Commons with Increasing Marginal Costs: Random Priority versus Average Cost*

Hervé Crès†and Hervé Moulin ‡

September 2000

Abstract

Indivisible units are produced with increasing marginal costs. Under Average Cost, each user pays average cost. Under Random Priority, users are randomly ordered (without bias) and successively offered to buy at the true marginal cost. Both AC and RP inefficiently overproduce. RP tends to overproduce less but which game collects more surplus depends much on the demand configuration. We show that a key to compare the welfare properties of the two mechanisms is the *crowding factor*, i.e., the number of potential users over the number of units of output users can afford: the more crowded the commons, the more RP outperforms AC. In the quadratic cost case, beyond the *threshold value* of 2.4 for the crowding factor, RP strongly outperforms AC, beneath it AC only mildly outperforms RP. Thus the Random Priority mechanism manages crowded commons better than Average Cost.

Keywords: Tragedy of the Commons, Scheduling, Priority, Increasing marginal costs, Cooperative production.

JEL Classification Number: D60, D62, D72.

1 Introduction

We consider a technology producing indivisible units of a homogeneous good that we call a *service* at an increasing marginal cost: c_q is the cost of producing the q-th unit and we assume $0 < c_1 < c_2 < \ldots < c_q < c_{q+1} < \ldots$. Each agent consumes at most one unit of service; agent i is willing to pay u_i, $u_i \geq 0$, for service.

We compare the welfare consequences of two simple and natural mechanisms to exploit the technology —namely, the commons.

The *Average Cost mechanism* (AC) is a one shot game where each agent chooses to buy the good or not and where total cost is equally divided among all the buyers.

The *Random Priority mechanism* (RP) is a sequential game where first, Nature chooses at random (with uniform probability) an ordering of the agents and offers them to buy the good at the successive marginal costs: the agent ranked first is offered the good at price c_1, and if exactly q among those ranked before i did buy, agent i is offered the price c_{q+1}.

The RP and AC games are plausible descriptions of certain commons utilized in the free access regime[1]. Our motivation however is exclusively normative: if both options (to charge average cost or to charge marginal price at random) are available, which one should we recommend? Our answer bears on distributive properties (which agents receive more surplus in what mechanism?) and on efficiency properties (which mechanism induces more overproduction, and which one collects more surplus?). The examples we have in mind are scheduling problems where the users have the option to leave the queue (on which more below), such as congested roads (Naor (1968), Mendelson (1985)) and the Internet (Demers et alii (1990), Shenker (1995)). See also the discussion of average cost versus incremental cost policies in Spulber (1994)[2].

---

[1]Think about our agent as walking around randomly in the forest looking for mushrooms or fruits. Someone will be lucky enough to find the fruit hanging lowest and pick it if the fruit is worth this person's effort to reach for that low branch; the next luckiest agent will find the next lowest hanging fruit and decide similarly whether or not to spend the effort to get it; and so on.

Another example is R & D competition where input is research effort and output is (the present value of) a patent. Think of all teams as supplying the same research effort, and of the lucky one (first in line) as the team which discovers the first and most profitable patent, and so on.

By contrast, the AC mechanism is a plausible description of free access to a commons when the returns of the technology are equal out of necessity, as when agents pump water form a common well and the pressure must be the same for all.

[2]Spulber (1994) discusses a case involving power utilities: at issue is the allocation of the costs of new investments necessary to serve new customers; this cost can be rolled over in the general budget (average

In equilibrium, both mechanisms inefficiently overproduce, but the inefficiencies in AC and RP differ both in size (surplus loss) and nature (surplus distribution). These differences are the subject of this paper: we ask which configurations of the demand (the distribution of the individual willingness to wait) are better handled by the AC or the RP mechanisms, in the sense of generating more surplus and minimizing overproduction. We show examples where one collects the efficient surplus whereas the other collects almost no surplus and vice-versa.

An important example in favor of RP against AC is the case of a perfectly homogeneous demand (all agents are willing to pay the same amont for service). There the RP equilibrium outcome is first best efficient, whereas the AC outcome dissipates all surplus (an extreme case of the tragedy of the commons): everyone is better off under RP than under AC.

The converse configuration, where everyone prefers the AC mechanism to the RP one turns out to be impossible: Proposition 3. We can, however, have a situation where the total surplus collected by AC exceeds that collected by RP: Proposition 5.

Our second main finding is the role of an index that we call the *crowding factor*, that affects the relative welfarewise performance of the two mechanisms. The crowding factor is the ratio of the potential demand (the number of agents who would like to be served if service is free) over the maximal production (the largest willingness to pay, i.e., the largest q such that $c_q \leq \max_i u_i$). This factor is a measure of the level of congestion of the system. We show that the more crowded the commons, the more the RP mechanism outperforms the AC one. In general, when the crowding factor is very large, the latter collects no surplus whereas the former collects a positive fraction of the efficient surplus (Proposition 8). In the canonical model with quadratic costs, the crowding factor is an exact measure of the remative performance of our two mechanisms. See Sections 3, 4 and 6 for precise statements.

1.1   A scheduling example

Two often observed protocols for managing queues are (1) the *organized* queue where a newcomer 'takes a number' and agents are served in turn (this is the familiar 'first in, first out' system) and (2) the *unorganized* queue where the server picks at random someone in the waiting crowd without paying attention to how long this person has been waiting.

cost sharing) or imputed solely to the beneficiaries of the investments (incremental cost sharing). He also argues in favor of the latter on incentives and efficiency grounds.

Most queues among real people (at the post office, in shops, etc.) are 'organized', and the actual distribution of numbered tickets is common when the wait is expected to be long (e.g., at the US Immigration and Naturalization Services). Many mechanical queues are 'unorganized', as when we try to get a line from a busy phone system, access the Internet and so on[3].

In both type of queues the key strategic decision of each agent is to opt out (leave the queue, also called balking in the queuing literature) or stay in the queue. Assume for simplicity that the server processes exactly one agent per unit of time; that agent i's utility for being served after waiting q periods is $u_i - q$, and is zero for balking at time $t = 0$; and that agents are risk neutral. We imagine that all agents show up at time $t = 0$, and then decide to stay or not in the queue. In the inorganized protocol, the expected delay of anyone agent when q agents choose to stay is $(1 + 2 + \ldots + q)/q = (q + 1)/2$ and the corresponding game is precisely the AC mechanism for the technology $c_q = q^4$. In the organized protocol, a priority ordering of all agents is drawn randomly and without bias at time $t = 0$ and the agents decide to stay or not after learning their number. Thus the agent who draws the number 1 is offered service with a delay $c_1 = 1$ and decides at once to stay or not; the agent drawing number 2 is offered a delay $c_2 = 