P(q)) = q > 0$, and $\tilde{D}'(p) = \tilde{D}'(P(q)) = 1/P'(q) < 0$ by equation (A1). *Q.E.D.*

Proof of Proposition 2

To complete the proof, we need to fill in several details omitted from the argument in the text. First, we will verify that $Q^{NA} < D(c)$, where $Q^{NA} = D(c+\kappa/\beta)$. There are two cases to consider. If $c+\kappa/\beta \geq \bar{p}$ then $D(c+\kappa/\beta) = 0 < D(c)$, where the last inequality was argued in the text preceding equation (6). On the other hand, if $c+\kappa/\beta < \bar{p}$, then by Lemma 1, $D'(p) < 0$ for all $p \in [c, c+\kappa/\beta]$, implying $D(c+\kappa/\beta) < D(c)$ by the fundamental theorem of calculus. In either event, we have verified $Q^{NA} < D(c)$.

Second, we need to verify that $D(c+\kappa/\beta) < D(c+\kappa)$. The argument follows exactly that in the previous paragraph and is thus omitted.

Finally, we need to verify the limits. The limit $\lim_{\beta\to 1} Q^{NA} = Q^{IN}$ is obvious. To see the other limit, $\lim_{\beta\to 0} Q^{NA} = \lim_{p\to\infty} D(p) = D(\bar{p}) = 0$. *Q.E.D.*

Proof of Proposition 3

We need to show that the output q^A during the AMC period that maximizes the present discounted value of the donor's and firm's joint surplus equals capacity Q. This present discounted value is

$$\int_0^T [\alpha B(q^A) - cq^A + sq^A] e^{-rT} dt + \int_T^\infty [\alpha B(Q) - CQ] e^{-rT} dt. \tag{A2}$$

We argued in the text that the principle of full capacity utilization holds in the tail period, hence the substitution of Q for output in the second term of (A2).

Before proceeding, we need to provide some analysis of AMC-period length T appearing in (A2), which is an implicit function of q^A according to (10). Solving (10) to obtain T as an explicit function of q^A,

$$T = \frac{1}{r} \ln\left(\frac{sq^A}{sq^A - rF}\right). \tag{A3}$$

Differentiating,

$$\frac{\partial T}{\partial q^A} = \frac{-F}{q^A(sq^A - rF)}. \tag{A4}$$

A temporary subsidy satisfies $sq^A > rF$, implying (A4) is negative.

Returning to (A2), we will show it is increasing in q^A for all $q^A \leq Q$ such that $Q < D(c)$, which holds in equilibrium. Differentiating (A2) with respect to q^A,

$$\frac{1}{r}[P(q^A) - c + s](1 - e^{-rT}) + sq^A e^{-rT} \cdot \frac{\partial T}{\partial q^A} + \left\{[\alpha B(q^A) - cq^A] - [\alpha B(Q) - cQ]\right\} e^{-rT} \cdot \frac{\partial T}{\partial q^A}. \tag{A5}$$

The last term is non-negative. To see this, note $\alpha B(q) - cq$ is concave in q, maximized for $q = D(c)$. This implies $\alpha B(q^A) - cq^A \leq \alpha B(Q) - cQ$ for $q^A \leq Q < D(c)$. Combined with $\partial T/\partial q^A < 0$, we have that the last term in (A5) is non-negative. Thus (A5) weakly exceeds

$$\frac{1}{r}[P(q^A)-c+s](1-e^{-rT})+sq^A e^{-rT}\cdot\frac{\partial T}{\partial q^A} \tag{A6}$$

$$= \frac{F}{sq^A}[P(q^A)-c], \tag{A7}$$

where (A7) holds by substituting for $\partial T/\partial q^A$ from (A4); substituting $s(1-e^{-rT})/r = F/q$, which holds by (10), into the first term in (A6); and substituting $e^{-rT} = (sq^A - rF)/sq^A$, which again holds by (10), into the second term in (A6). For $q^A \le Q < D(c)$, $P(q^A) > c$, implying (A7) is positive for all $q^A \le Q < D(c)$. *Q.E.D.*

Proof of Proposition 5

We first verify the claim in the text that the firm's participation constraint is equivalent to the non-negativity constraint on the subsidy. The firm participates if its ex ante profit under the framework AMC, given in equation (16), weakly exceeds its ex ante profit in the absence of an AMC, given by (7), i.e., $\pi(Q,\theta)+\beta sQ/r \ge \pi(Q,\theta)$, which reduces to $s \ge 0$.

We next verify that an implicit solution Q^{PS} to (21) exists, which is equivalent to the existence of a root Q^* of (17). We will show that the left-hand side of (17) is positive for $Q = 0$, negative in the limit as $Q \to \infty$, so by continuity there exists a root. Since $B(q)$ is twice continuously differentiable for $q \ge 0$, $B''(0)$ is finite, implying $P'(0) = \alpha B''(0)$ is finite. Thus at $Q = 0$ the left-hand side of (17) is $P(0)-(c+\kappa) = \bar{p}-(c+\kappa)$, which is positive by Assumption (1). At the other extreme,

$$\lim_{Q\to\infty}[P(Q)+\beta QP'(Q)] \le \lim_{Q\to\infty} P(Q) = \alpha \lim_{Q\to\infty} B'(Q) = 0,$$

where the first step follows from $P'(Q) < 0$, the second from the definition of $P(Q)$, the third from an assumed property of the health benefit function. Since $c+\kappa > 0$, the left-hand side of (17) is negative in the limit as $Q \to \infty$.

With those facts verified, the text preceding the proposition proves that an nontrivial, feasible framework AMC (NFFA) exists if and only if $s^{PS} > 0$. We have

$$s^{PS} = c+\frac{\kappa}{\beta}-P(Q^{PS}) = \left(\frac{1-\beta}{\beta}\right)\kappa+\beta Q^{PS}P'(Q^{PS}), \tag{A8}$$

where the first equality follows from (17) and the second from (21). Taking limits in (A8), $\lim_{\beta\to1} s^{PS} = \lim_{\beta\to1}[Q^{PS}P'(Q^{PS})] < 0$. Thus $s^{PS} < 0$—implying a NFFA does not exist—for β sufficiently close to 1. Examining the limit at the opposite extreme, $\lim_{\beta\to0} Q^{PS} = D(c+\kappa) = Q^{IN}$, implying

$$\lim_{\beta\to0} s^{PS} = \kappa \lim_{\beta\to0}\left(\frac{1-\beta}{\beta}\right)+0\cdot Q^{IN}P'(Q^{IN}) = \infty.$$

Thus $s^{PS} > 0$—implying an NFFA exists—for β sufficiently close to 0.

To establish the non-existence of an NFFA in the limit $\kappa \to 0$, arguments used the second paragraph of this proof can be used to shows that (21) has a finite solution for $\kappa = 0$; call this solution Q^0. Taking limits in (A8), $\lim_{\kappa\to0} s^{PS} = \beta Q^0 P'(Q^0) < 0$. Thus $s^{PS} < 0$—implying a NFFA does not exist—for κ sufficiently close to 0.

To complete the proof, we establish the remaining claims, which require assumptions on the concavity of the donor's objective function. We will prove the result about the threshold value $\hat{\beta}$. The proof for $\hat{\kappa}$ is similar and omitted. Denote the donor's objective function in (19) by $\Delta(Q,\theta)$, where θ is the vector of model parameters. The requisite concavity assumption is $\Delta_{QQ}(Q,\theta) < 0$, where subscripts indicate partial derivatives. We will show that under this assumption $\partial s^{PS}/\partial\beta < 0$. Combined with the limits established above, $\lim_{\beta\to0} s^{PS} > 0$ and $\lim_{\beta\to1} s^{PS} < 0$, these results are sufficient for the existence of a threshold $\hat{\beta} \in (0,1)$ such that $s^{PS} > 0$ for $\beta < \hat{\beta}$ and $s^{PS} \le 0$ for $\beta \ge \hat{\beta}$.

We will use the implicit function rule to derive the sign of $\partial s^{PS}/\partial\beta$. The first-order condition $\Delta_Q(Q^{PS},\theta) = 0$ gives Q^{PS} as an implicit function of β. This first-order condition can be transformed so that s^{PS} appears as an implicit function of β. To do so, solve the incentive-compatibility constraint (17) for Q as a function of the other variables, obtaining $Q^*(s,\theta) = D(c+\kappa/\beta-s)$. For reference, we can sign the following partial derivatives:

$$Q^*_s(s,\theta) = -D'\left(c+\frac{\kappa}{\beta}-s\right) > 0 \tag{A9}$$

$$Q^*_\beta(s,\theta) = -D'\left(c+\frac{\kappa}{\beta}-s\right)\frac{\kappa}{\beta^2} > 0. \tag{A10}$$

Substituting $Q^*(s,\theta)$ into the donor's first-order condition for Q gives

$$\Delta_Q(Q^*(s^{PS},\theta),\theta) = 0 \tag{A11}$$

Applying the implicit function rule to (A11) yields

$$\frac{\partial s^{PS}}{\partial \beta} = -\frac{\Delta_{QQ}(Q^*(s^{PS},\theta),\theta)Q^*_{\beta}(s^{PS},\theta) + \Delta_{Q\beta}(Q^*(s^{PS},\theta),\theta)}{\Delta_{QQ}(Q^*(s^{PS},\theta),\theta)Q^*_{s}(s^{PS},\theta)}. \tag{A12}$$

All the terms in (A12) have been signed except for one, which we can sign by differentiating (A11) with respect to β, yielding

$$\Delta_{Q\beta}(Q^*(s^{PS},\theta),\theta) = \frac{1}{r}Q^*(s^{PS},\theta)P'(Q^*(s^{PS},\theta)) < 0. \tag{A13}$$

Substituting the signs in (A9), (A10), and (A13) as well as the concavity of $\Delta(Q,\beta)$ into (A12) implies $\partial s^{PS}/\partial\beta < 0$. *Q.E.D.*

Proof of Proposition 6

We need to fill in the remaining detail that the fund size F^{SC} under the optimal supply commitment equals the fund size F^{PS} under the optimal framework AMC with perpetual subsidy. We first compute F^{SC}. Substituting equilibrium capacity under the supply commitment, which we showed in the text satisfies $Q^{SC} = Q^{PS}$, for Q in (28) and solving for F yields

$$F^{SC} = \frac{Q^{PS}}{r}\left[c + \frac{\kappa}{\beta} - P(Q^{PS})\right]. \tag{A14}$$

According to the text following Proposition 4

$$F^{PS} = \frac{s^{PS}Q^{PS}}{r} = \frac{Q^{PS}}{r}\left[c + \frac{\kappa}{\beta} - P(Q^{PS})\right]. \tag{A15}$$

Equations (A14) and (A15) together imply $F^{SC} = F^{PS}$. *Q.E.D.*

Proof of Proposition 8

The optimal direct revelation mechanism is the choice of $\underline{Q}$, $\bar{Q}$, $\underline{F}$, $\bar{F}$ maximizing expected donor surplus

$$\nu\left\{\frac{1-\beta}{r}\left[\alpha B(\underline{Q}) - \underline{c}\underline{Q}\right] - \beta\underline{F}\right\} + (1-\nu)\left\{\frac{1-\beta}{r}\left[\alpha B(\bar{Q}) - \bar{c}\bar{Q}\right] - \beta\bar{F}\right\} \tag{A16}$$

subject to participation constraints for types $\underline{c}$ and $\bar{c}$, respectively,

$$\pi(\underline{Q},\underline{c}) + \beta\underline{F} \geq \pi(Q^{NA}(\underline{c})),\underline{c}) \tag{A17}$$

$$\pi(\bar{Q},\bar{c}) + \beta\bar{F} \geq \pi(Q^{NA}(\bar{c})),\bar{c}) \tag{A18}$$

and truth-telling constraints for types $\underline{c}$ and $\bar{c}$, respectively,

$$\pi(\underline{Q},\underline{c}) + \beta\underline{F} \geq \pi(\bar{Q},\underline{c}) + \beta\bar{F} \tag{A19}$$

$$\pi(\bar{Q},\bar{c}) + \beta\bar{F} \geq \pi(\underline{Q},\bar{c}) + \beta\underline{F}. \tag{A20}$$

The left-hand side of constraints (A17)–(A20) use the expression for ex ante firm profit under a forcing contract provided in (36).

Begin by supposing that constraints (A17) and (A20) do not bind. Then the remaining constraints obviously bind. We can use (A18) and (A19) to solve for $\beta\underline{F}$ and $\beta\bar{F}$. Substituting these values into (A16) leaves the following unconstrained problem in two choice variables, $\underline{Q}$ and $\bar{Q}$:

$$\begin{aligned}\nu\left\{\frac{1-\beta}{r}\left[\alpha B(\underline{Q}) - \underline{c}\underline{Q}\right] - \pi(Q^{NA}(\bar{c}),\bar{c}) + \pi(\bar{Q},\bar{c}) - \pi(\bar{Q},\underline{c}) + \pi(\underline{Q},\underline{c})\right\}\\ + (1-\nu)\left\{\frac{1-\beta}{r}\left[\alpha B(\bar{Q}) - \bar{c}\bar{Q}\right] - \pi(Q^{NA}(\bar{c}),\bar{c}) + \pi(\bar{Q},\bar{c})\right\}\end{aligned} \tag{A21}$$

The first-order condition with respect to $\underline{Q}$ yields $\underline{Q}^* = D(\underline{c}+\kappa) = Q^{IN}(\underline{c})$. The first-order condition with respect to $\bar{Q}$ after rearranging yields the expression in (38). Substituting $\underline{Q}^*$ and $\bar{Q}^*$ back into (A18) and (A19) yields

$$\bar{F}^* = \frac{1}{\beta}\left[\pi(Q^{NA}(\bar{c}),\bar{c}) - \pi(\bar{Q}^*,\bar{c})\right] \tag{A22}$$

$$\underline{F}^* = \bar{F}^* + \frac{1}{\beta}\left[\pi(\bar{Q}^*,\underline{c}) - \pi(\underline{Q}^*,\underline{c})\right]. \tag{A23}$$

We are left to check the ignored constraints. It is easy to see that (A20) is satisfied by the solution. It remains to check (A17). Since (A19) holds with equality and the left-hand sides of (A17) and (A19) are the same, (A17) is satisfied if and only if the right-hand side of (A19) is at least as high as the right-hand side of (A17):

$$\pi(\bar{Q}^*,\underline{c})+\beta\bar{F}^* \geq \pi(Q^{NA}(\underline{c}),\underline{c}). \tag{A24}$$

Substituting from (A22) for $\bar{F}^*$ and rearranging, we can write (A24) as

$$\begin{aligned}
0 \quad &\leq \quad \left[\pi(Q^{NA}(\bar{c}),\bar{c})-\pi(Q^{NA}(\underline{c}),\underline{c})\right]-\left[\pi(\bar{Q}^*,\bar{c})-\pi(\bar{Q}^*,\underline{c})\right] && \text{(A25)}\\
&= \quad \int_{\underline{c}}^{\bar{c}} \frac{d\pi(Q^{NA}(c),c)}{dc}dc-\int_{\underline{c}}^{\bar{c}} \frac{\partial\pi(\bar{Q}^*,c)}{\partial c}dc && \text{(A26)}\\
&= \quad -\int_{\underline{c}}^{\bar{c}} Q^{NA}(c)dc+\int_{\underline{c}}^{\bar{c}} \bar{Q}^* dc && \text{(A27)}\\
&= \quad \int_{\underline{c}}^{\bar{c}} \left[D\left(\bar{c}+\kappa+\frac{\nu}{1-\nu}\beta(\bar{c}-\underline{c})\right)-D\left(c+\frac{\kappa}{\beta}\right)\right]dc. && \text{(A28)}
\end{aligned}$$

Equation (A26) follows from the fundamental theorem of calculus; (A27) follows from differentiating (8) and applying the envelope theorem; (A28) follows from (38) and the formula $Q^{NA}(c)=D(c+\kappa/\beta)$.

Hence (A17) is satisfied if (A28) is non-negative. One can show that (A28) is decreasing in β, positive in the limit as $\beta \to 0$, negative in the limit as $\beta \to 1$. Hence there exists $\beta \in (0,1)$ for which the left-hand side of (A28) is exactly 0. This is our threshold $\hat{\beta}$.

The proof is completed by deriving the difference in expected donor surplus (39). Expected donor surplus under symmetric information is

$$\nu\left[w(Q^{IN}(\underline{c}),\underline{c})-\pi(Q^{NA}(\underline{c}),\underline{c})\right]+(1-\nu)\left[w(Q^{IN}(\bar{c}),\bar{c})-\pi(Q^{NA}(\bar{c}),\bar{c})\right] \tag{A29}$$

because for each cost realization c, the donor can offer a forcing contract allowing it to extract the integrated surplus $w(Q^{IN}(c),c)$ defined in (3) less the surplus $\pi(Q^{IN}(c),c)$ that must be offered to the firm to gain its participation. Substituting the equilibrium contract terms into (A16) and subtracting from (A29), after rearranging, yields the difference (39) stated in the proposition. *Q.E.D.*

Proof of Lemma 2

Suppose that equilibrium capacity in the absence of an AMC Q_N^{NA} satisfies $Q_N^{NA} < D(c)$. Then we can invoke the principle of full capacity utilization. The ex ante profit of firm *i*, equal to the present discounted value of its Shapley value each ex post instant less ex ante capacity cost, is

$$\frac{1}{r}\left\{\sum_{n=1}^{N} w(n,N)\left[\alpha B(Q_i+(n-1)Q_j)-\alpha B((n-1)Q_j)-cQ_i\right]-\kappa Q_i\right\}, \tag{A30}$$

Shapley-value weight $w(n,N)$ equals the unconditional probability that a coalition including *i*—randomly formed by permuting a grand coalition of *N* firms and the donor—happens to have exactly $n-1$ other firms (so *n* firms total) as well as the donor. We prove below that $w(n,N)$ is given by equation (48). Taking the first-order condition with respect to Q_i and imposing $Q_i = Q_j = Q/N$ in a symmetric equilibrium yields, after rearranging, the condition for equilibrium industry capacity conditional on the AMC terms:

$$\sum_{n=1}^{N} w(n,N)\left[P(nQ/N)-c\right]-\kappa = 0. \tag{A31}$$

Multiplying both sides by 2, using the facts that $2w(n,N)=\tilde{w}(n,N)$ and $\sum_{n=1}^{N}\tilde{w}(n,N)=1$, and using the definition of $\mathbf{P}(Q,N)$, equation (A31) can be shown to be equivalent to (52).

Let $\mathbf{D}(P,N)$ be the inverse of $\mathbf{P}(Q,N)$, i.e., $\mathbf{D}(\mathbf{P}(Q,N),N)=Q$, which can be interpreted as the virtual demand arising from Shapley bargaining. This virtual demand exists because $\mathbf{P}(Q,N)$ is a linear combination of invertible functions $P(Q)$. Equilibrium quantity in the absence of an AMC can be derived by substituting $\theta = 0$ into (53) and inverting, yielding $\mathbf{D}(c+2\kappa)$. The principle of full capacity utilization is thus

$$\mathbf{D}(c+2\kappa,N) < D(c). \tag{A32}$$

One can apply the inverse function to both sides of (A32) to show it is equivalent to (51).

To show (51) is non-vacuous, note that the text preceding equation (6) argued that $D(c)$ is a finite, positive number, implying $\mathbf{P}(D(c),N)$ is as well. Therefore, (A32) holds for sufficiently large κ and is thus non-vacuous.

The proof is completed by deriving the formula (48) posited for $w(n,N)$. Shapley weight $w(n,N)$ equals a fraction, the numerator of which equals the number of permutations of the players in which *i* is preceded by the donor and exactly $n-1$ rivals, the denominator of which equals the $(N+1)!$ total permutations of the players. We need to count the permutations in the numerator.

The $N-1$ rivals can fill the $N-n$ positions after i in

$$\frac{(N-1)!}{[(N-1)-(N-n)]!} = \frac{(N-1)!}{(n-1)!}$$

ways. The remaining $n-1$ rivals and the donor can fill the n positions preceding i in $n!$ ways. Thus,

$$w(n,N) = \frac{[(N-1)!/(n-1)!]n!}{(N+1)!},$$

which reduces to (48). *Q.E.D.*

Proof of Proposition 9

Differentiating,

$$\frac{\partial \mathbf{P}(Q,N)}{\partial Q} = \frac{2}{N^2(N+1)} \sum_{n=1}^{N} n^2 P'(nQ/N) < 0,$$

where the inequality holds because $P'(Q) < 0$ for all $Q \geq 0$. Thus

$$\begin{aligned}
\mathbf{P}(Q,N) &= \tilde{w}(N,N)P(Q) + \sum_{n=1}^{N-1} \tilde{w}(n,N)P(nQ/N) \\
&> \tilde{w}(N,N)P(Q) + \sum_{n=1}^{N-1} \tilde{w}(n,N)P(Q) \\
&= \sum_{n=1}^{N} \tilde{w}(n,N)P(Q) \\
&= P(Q) \\
&= \mathbf{P}(Q,1),
\end{aligned}$$

The second follows from $P'(Q) < 0$ and $n/N < 1$ for $n \leq N-1$. The remaining lines are either rearrangements or hold by definition. Thus $\mathbf{P}(Q_N^{NA},1) < \mathbf{P}(Q_N^{NA},N) = c+2\kappa = \mathbf{P}(Q_1^{NA},1)$, where the inequality was just established and the equalities follow from (52). But $\mathbf{P}(Q_N^{NA},1) < \mathbf{P}(Q_1^{NA},1)$ together with $\partial\mathbf{P}/\partial Q < 0$ implies $Q_N^{NA} > Q_1^{NA}$. *Q.E.D.*

Proof of Proposition 10

Assuming $P(Q) = a - bQ$ for Q in the relevant range,

$$\begin{aligned}
\mathbf{P}(Q,N) &= \sum_{n=1}^{N} \tilde{w}(n,N)\left[a - b\left(\frac{nQ}{N}\right)\right] \\
&= a - \frac{2bQ}{N^2(N+1)} \sum_{n=1}^{N} n^2 \\
&= a - \left[\frac{2N+3}{3(N+1)}\right] bQ,
\end{aligned}$$

where the last line follows from $\sum_{n=1}^{N} n^2 = n(N+1)(2N+1)/6$. Similarly,

$$\mathbf{P}(Q,N+1) = a - \left[\frac{2N+3}{3(N+1)}\right] bQ.$$

Combining these results and rearranging,

$$\mathbf{P}(Q,N+1) - \mathbf{P}(Q,N) = \frac{bQ}{3N(N+1)} > 0.$$

By induction, $\mathbf{P}(Q,N'') > \mathbf{P}(Q,N')$ for all natural numbers $N'' > N'$. Thus $\mathbf{P}(Q_{N''}^{NA},N') < \mathbf{P}(Q_{N''}^{NA},N'') = c+2\kappa = \mathbf{P}(Q_{N'}^{NA},N')$, where the inequality was just established and the equalities follow from (52). But $\mathbf{P}(Q_{N''}^{NA},N') < \mathbf{P}(Q_{N'}^{NA},N')$ together with $\partial\mathbf{P}/\partial Q < 0$ by the proof of Proposition 9 implies $Q_{N''}^{NA} > Q_{N'}^{NA}$. *Q.E.D.*

Properties of Figure 4

Let N be a natural number and define

$$Z(N) = \frac{2(N - H_N)}{N+1}. \tag{A33}$$

The text stated three claims about $Z(N)$ based on inspection of Figure 4: $Z(N+1) > Z(N)$ for all $N \geq 1$, $Z(N) > 0$ for all $N \geq 2$, and $\lim_{N\to\infty} Z(N) = 2$. We prove each in turn here. To prove the first claim,

$$\begin{aligned}
Z(N+1) - Z(N) &= \frac{2(N+1-H_{N+1})}{N+2} - \frac{2(N-H_N)}{N+1} \\
&= \frac{2}{(N+1)(N+2)}[1+(N+2)H_N - (N+1)H_{N+1}] \\
&= \frac{2}{(N+1)(N+2)}\left[1+(N+2)H_N - (N+1)\left(H_N + \frac{1}{N+1}\right)\right] \\
&= \frac{2H_N}{(N+1)(N+2)},
\end{aligned}$$

which is positive for all $N \geq 1$, proving $Z(N+1) > Z(N)$.

To prove the second claim, note $Z(1) = 0$. Then by the first claim $Z(N) > 0$ for all $N \geq 2$.

The third claim can be proved using a result due to Euler that $H_N = \ln N + \gamma + o(1)$, where $\gamma \approx 0.577$ is the Euler-Mascheroni constant. Then

$$\lim_{N\to\infty} Z(N) = \lim_{N\to\infty} \frac{2[N - \ln N - \gamma - o(1)]}{N+1} = 2$$

by l'Hôpital's Rule. *Q.E.D.*

Proof of Proposition 11

If the AMC leads the N firms to produce more than $D(c)$, by assumption that capacity in the absence of an AMC is less than $D(c)$, we are done. Therefore, suppose equilibrium capacity with an AMC is less than $D(c)$. In that case, we can invoke the principle of full capacity utilization. The ex ante profit of firm i, equal to the present discounted value of its Shapley value each ex post instant less ex ante capacity cost, is

$$\frac{1}{r}\left\{\sum_{n=1}^{N} w(n,N)\left[\alpha B(Q_i + (n-1)Q_j) - \alpha B((n-1)Q_j) - cQ_i + sQ_i(1-e^{-rT_n})\right] - \kappa Q_i\right\}. \tag{A34}$$

The variable T_n denotes the duration of an AMC when n firms including i operate, given implicitly by the accounting identity

$$F = \frac{s[Q_i + (n-1)Q_j]}{r}\left(1 - e^{-rT_n}\right). \tag{A35}$$

Substituting for T_n from (A35) in (A34) yields

$$\frac{1}{r}\left\{\sum_{n=1}^{N} w(n,N)\left[\alpha B(Q_i + (n-1)Q_j) - \alpha B((n-1)Q_j) - cQ_i + \frac{rFQ_i}{Q_i + (n-1)Q_j}\right] - \kappa Q_i\right\}. \tag{A36}$$

Taking the first-order condition with respect to Q_i, and imposing symmetry $Q_i = Q_j = Q/N$, after rearranging, yields the equilibrium condition for industry capacity conditional on the AMC terms:

$$\sum_{n=1}^{N} w(n,N)\left[P(nQ/N) - c + \frac{rFN(n-1)}{Qn^2}\right] - \kappa = 0. \tag{A37}$$

Multiplying through by 2 and rearranging, (A37) implies

$$\sum_{n=1}^{N} \tilde{w}(n,N)P(nQ/N) = c\sum_{n=1}^{N} \tilde{w}(n,N) + 2\kappa - \frac{2rF}{Q}\sum_{n=1}^{N} \frac{n}{N(N+1)}\left(\frac{n-1}{n^2}\right). \tag{A38}$$

Using equations (49) and (50) and the definition of $\mathbf{P}(Q,N)$, (A38) implies

$$\mathbf{P}(Q,N) = c + 2\kappa + \frac{2rF}{Q(N+1)}\sum_{n=1}^{N}\left(1 - \frac{1}{n}\right),$$

which reduces to equation (53) reported in the text. *Q.E.D.*

Proof of Proposition 12

We first derive equation (56), characterizing equilibrium capacity under the optimal framework AMC. The donor's ex ante profit surplus, equal to the present discounted value of its Shapley value each ex post instant less its ex ante contribution to the fund F, is

$$\frac{1}{r}\sum_{n=1}^{N} v(n,N)\left[\alpha B(nQ_i) - cnQ_i + snQ_i(1-e^{-rT_n})\right] - F. \tag{A39}$$

The donor's marginal contribution to surplus in each coalition equals total surplus because the donor is pivotal. The donor's Shapley-value weight $v(n,N)$ equals the unconditional probability that a coalition including the donor—randomly formed by permuting the grand coalition—happens to have n of the N total firms. This probability equals the ratio of permutations in which n firms appear before and $N-n$ appear after the donor, which is simply $N!$, to the total $(N+1)!$ permutations of the grand coalition. Thus

$$v(n,N) = \frac{N!}{(N+1)!} = \frac{1}{N+1}. \tag{A40}$$

Substituting for T_n from (A35) in (A39) and imposing symmetry $Q_i = Q_j = Q/N$ yields

$$\frac{1}{r(N+1)}\sum_{n=1}^{N}\left[\alpha B(nQ/N) - \frac{cnQ}{N}\right] - \frac{F}{N+1}. \tag{A41}$$

The optimal framework AMC for the donor maximizes (A41) subject to incentive-compatibility constraint (53). This constrained optimization problem can be transformed into an equivalent unconstrained one by solving (53) for F, substituting into (A41), and rearranging to obtain the new objective function

$$\frac{1}{r(N+1)}\sum_{n=1}^{N}\left[\alpha B(nQ/N) - \frac{cnQ}{N}\right] - \frac{Q}{r(N+1)Z(N)}[c+2\kappa-\mathbf{P}(Q,N)], \tag{A42}$$

where $Z(N)$ is defined in (A33). The first-order condition, after rearranging, is

$$\sum_{n=1}^{N}\frac{n}{N}P(nQ/N) - \frac{N+1-H_N}{Z(N)}c - \frac{2\kappa}{Z(N)} + \frac{\mathbf{P}(Q,N)}{Z(N)} + \frac{Q}{Z(N)}\cdot\frac{\partial\mathbf{P}(Q,N)}{\partial Q} = 0, \tag{A43}$$

Rearranging (A43) yields

$$(N+1-H_N)[\mathbf{P}(Q,N)-c] - 2\kappa + \frac{\partial\mathbf{P}(Q,N)}{\partial Q} = 0, \tag{A44}$$

which upon further rearranging yields (56).

We next analyze equilibrium with a supply commitment. The ex ante profit of firm i continues to be given by (A34), but the AMC duration with a supply commitment is now given implicitly by the accounting identity

$$F = \frac{s\hat{q}}{r}\left(1-e^{-rT_n}\right). \tag{A45}$$

Substituting for T_n from (A45) in (A34) yields

$$\frac{1}{r}\left\{\sum_{n=1}^{N} w(n,N)\left[\alpha B(Q_i+(n-1)Q_j) - \alpha B((n-1)Q_j) - cQ_i + \frac{rFQ_i}{\hat{q}}\right] - \kappa Q_i\right\}. \tag{A46}$$

Taking the first-order condition with respect to Q_i, imposing symmetry $Q_i = Q_j = Q/N$, imposing that target optimally equals equilibrium capacity $\hat{q} = Q$, and rearranging, yields the equilibrium condition for industry capacity conditional on the AMC terms:

$$\sum_{n=1}^{N} w(n,N)\left[P(nQ/N) - c + \frac{rF)}{Q}\right] - \kappa = 0. \tag{A47}$$

Multiplying through by 2, noting $\tilde{w}(n,N) = 2w(n,N)$, and rearranging yields

$$\mathbf{P}(Q,N) = c + 2\kappa + \frac{rF}{Q}. \tag{A48}$$

Turning to the donor's problem of designing the optimal supply commitment, its objective function again starts off as (A39). Substituting for T_n from (A45), imposing symmetry $Q_i = Q_j = Q/N$, imposing that target optimally equals equilibrium capacity

$\hat{q} = Q$, and rearranging, yields objective function

$$\frac{1}{r(N+1)} \sum_{n=1}^{N} \left[\alpha B(nQ/N) - \frac{cnQ}{N} \right] - \frac{F}{2}. \tag{A49}$$

The optimal supply commitment for the donor maximizes (A49) subject to incentive-compatibility constraint (A48). This constrained optimization problem can be transformed into an equivalent unconstrained one by solving (A48) for F, substituting into (A49), and rearranging to obtain the new objective function

$$\frac{1}{r(N+1)} \sum_{n=1}^{N} \left[\alpha B(nQ/N) - \frac{cnQ}{N} \right] - \frac{Q}{2r}[c+2\kappa - \mathbf{P}(Q,N)], \tag{A50}$$

Transforming the first-order condition from (A50) into Lerner-index condition (57) follows steps similar to (A43)–(A44) above.

We are now set to compare the efficiency of the two policies by nesting the objective functions (A42) and (A50) as

$$\frac{1}{r(N+1)} \sum_{n=1}^{N} \left[\alpha B(nQ/N) - \frac{cnQ}{N} \right] - \frac{\theta Q}{r}[c+2\kappa - \mathbf{P}(Q,N)], \tag{A51}$$

where $\theta = 1/(N+1)Z(N) = 1/2(N-H_N)$ for a framework AMC and $\theta = 1/2$ for a supply commitment. By the Envelope Theorem the donor's welfare is decreasing in θ. Therefore, the donor strictly prefers a framework AMC if and only if $1/2(N-H_N) < 1/2$, equivalent to $H_N + 1 - N < 0$, and a supply commitment if the reverse inequality holds.

Define $f(N) = H_N + 1 - N$. One can verify $f(2) = 1/2 > 0$ and $f(3) = -1/6 < 0$. Further $f(N+1) - f(N) = H_{N+1} - H_N - 1 = -N/(N+1)$ for all natural numbers N, implying $f(N+1) < f(N)$ implying $f(N) < 0$ for all $N \geq 3$. Hence the donor strictly prefers a supply commitment if $N = 2$ and a framework AMC if $N \geq 3$. *Q.E.D.*

Proof of Proposition 13

It remains to show that an efficient forcing contract with infinitesimal country copayment s_c can deliver the integrated surplus to the donor. We saw from the text that that AMC design induced the firm to invest in R&D. The firm's ex ante profit is $\pi(Q^{IN}, \theta) + \beta F^{DP} - R = 0$ by construction of F^{DP}. The AMC results in efficient investment and output, generating joint surplus $w(Q^{IN}, \theta) - R$. The donor appropriates all but the 0 earned by the firm. *Q.E.D.*

Proof of Proposition 14

Assume $R > \pi(Q^{NA}, \theta)$. Suppose the donor implements an efficient framework AMC with temporary subsidy, which the text showed involves fund size $F^{DP} = [R - \pi(Q^{NA}, \theta)]/\beta$. It remains to derive the condition under which the donor benefits from this policy relative to no AMC. The donor's surplus from this AMC design is

$$\begin{aligned}
& \frac{1-\beta}{r} \left[\alpha B(Q^{NA}) - cQ^{NA} + sQ^{NA}(1 - e^{-rT}) \right] - F^{DP} && \text{(A52)} \\
= \quad & \frac{1-\beta}{r} \left[\alpha B(Q^{NA}) - cQ^{NA} \right] - \beta F^{DP} && \text{(A53)} \\
= \quad & \frac{1-\beta}{r} \left[\alpha B(Q^{NA}) - cQ^{NA} \right] - R + \pi(Q^{NA}, \theta) && \text{(A54)} \\
= \quad & w(Q^{NA}, \theta) - R. && \text{(A55)}
\end{aligned}$$

Equation (A52) is derived by subtracting the donor's initial expenditure F^{DP} on the AMC fund from its share $1-\beta$ of gains from Nash bargaining from equation (11). We have further substituted Q^{NA} for Q following arguments in the text that this design of AMC does not enhance capacity incentives beyond those without an AMC. Equation (A53) follows by substituting from (10), (A54) by substituting for F^{DP}, and (A55) by first substituting for $\pi(Q^{NA}, \theta)$ from (8) and then using the definition of $w(Q, \theta)$ from (3). The donor strictly benefits from this design of AMC if and only if (A55) is positive. *Q.E.D.*

Proof of Proposition 15

The first step is to derive the optimal forcing contracts for close and distant products. Proposition 8 characterizes the optimal contract for a close product. It remains to derive the optimal contract for a distant product. As with a close product, we will consider a direct revelation mechanism in which the firm announces its cost type before investing in capacity or producing. Quantity and fund size are $\underline{Q}$ and $\underline{F}$ if it announces type $\underline{c}$ and $\bar{Q}$ and $\bar{F}$ if it announces $\bar{c}$. These terms are set to maximize donor surplus (A16) subject to truth-telling constraints (A19) and (A20) as well as the individual-rationality constraint

$$\nu \left[\pi(\underline{Q}, \underline{c}) + \beta \underline{F} \right] + (1-\nu) \left[\pi(\bar{Q}, \bar{c}) + \beta \bar{F} \right] \geq \max \left\{ R, \mathrm{E}\left[\pi(Q^{NA}(c), c) \right] \right\}. \tag{A56}$$

Following Section 4, we slightly abuse notation by writing the profit function from (8) as $\pi(Q,c)$, the second argument stressing its dependence on the parameter c of interest rather than the whole vector of model parameters θ. The two terms in the max operator in (A56) ensure the firm makes two participation decisions correctly, the first ensuring the firm invests in R&D and the second ensuring it accepts the AMC rather that proceeding without one. Because the contract is signed at a point of symmetric uncertainty, and the R&D investment is made before uncertainty is resolved, the individual-rationality constraint need only hold in expectation across states.

We will analyze the following candidate solution (two stars distinguishing the optimum for a distant product from that for a close product). Set outputs at their levels in the integrated optimum: $\underline{Q}^{**} = Q^{IN}(\underline{c})$ and $\bar{Q}^{**} = Q^{IN}(\bar{c})$. Set fund sizes so that they solve the system of equations treating constraints (A19) and (A56) as equalities:

$$\bar{F}^{**} = \frac{1}{\beta}\left\{\max\left\{R, \mathrm{E}\left[\pi(Q^{NA}(c),c)\right]\right\} - \nu\pi(Q^{IN}(\bar{c}),\underline{c}) - (1-\nu)\pi(Q^{IN}(\bar{c}),\bar{c})\right\} \tag{A57}$$

$$\underline{F}^{**} = \bar{F}^{**} - \frac{1}{\beta}\left[\pi(Q^{IN}(\underline{c}),\underline{c}) - \pi(Q^{IN}(\bar{c}),\underline{c})\right]. \tag{A58}$$

One can verify that these fund sizes satisfy the remaining constraint (A20):

$$\begin{aligned}
\pi(Q^{IN}(\bar{c}),\bar{c}) + \beta\bar{F}^{**} &= \pi(Q^{IN}(\bar{c}),\bar{c}) - \pi(Q^{IN}(\bar{c}),\underline{c}) + \pi(Q^{IN}(\underline{c}),\underline{c}) + \beta\underline{F}^{**} &\text{(A59)}\\
&= -\frac{\beta}{r}(\bar{c}-\underline{c})Q^{IN}(\bar{c}) + \pi(Q^{IN}(\underline{c}),\underline{c}) + \beta\underline{F}^{**} &\text{(A60)}\\
&= \frac{\beta}{r}(\bar{c}-\underline{c})\left[Q^{IN}(\underline{c}) - Q^{IN}(\bar{c})\right] + \pi(Q^{IN}(\underline{c}),\bar{c}) + \beta\underline{F}^{**} &\text{(A61)}\\
&> \pi(Q^{IN}(\underline{c}),\bar{c}) + \beta\underline{F}^{**}. &\text{(A62)}
\end{aligned}$$

Equation (A59) follows from substituting for $\bar{F}^{**}$ from (A57), (A60) and (A61) using the definition of the profit function (8), and (A62) from $\bar{c} > \underline{c}$ and $Q^{IN}(c)$ decreasing in c. Equations (A59)–(A62) verify (A20) holds.

The proposed solution must solve our proposed constrained optimization problem because it involves outputs maximizing the integrated surplus and forces the individual-rationality constraint to hold with equality. Indeed, the proposed solution must be the optimum over the entire set of contracts—forcing or not, direct revelation or not—satisfying the firm's individual rationality from a position of symmetric uncertainty. The subsequent analysis will not focus on the terms of this optimal contract but just on the donor's equilibrium surplus, which equals the integrated surplus less the individually rational ex ante profit for the firm:

$$\mathrm{E}\left[w(Q^{IN}(c),c)\right] - \max\left\{R, \mathrm{E}\left[\pi(Q^{NA}(c),c)\right]\right\}, \tag{A63}$$

where $\mathrm{E}\left[Q^{IN}(c),c)\right]$ is defined analogously to (58).

With the optimal contracts for close and distant products in hand, we turn to a comparison of donor surpluses from the two products. First suppose $\beta > \hat{\beta}$. With a close product, by Proposition 8, the donor obtains $w(Q^{IN}(c),c) - \pi(Q^{NA}(c),c)$ in each state c, for an expected surplus of $\mathrm{E}\left[w(Q^{IN}(c),c)\right] - \mathrm{E}\left[\pi(Q^{NA}(c),c)\right]$, which equals (A63) if (59) is negative and exceeds (A63) if (59) is positive.

Next suppose $\beta < \hat{\beta}$. With a close product, the donor's expected surplus is as computed in the previous paragraph less the distortion term (39), denoted ξ for brevity:

$$\mathrm{E}\left[w(Q^{IN}(c),c)\right] - \mathrm{E}\left[\pi(Q^{NA}(c),c)\right] - \xi. \tag{A64}$$

The difference between (A63) and (A64) equals

$$\xi - \max\left\{0, R - \mathrm{E}\left[\pi(Q^{NA}(c),c)\right]\right\} = \min\left\{\xi, \xi - \left\{R - \mathrm{E}\left[\pi(Q^{NA}(c),c)\right]\right\}\right\}. \tag{A65}$$

Because $\xi > 0$ when $\beta < \hat{\beta}$ by Proposition 8, (A65) is positive if (39) exceeds (59) and is negative if (59) exceeds (39). *Q.E.D.*

Appendix B: Design Refinements for Technologically Close Products

This appendix analyzes a series of refinements to the design of an AMC for a technologically close product. We examine how the performance of the basic AMC studied in Section 3 changes when various features are added such as price caps, purchase guarantees, and country copayments. We also examine the effect of adding an outside party that collects the accrued interest in the AMC fund or acts as the procurement agent in bargaining.

B.1. Price Cap

In the original proposals for AMCs (Kremer and Glennerster 2004; Kremer, Barder, and Levine 2005) price caps played a prominent role in AMC design. The idea was to use the AMC subsidy for a dual purpose. In addition to stimulating ex ante investment, the extra surplus provided during the AMC period could be exchanged for lower prices during the tail period, avoiding static deadweight loss then. To streamline the analysis, we have so far ignored price caps but return to them now.

Price caps will turn out to be quite beneficial in our model, but for subtly different reasons than in the original proposals. There is no problem of static deadweight loss in our model because trade is characterized by Nash bargaining, which is ex post efficient. Instead, price caps will help solve hold-up problem. In the absence of a price cap, the firm has an incentive to cut back capacity to put the donor at a bargaining disadvantage by increasing its average value for the product, raising the price the firm can extract from the donor. By limiting this price increase, a price cap reduces the ex ante underinvestment distortion.

Indeed, price caps can be be so powerful that an AMC with a suitably designed price cap can completely solve the hold-up problem, achieving the integrated outcome. The efficient policy specifies that the donor pays the firm a lump sum F^{FC} ex ante in exchange for an agreement that the per-unit price over the whole ex post period be no higher than $c+\kappa$. It can be shown that the price cap binds, leading to equilibrium capacity and output $Q^{IN} = q^{IN} = D(c+\kappa)$. Lump-sum payment F^{FC} was fine-tuned in (37) to provide enough surplus to make the firm indifferent between accepting the contract and rejecting it and moving to the subgame without an AMC.

A subtle commitment problem with a price cap needs to be addressed. The firm can threaten not to supply at that price to induce the donor to renegotiate, deleting the price-cap provision. The donor would like to issue a counterthreat not to renegotiate, but because trade requires both parties' assent, this counterthreat would merely enforce the firm's initial threat. To circumvent the commitment problem, the requirement that both parties must assent to trade needs to be altered. The contract could specify a supply guarantee, that the firm must agree to supply a certain amount each instant in exchange for F^{FC}. The efficient level for supply guarantee is the integrated capacity/output Q^{IN}.

The design of an AMC with price caps is somewhat delicate. The policy just outlined puts the firm on the knife edge between accepting and not, which leaves the possibility that the AMC is rejected if one of the terms is set incorrectly. Raising the lump-sum payment slightly above F^{FC} or the price cap slightly above $c+\kappa$ can provide the needed cushion.

B.2. Purchase Guarantee

One modification to the framework AMC that was considered by the Economics Expert Group but ultimately not included in the design of the pneumococcus pilot was a purchase guarantee, i.e., a right given to the firm to be able to sell as much as it wants up to some limit, $\hat{q}$. This subsection is devoted to an analysis of how this modification would affect investment incentives under a framework AMC.

Return to the model in which the donor and buyer roles are combined in a single player called the donor and assume interest accrues to the escrow. A purchase guarantee can be captured in the model by a change in Nash bargaining threat points. In the absence of a purchase guarantee, both bargaining parties must assent to trade, leading to threat points of 0 for both. A purchase guarantee allows the firm the option to force trade. If $s < c$, this option is worthless because, if bargaining breaks down, the donor makes no payment above and beyond s, and s does not even cover c. If $s > c$, this option is valuable and so would be exercised.

To rule out a proliferation of subcases, before turning to an analysis of the case in which $s > c$, we will posit a number of additional conditions. Posit, first, that $Q \le \hat{q}$, i.e., output q is constrained by Q rather than $\hat{q}$. If $Q > \hat{q}$, one can show that the AMC does not provide investment incentives because the purchase guarantee applies to inframarginal units. Posit, second, that F is sufficiently small that the subsidy period is temporary. Then T is determined by (10). We will explore the case of a permanent subsidy below. Invoking the principle of full capacity utilization, the firm's threat point is

$$\int_0^T (s-c)Qe^{-rt}dt = \frac{(s-c)Q}{r}(1-e^{-rT}) \tag{B1}$$

and the donor's is

$$\int_0^T \alpha B(Q)e^{-rt}dt = \frac{\alpha B(Q)}{r}(1-e^{-rT}). \tag{B2}$$

Paradoxically, the firm's positive threat point may put it in a worse bargaining position because the donor's positive threat point may be even better. However, the firm cannot credibly threaten *not* to exercise its option in a subgame perfect equilibrium. So what is a valuable option in isolation may be damaging in a strategic situation.

Indeed, we can show that incentives can be impaired by a framework AMC with purchase guarantee compared to no AMC. Subtracting the sum threat-point surpluses from the present discounted value of joint surplus over the ex post continuation game given in (11) after rearranging yields gains from trade

$$\frac{e^{-rT}}{r}[\alpha B(Q) - cQ]. \tag{B3}$$

Adding the firm's β share of these gains from trade to the firm's threat point gives its Nash bargaining surplus. Subtracting capacity costs yields the following expression for ex ante profits,

$$\frac{\beta e^{-rT}}{r}[\alpha B(Q) - cQ] + \frac{(s-c)Q}{r}(1 - e^{-rT}) - KQ = \frac{1}{r}\left\{\beta\left(1 - \frac{rF}{sQ}\right)[\alpha B(Q) - cQ] + \frac{s-c}{s}rF - \kappa Q\right\}, \tag{B4}$$

where the right-hand side follows from substituting for T from (10). Taking the first-order condition for Q and rearranging gives an expression for equilibrium capacity conditional on the AMC terms:

$$Q = D\left(c + \left(\frac{sQ}{sQ - rF}\right)\frac{\kappa}{\beta} + \left[\frac{\alpha B(Q)}{Q} - c\right]\frac{rF}{sQ}\right). \tag{B5}$$

It is easy to see that the argument of D in (B5) is strictly greater than $c + \kappa/\beta$. Thus if $D(c + \kappa/\beta) > 0$, then $D'(p) < 0$ for all $p \in (c + \kappa/\beta, \bar{p})$, implying (B5) is strictly less than $D(c + \kappa/\beta) = Q^{NA}$. Hence for any F and s satisfying the posited conditions, we have the surprising result that incentives can be impaired by such an AMC. Incentives are destroyed during the AMC period because the only surplus is the present discounted value F provided by the AMC subsidy, which is effectively lump sum. Nash bargaining during the tail period does provide incentives, but the same each instant as provided by Nash bargaining in the absence of an AMC, and the latter begins instantly ex ante, not after a delay of T as with the AMC with purchase guarantee.

Adding a purchase guarantee can enhance incentives if layered on a framework AMC with a perpetual subsidy. Indeed, it is easy to construct a policy of this form attaining the integrated benchmark. Set the subsidy as $s = c + \kappa$ and the limit to the guarantee as $\hat{q} = Q^{IN}$. Take the fund size to be large enough that the subsidy is perpetual: $F \geq (c + \kappa)Q^{IN}/r$ (this is an inequality because a larger fund just means more interest flowing back to the donor, so the funds are not wasted). In equilibrium, there is no Nash bargaining because the firm cannot commit not to exercise its option to produce. To construct the equilibrium, we assume the firm, indifferent as to which capacity it installs in $[0, Q^{IN}]$, installs the maximum in this range. The equilibrium is robust in that the donor could always break the firm's indifference by setting s to be slightly greater than $c + \kappa$.

The purchase guarantee layered on a perpetual subsidy functions much like the efficient price cap from the previous subsection. The firm is offered a high enough s that its cost of capacity and production for the efficient output are covered but no more. The firm does not gain a bargaining advantage from restricting output; there is no need for the donor to bargain with the firm because the firm cannot commit not to chase the subsidy by supplying up to capacity. The subsidy effectively becomes the fixed price in the ex post market.

Overall, the results from this subsection are somewhat chaotic. A price guarantee layered on an AMC with a temporary subsidy could be worse than no AMC at all. On the other hand, a price guarantee layered on an AMC with a permanent subsidy can achieve the integrated outcome. Price guarantees are thus potentially powerful tools that should be used with caution.

B.3. Interest Accrual

The baseline specification behind the policies analyzed so far was that interest accrues in the escrow and, like the principal in the escrow, cannot be used for other purposes. Here we explore two alternatives: interest flows back to the donor or flows to some third party. We will see that allowing interest to flow back to the donor completely undermines the investment incentives provided by a framework AMC. In contrast, having the interest flow to a third party helps with incentives. The bargaining parties prefer to get the money out of the fund sooner rather than later, when discounting has destroyed much of its value. The way parties extract money from the fund more quickly is to increase quantity traded, which requires the firm to have installed greater capacity, hence the source of increased incentives. The donor can do better than just giving the interest away to the third party; it can capture the value by charging the third party a lump sum ex ante for the right to collect interest on the fund ex post. In the limit, the donor's surplus approaches that from a framework AMC with interest accruing to the escrow involving a perpetual subsidy.

Because interest does not accrue in the escrow in either alternative, there is no way with a finite F to endow a perpetual subsidy. Thus both alternatives for interest accrual must be associated with a temporary subsidy. Both alternatives involve a change to the accounting identity determining the length T of the AMC period, before given by equation (10), now given by

$$F = sqT. \tag{B6}$$

The identity says that the nominal amount paid into the escrow on the left-hand side equals the nominal amount paid out on the right-hand side.

To proceed, we focus first on the alternative in which interest on the AMC fund flows back to the donor rather than accruing in the escrow. The results from Proposition 3 lead one to expect this AMC design to also be useless for incentives, and we will see this is the case. Considerable new analysis is required because the subtle change in interest accrual causes a fundamental change in the economic environment, in particular to how Nash bargaining works. Rather than zero threat points, the donor's is now positive because vetoing trade leaves all the money in the escrow from which it earns interest. Trade has the opportunity cost of drawing down the fund, reducing the interest earned by the donor, which must be accounted for. We undertake the requisite analysis in the appendix, where we provide the proof of the following proposition.

Proposition 16. *Suppose that interest from the AMC fund flows back to the donor rather than staying in the escrow. For all terms $s > 0$ and $F > 0$, the framework AMC adds no incentives for a monopoly firm to invest in capacity.*

Proof. Consider a framework AMC letting interest accrue to the donor rather than the escrow. The nature of Nash bargaining differs somewhat from in Section 3.4, so some new analysis is needed to compute surplus allocations from Nash bargaining. We proceed by computing threat points and gains from trade.

To compute threat points, note that if bargaining breaks down, the firm obtains no continuation surplus. The donor obtains interest rF each instant for a present discounted value of $\int_0^\infty rFe^{-rt}dt = F$. Thus the donor's threat point is F.

To compute gains from trade, start by computing the present discounted value of joint continuation surplus from trade is

$$\int_0^T [\alpha B(Q) - cQ + sQ]e^{-rt}dt + \int_T^\infty [\alpha B(Q) - cQ]e^{-rt}dt + I. \tag{B7}$$

Capacity Q has been substituted for output on the basis of the principle of full capacity utilization. The term I is the present discounted value of the interest earned by the donor. This is determined by accounting identity (B11). Solving (B11) for I, substituting the resulting expression for I in (B7), and subtracting off the sum F of parties' threat point surpluses yields gains from trade

$$\frac{1}{r}[\alpha B(Q) - cQ],$$

the same as in equation (5) for the case without an AMC. The rest of the derivation of equilibrium is thus the same here as without an AMC. *Q.E.D.*

While the result is the same for framework AMCs of finite duration whether interest accrues in the fund or flows back to the donor—both designs useless for incentives—the result holds for different reasons. When interest accrues to the fund, the AMC does contribute to ex post gains from trade. The reason the gains do not contribute to investment incentives is that they do not vary with capacity. When interest flows back to the donor, on the other hand, the AMC contributes nothing to gains from trade. The opportunity cost of \$1 extra subsidy to the firm is \$1 of interest that would have flowed back to the donor. Thus an AMC that lets interest flow out of the fund to the donor does not contribute to incentives for quite stark reasons.

Next consider the alternative in which interest flows to a third party. For concreteness, we will call the third party a bank and suppose it is the same party that sets up and holds the escrow. A crucial assumption is that the bank is a non-strategic player. If the bank were strategic and in particular were able to participate in bargaining ex post, the outcome could collapse to the one just analyzed with interest flowing back to the donor. An added virtue of the analysis of this variant is that it will provide a building block for the next subsection, where we extend the model to the case in which the AMC is set up by a donor but then administered by a different procurement agency, which undertakes ex post bargaining. The pilot AMC for pneumococcus had this sort of principal-agent structure, with the Gates Foundation and donor countries providing funds and designing the AMC but with GAVI carrying out the procurement. We will be able to port much of the analysis from this subsection directly to the next.

Nash bargaining when interest accrues to a third party is similar in nature to that when interest accrues to the escrow in Section 3.4. In both cases, parties' threat points are 0, and the donor does not directly gain from leaving money in the fund. Thus we can borrow much of the analysis from Section 3.4 here, in particular the analysis up through equation (12). The analysis diverges when we consider the expression to substitute for the AMC period length T in (12). Using (B6) rather than (10) to substitute for T, after rearranging, yields the following expression for the firm's ex ante profits:

$$\frac{\beta}{r}\left[\alpha B(Q) - \left(c + \frac{\kappa}{\beta}\right)Q + sQ(1 - e^{-rF/sQ})\right] = \pi(Q,\theta) + \frac{\beta sQ}{r}(1 - e^{-rF/sQ}). \tag{B8}$$

The first-order condition with respect to Q is

$$P(Q) - \left(c + \frac{\kappa}{\beta}\right) + s - s\left(1 + \frac{rF}{sQ}\right)e^{-rF/sQ} = 0. \tag{B9}$$

Equation (B9) can be readily compared to the first-order conditions derived in earlier settings. The first three terms, $P(Q) - c - \kappa/\beta$, constitute the first-order condition from the benchmark with no AMC. The last two terms reflect the incentive effects provided by the AMC subsidy. Of these two terms, the first, simply s, is the same incentive effect provided by a perpetual subsidy. The last term reflects the fact that with a fixed fund, an increase in production shrinks the AMC period, dampening the incentives provided

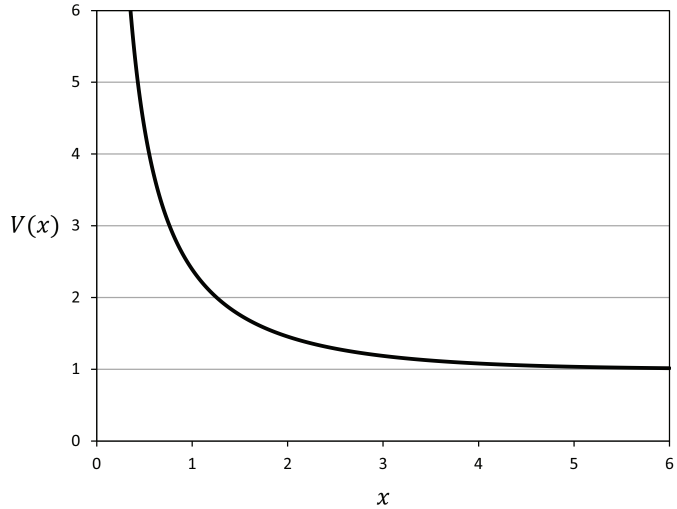

Figure 6: Ratio of marginal cost to marginal benefit of lengthening AMC

by this AMC design. Because interest does not accrue, the current subsidy gain is not completely offset by the future subsidy loss, so the incentive effect does not completely wash out as it did in equation (13).

Folding the game back to the donor's design of the optimal AMC, the present discounted value of donor surplus equals its share $1-\beta$ of joint continuation surplus (11), less its initial contribution F to the escrow, plus the bank's lump-sum payment I from the bank for the right to collect interest ex post:

$$\frac{1-\beta}{r}[\alpha B(Q)-cQ+sQ(1-e^{-rT})]-F+I=\frac{1-\beta}{r}[\alpha B(Q)-cQ]-\frac{\beta}{r}sQ(1-e^{-rF/sQ}). \tag{B10}$$

The right-hand side requires the derivation of an expression to substitute for I. Assuming banks bid competitively for the right to set up the escrow and collect interest, and these activities are carried out at zero cost, I will equal the present discounted value of payments out of the AMC fund, which can be derived from the accounting identity

$$F=\int_0^T sqe^{-rt}dt+I. \tag{B11}$$

The left-hand side is the present discounted value of payments into the fund, in this case simply the lump-sum endowment. The right-hand side is the present discounted value of all payments out of the fund, both for interest and the subsidy. Substituting for I from (B11) and for T from (B6) into the left-hand side of (B10), after rearranging, gives the right-hand side.

The optimal AMC for the donor is given by the values of F and s maximizing (B10) subject to (B9). This constrained optimization problem can be simplified in several steps. First note that F appears with the same group of terms in (B9) and (B10). We thus substitute the change of variables $x=rF/sQ$ and maximize with respect to x rather than F. Because we can also write $x=rT$, x can be interpreted as a rescaling of the length of the AMC period. After making the change of variables, we can solve (B9) for s and substitute out for s in (B10), yielding

$$\frac{1-\beta}{r}[\alpha B(Q)-cQ]-\frac{\beta}{r}Q\left[c+\frac{\kappa}{\beta}-P(Q)\right]V(x), \tag{B12}$$

where for brevity we have introduced the function $V(x)$, graphed in Figure 6, defined as

$$V(x)=\frac{1-e^{-x}}{1-(1+x)e^{-x}}. \tag{B13}$$

$V(x)$ can be interpreted as the ratio of marginal cost to marginal benefit (in terms of increased investment incentives) of lengthening AMC holding other variables constant.

We have thus transformed the donor's original constrained optimization problem into the unconstrained problem of choosing Q and x to maximize (B12). To complete the solution for the donor's optimum, as the figure shows, and the proof of the next proposition makes rigorous, $V(x)$ is minimized in the limit as $x\to\infty$, approaching a value of 1. Thus the optimal x is as large as possible, implying given the change of variables that the optimal AMC lasts as long as possible. Substituting the limit of 1 for $V(x)$ in (B12) and taking the derivative with respect to the remaining choice variable, Q, yields the same first-order condition as (20).

This establishes the following proposition.

Proposition 17. *The donor's surplus from a framework AMC having interest accruing to a third party can come arbitrarily close to that from the optimal framework AMC with interest accrual and a perpetual subsidy, approaching the outcome described in Proposition 4.*

Proof. This proof fills in the remaining detail from the analysis leading up to the proposition of verifying the properties of $V(x)$ seen in Figure 6. In particular, we are left to verify that $V(x)$ approaches its infimum over the set of $x > 0$ in the limit as $x \to \infty$. We will do this by showing that $V'(x) < 0$ for all $x > 0$ and that $\lim_{x\to\infty} V(x) = 1$. We have

$$V'(x) = \frac{Y(x)e^{-x}}{[1-(1+x)e^{-x}]^2}.$$

where $Y(x) = 1-x-e^{-x}$. The sign of $V'(x)$ is determined by the sign of $Y(x)$. We have $Y'(x) = e^{-x}-1$, which is negative for all $x > 0$. Given $Y(0) = 0$, $Y'(x) < 0$ for all $x > 0$ implies $Y(x) < 0$ for all $x < 0$. Thus $V'(x) < 0$ for all $x < 0$.

To find the limit, rewrite $V(x)$ as

$$V(x) = \frac{1}{1-xe^{-x}/(1-e^{-x})}.$$

Then

$$\lim_{x\to\infty} V(x) = \frac{1}{1-\lim_{x\to\infty} x/e^x} = \frac{1}{1-\lim_{x\to\infty} 1/e^x} = 1,$$

where the second equality uses l'Hôpital's Rule. *Q.E.D.*

Intuitively, the AMC without interest accrual achieves efficiency by having an increasingly long duration. This helps avoid the perverse incentives associated with AMCs of finite duration, whereby an increase in the firm's capacity and output shortens the AMC period. An increasingly long duration pushes these disincentive effects off into the far future in which they are heavily discounted and thus matter less for ex ante investment. The fund F needed to endow an arbitrarily long AMC becomes arbitrarily large. The donor is willing to contribute an arbitrarily large F because the fee I it earns from the the bank for the right to collect interest ex post also becomes arbitrarily large.

Proposition 17 focuses on AMC performance in the limit. Results for AMCs short of the limit—for given finite values of x and thus F—deserve highlighting. While the proposition says that a framework AMC with interest accruing to a third party can approach the performance of one with interest accruing to the escrow, careful dissection of the analysis shows that the former does strictly worse than the latter for any finite F. Thus, *at best*, third-party interest can only approach a tie with escrow interest in the limit. Propositions 3 and 16 give the impression that a framework AMC cannot add to incentives if it is of finite duration. However, this impression is false for AMCs with interest accruing to a third party, according to Proposition 17. The proposition says that concrete AMCs of that design can be constructed providing donor surplus arbitrarily close to that from an AMC with a perpetual subsidy, which strictly improve incentives by Proposition 4.

So far we have focused on the effect of interest accrual on the efficiency of framwork AMCs. For completeness, it is worth expanding the discussion to other policies including supply commitments and forcing contracts. It is immediate from existing results that any loss from having interest accrue to the donor or third party rather than the escrow can be eliminated in the limit with a suitably designed supply commitment or forcing contract. For both of those policies, we found that the optimal subsidy rate s was not pinned down. A whole range of s could equivalently be used to implement the optimum, including an arbitrarily high s, which in effect results in the whole fund F being paid out almost in the first instant ex post. Such a policy would render interest payments irrelevant because there would be no time for interest to accrue. Thus it is immediate that any inefficiencies from alternative interest accrual assumptions vanish in the limit for supply commitments and forcing contracts.

B.4. Procurement Agent

In this subsection, we modify the model to represent the realities of the pilot for the pneumococcus vaccine more closely. The donors, the Gates Foundation and five countries, who contributed to the design and funding of the AMC, then stepped aside, passing off the procurement to GAVI. To this point in the analysis, we have modeled those functions as being combined in a player called the donor, but in this subsection we will model them as separate players. The analysis in this section will provide a clearer theoretical understanding of the incentive effects of the pilot's design and whether, for future AMCs, introducing a procurement agent is a design feature that should be copied or avoided if possible.

As before, a donor participates in funding and designing the AMC. It receives flow utility $\alpha_d B(q) - X_d$ each instant, where $\alpha_d \in (0,1)$ represents the degree of donor altruism and X_d is its expenditure on vaccines. Introduce a new party, a buyer, who takes on the role of the procurement agent, engaging in Nash bargaining with the firm ex post. Its flow utility is $\alpha_b B(q) - X_b$, where $\alpha_b \in (0,1)$ represents the degree of buyer altruism and X_b is its expenditure on vaccines. Allowing α_d to differ from α_b allows the parties' altruism to differ. Redefine α as the sum $\alpha = \alpha_d + \alpha_b$, and let $P(Q) = \alpha B'(Q)$, i.e., $P(Q)$ is inverse demand for the two altruistic parties together. Suppose, as was the case in the pilot, that the buyer participates in the design of the AMC along with the donor. Thus the AMC will be set to maximize their ex ante joint surplus. As was the case in the pilot, suppose interest from the AMC fund accrues to the donor. We will focus on a framework AMC design.

As before, we solve for equilibrium using backward induction. Nash bargaining between the firm and buyer is isomorphic to Nash bargaining between the firm and donor in the variant from the previous subsection in which interest accrued to a third party. In both cases, interest accrues to a party external to bargaining, and so interest does not factor into the bargaining process. Thus, we can borrow much of the analysis from the previous subsection. Because interest does not accrue to the fund, a finite F cannot fund a perpetual subsidy. AMC duration T must therefore be finite. The relevant accounting identity determining T is again given by equation (B6). One can derive parties' Nash bargaining surpluses then fold the game back to the ex ante period, subtracting off the firm's investment cost, to obtain the following expression for the firm's ex ante profits:

$$\frac{\beta}{r}\left[\alpha_b B(Q)-\left(c+\frac{\kappa}{\beta}\right)Q+sQ(1-e^{-rF/sQ})\right]. \tag{B14}$$

This is identical to (B8) except α_b, the altruism parameter of the buyer who is here the bargaining party, appears in place of α. The first-order condition with respect to Q is

$$\alpha_b B'(Q)-c-\frac{\kappa}{\beta}+s-s\left(1+\frac{rF}{sQ}\right)e^{-rF/sQ}=0. \tag{B15}$$

Folding the game back to the design of the optimal framework AMC, in which both donor and buyer participate. The ex ante present discounted value of the buyer's surplus is its $1-\beta$ share of the Nash bargaining surplus

$$\frac{1-\beta}{r}\left[\alpha_b B(Q)-\left(c+\frac{\kappa}{\beta}\right)Q+sQ(1-e^{-rF/sQ})\right]. \tag{B16}$$

The ex ante present discounted value of the donor's surplus is

$$\frac{\alpha_d B(Q)}{r}-F+I=\frac{1}{r}\left[\alpha_d B(Q)-sQ(1-e^{-rF/sQ})\right]. \tag{B17}$$

As an outsider to the Nash bargain, the donor receives flow surplus $\alpha_d B(Q)$ each instant in effect as an externality. The donor endows the fund with F but receives the present discounted value of accrued interest I, which can be derived using accounting identity (B11). Adding (B16) and (B17), substituting for I from (B11), and rearranging yields the following expression for the ex ante present discounted value of the designers' combined surplus:

$$\frac{1}{r}\left\{[(1-\beta)\alpha_b+\alpha_d]B(Q)-(1-\beta)cQ-\beta sQ(1-e^{-rF/sQ})\right\}. \tag{B18}$$

We can solve for the optimal AMC exactly as in the previous subsection, introducing change of variables $x=rF/sQ$ into (B15) and (B18), solving (B15) for s, substituting for s in (B18), factoring out $V(x)$, taking the limit $x\to\infty$, which is again optimal here, leading to the limit $V(x)\to 1$. After those manipulations, we can take the first-order condition of the resulting expression with respect to Q, providing the following condition for the equilibrium capacity in an optimum:

$$P(Q^{DB})+\beta\alpha_b Q^{DB}B''(Q^{DB})-c-\kappa=0, \tag{B19}$$

where the superscript DB designating this case refers to the fact that the donor and buyer are separate parties. Equation (B19) can be expressed in a Lerner-index form facilitating comparison with previous results:

$$L^{DB}=\frac{P(Q^{DB})-(c+\kappa)}{P(Q^{DB})}=\left(\frac{\alpha_b}{\alpha}\right)\frac{\beta}{|\eta^{DB}|}, \tag{B20}$$

where $\eta^{DB}=Q^{DB}P'(Q^{DB})/P(Q^{DB})$ is the elasticity of the altruists' combined vaccine demand at the equilibrium quantity.

This analysis suggests that the agency structure used in the pneumococcus pilot, with a separate donor and buyer, contributes to incentives in several ways. First, it has interest accrue to a party not involved in bargaining. This allows a finite-duration AMC to provide incentives just as having interest accrue to a third party did in the previous subsection. But separating donor and buyer goes much further, allowing for stronger incentives than a perpetual AMC did when donor and buyer functions were combined in the donor. Compare the Lerner index formula from that case, (22), to that in (B20). The rightmost-hand of both involve the factor $\beta/|\eta|$, but (B20) is multipled by the fraction $\alpha_b/\alpha<1$. Hence, (B20) is closer to the integrated optimum, 0, than (22). For example, suppose the altruistic parties are symmetric, i.e., $\alpha_d=\alpha_b$, implying $\alpha_d/\alpha=1/2$. Then the Lerner index in (B20) would be halfway between (22) and the integrated optimum.

The intuition for this improvement in incentives is that the hold-up problem turns out to be less severe when one altruistic party carries on the bargaining for many involved in the design of the AMC. There are different lenses through which to view the hold-up problem, but one way to view it is that by underinvesting in capacity, the firm raises its counterparty's value per unit, thus increasing the surplus per unit it can extract from the counterparty. Having multiple counterparties would only multiply this underinvestment distortion. If additional altruistic parties participate ex ante instead of ex post, they can contribute to the committed subsidy, multiplying incentives, without multiplying the underinvestment distortion.

Using this intuition, one can better understand how the altruistic parties would structure the procurement-agency relationship if this were endogenous. Suppose there were two altruistic parties, either or whom could feasibly carry out ex post negotiations, with altruism parameters $\alpha_1 < \alpha_2$. Which would be better to designate as the procurement agent? The formula (B20) indicates that the less altruistic one would be better. Lowering the surplus at stake in ex post negotiations reduces the firm's gains from strategically underinvesting relative to a given AMC subsidy, thus reducing the severity of the hold-up problem. In the limit, having a completely non-altruistic procurement agent would lead to the integrated outcome.

B.5. Country Copayment

The buyer side of the pilot AMC for pneumococcus was yet more complicated in practice than modeled in the previous section. We modeled the buyer side as being divided into two players, a donor which set up the AMC (in practice, the Gates Foundation and country finance ministries) and a procurement agent which carried out the ex post negotiation (in practice, GAVI). The pilot also involved a third player, the country receiving the vaccine, which was required to contribute a copayment for every dose purchased. Adding a third player to the model on the buyer side—a fourth player in total considering the firm on the supplier side—multiplies the modeling alternatives and clouds the determination of the most natural alternative. We start with a simple alternative that will allow us to highlight a some basic issues. The subsection then moves to a more complicated model raising some additional issues. Section 6 highlights an additional first-order effect of country copayments that arises with technologically distant product rather than the technologically close product studied in this section.

Consider a model in which the donor, with altruism parameter α_d, sets up the AMC. We will consider a framework AMC with temporary subsidy paid at rate s_d, where the subscript will allow us to distinguish payments coming from the donor-contributed funds from others. We introduce the country receiving the vaccine as a player having the familiar form of utility function used for other players on the buyer side: $\alpha_c B(q) - X_c$, where X_c is its expenditure on vaccines and α_c is its weight on its own health benefit, inversely related to its marginal utility of income. Parameter α_c is free in the model but consistent with the donor making purchases on behalf of a very poor country we think of α_c as being close to 0 and thus the country's marginal utility of income being quite high. The AMC requires the country to make a copayment, denoted s_c. The country can also engage in Nash bargaining with the firm ex post. To keep the bargaining game simple we suppose that the donor sets up the AMC, which is designed to maximize the joint surplus of donor and country, but does not bargain ex post, leaving the country as the only active party from the buyer side ex post.

By construction, the model is similar to that from the previous subsection with the country now filling the role of the buyer. Besides the notational difference that the country's altruism parameter is α_c instead of the buyer's α_b, the only other difference is that the country makes payment s_c for each unit in addition to any payments resulting from Nash bargaining. However, the s_c is just an ex post transfer between the bargaining parties. Unlike the subsidy s_d coming from the AMC fund, which is sunk ex ante, there is nothing sunk about s_c. It ends up having no bearing on the outcome of Nash bargaining. If the equilibrium transfer from country to firm ends up being higher than s_c, the country has to top the payment up; if lower, the firm ends up forgiving some of the required payment. The firm's ex ante profit is identical to equation (B14) after substituting s_d for s and α_c for α_b. The analysis is then identical to that in the previous subsection, leading to the same Lerner index as (B20), with α_c substituted for α_b.

Taking α_c to be close to 0 reveals a benefit of structuring the AMC so that the country make the ex post payments. The Lerner index in (B20) would then be close to 0 and the equilibrium capacity close to the integrated outcome. This is an extension of our earlier finding that having the less-altruistic agent bargain reduces the severity of the hold-up problem. While odd to think of the country as being less altruistic toward itself than an outside party, it is natural to think of a poor country as having a high marginal utility of income. The firm's incentive to distort ex ante capacity investment is reduced when bargaining against a party with a high marginal utility of income. Overall we conclude that a framework AMC with country copayment can enhance efficiency relative to no AMC and for plausible parameters can be quite efficient.

Further efficiencies could be realized by capping the country copayment and tying it to a supply guarantee from the firm. This allows the country copayment to function exactly like a price cap from Section B.1. We saw there that a well-designed price cap can achieve the integrated outcome. The same can be true here: an appropriately specified copayment can not just approach in the limit $\alpha_c \to 0$ but attain the integrated outcome for all $\alpha_c > 0$. We saw that specifying the appropriate price cap is a delicate exercise. Specifying the appropriate copayment may be even more delicate because now the country as well as the firm may be on the knife edge between accepting and not. If the copayment is set too high, the country may decline to participate, but reducing it would lead the firm to reject. Lowering the copayment and increasing the size of the AMC fund to compensate the firm could provide the needed cushion.

The caveat that the whole AMC might fail if the copayment is set too high already serves to highlight possible drawbacks with country copayments without the need for complex additional analysis. An additional drawback arises in the more complex model we turn to next. In this model, the donor participates in ex post bargaining in addition to setting up the AMC. Introducing a country copayment entangles the country in ex post bargaining as well, possibly exacerbating the hold-up problem because by withholding the copayment it can veto the firm's receipt of the AMC subsidy.

To this end, consider an alternative model in which the donor designs the AMC ex ante, specifying a fund F it contributes to, a subsidy rate s_d from that fund, and a copayment s_c from the country, which can be toggled on or off. In the design in which the copayment is toggled on, assume that the country's making the copayment is a precondition for release of the AMC subsidy. we assume The analysis is simplest if the AMC involves a perpetual subsidy, so we will consider a framework AMC with a perpetual subsidy with interest accruing to the escrow. The donor, firm, and country engage in ex post bargaining characterized by Shapley value. The weights that the donor and country place on health benefit $B(q)$ in their utility functions are α_d and α_c respectively.

One complexity that arises with Shapley bargaining is that the principle of full capacity utilization may not hold for out-of-equilibrium coalitions with a subset of demanders. Whether or not the principle holds depends on the relative values of α_d, α_c, c, and κ, leading to a profusion of subcases. To reduce the profusion, we will treat the case in which α_c is close to 0, implying that α_d is close to the sum $\alpha = \alpha_d + \alpha_c$. This will ensure that the principle of full capacity utilization applies to any coalition including the donor and firm. To allow for the possibility that the principle does not apply to the coalition including the country and the firm, we introduce some notation. The present discounted value of that coalition's joint surplus ex post is $[\alpha_c B(q) - cq]/r$. Ignoring capacity constraints, the output maximizing that surplus satisfies first-order condition $\alpha_c B'(q) - c = 0$, implying $q = (B')^{-1}(c/\alpha_c)$. Let $\tilde{Q}$ denote the capacity maximzing that surplus subject to a capacity constraint, i.e., $\tilde{Q} = \max[Q, (B')^{-1}(c/\alpha_c)]$.

An input into the calculation of players' Shapley values, Tables B1 and B2 list player's marginal contributions to the coalition including them and all preceding players in the six possible permutations. A player's Shapley value is the average of its entries down a column. Table B2 lists them in the case in which the country is not charged a copayment and Table B2 when it is. Notice that the country copayment does not appear in any entry in the latter table. In coalitions without the country, the copayment is not made. In coalitions with the country, the copayment is just a transfer so is not a marginal contribution to joint surplus. Thus the copayment "washes out" of both tables.

The only differences between the two tables occur in the third line for permutation DFC and the fourth line for permutation FDC. In each line, a term $s_d Q$ appearing in another players' marginal contribution in Table B1, boxed for easy reference, has been reallocated to the country in Table B1. Coalitions excluding the country lose the AMC subsidy because the country's copayment is a precondition for s_d to be paid per unit. This makes the country pivotal for the realization of the $s_d Q$ term, shifting it into the country's column.

Further analysis is not needed to determine the effect of the copayment. Suppose the AMC is designed by the donor to maximize its ex ante surplus alone. Then the copayment reduces the efficiency of the AMC. The donor loses surplus to the country in the FDC permutation. The firm's loss of surplus in the DFC permutation dulls its marginal investment incentives. Both effects worsen the outcome for the donor. If the AMC is designed by the donor to maximize the combined surplus of donor and country, the copayment policy may be irrelevant. The donor and country offset any ex post transfers between themselves with ex ante transfers. The loss of investment incentives can be offset by increasing s_d in the optimal AMC design.

Table B1: Shapley value when country is not charged a copayment

	Marginal contribution to surplus		
Permutation	Firm, F	Donor, D	Country, C
CDF	$\alpha B(Q)-(c-s_d)Q$	0	0
DCF	$\alpha B(Q)-(c-s_d)Q$	0	0
DFC	$\alpha_d B(Q)-cQ+\boxed{s_d Q}$	0	$\alpha_c B(Q)$
FDC	0	$\alpha_d B(Q)-cQ+\boxed{s_d Q}$	$\alpha_c B(Q)$
FCD	0	$\alpha B(Q)-\alpha_c B(\tilde{Q})-(c-s_d)(Q-\tilde{Q})$	$\alpha_c B(\tilde{Q})-(c-s_d)\tilde{Q}$
CFD	$\alpha_c B(\tilde{Q})-(c-s_d)\tilde{Q}$	$\alpha B(Q)-\alpha_c B(\tilde{Q})-(c-s_d)(Q-\tilde{Q})$	0

Note: All entries are flow surpluses each instant. Dividing by r converts them into present discounted values.

Table B2: Shapley value when country is charged a copayment

	Marginal contribution to surplus		
Permutation	Firm, F	Donor, D	Country, C
CDF	$\alpha B(Q)-(c-s_d)Q$	0	0
DCF	$\alpha B(Q)-(c-s_d)Q$	0	0
DFC	$\alpha_d B(Q)-cQ$	0	$\alpha_c B(Q)+\boxed{s_d Q}$
FDC	0	$\alpha_d B(Q)-cQ$	$\alpha_c B(Q)+\boxed{s_d Q}$
FCD	0	$\alpha B(Q)-\alpha_c B(\tilde{Q})-(c-s_d)(Q-\tilde{Q})$	$\alpha_c B(\tilde{Q})-(c-s_d)\tilde{Q}$
CFD	$\alpha_c B(\tilde{Q})-(c-s_d)\tilde{Q}$	$\alpha B(Q)-\alpha_c B(\tilde{Q})-(c-s_d)(Q-\tilde{Q})$	0

Note: All entries are flow surpluses each instant. Dividing by r converts them into present discounted values.

References

Berndt, Ernst, Rachel Glennerster, Michael Kremer, Jean Lee, Ruth Levine, Georg Weizsäcker, and Heidi Williams, "Advance Market Commitments for Vaccines Against Neglected Diseases: Estimating Costs and Effectiveness," *Health Economics* vol. 16, no. 5 (May 2007) pp. 491–511.

Kremer, Michael, Owen Barder, and Ruth Levine, *Making Markets for Vaccines: Ideas to Action*. Washington, DC: Center for Global Development, 2005.

Kremer, Michael and Rachael Glennerster, *Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases*. Princeton, NJ: Princeton University Press, 2004.

Levine, Orin, "GAVI Alliance Investment Case: Accelerating the Introduction of Pneumococcal Vaccine into GAVI-Eligible Countries," Pneumo-ADIP, October 26, 2006.

Scudellari, Megan. "Are Advance Market Commitments for Drugs a Real Advance?" *Nature Medicine*, vol. 17, no. 2, February 2011, p. 139.

Sinha, Anushua, *et al.*, "Cost Effectiveness of Pneumococcal Conjugate Vaccination in the Prevention of Childhood Mortality: An International Economic Analysis," *Lancet* vol. 369, February 3, 2007, pp. 389–396.

Snyder, Christopher M., Wills Begor, and Ernst R. Berndt, "Economic Perspectives on the Advance Market Commitment for Pneumococcal Vaccines," *Health Affairs* vol. 30, no. 8, 2011, pp. 1508–1517.

World Bank and GAVI (2006) need cite.

World Health Organization, "Pnuemococcal Conjugate Vaccine for Childhood Immunization—WHO Position Paper," *Weekly Epidemiological Record* vol. 82, no. 12, March 23, 2007, pp. 93–104.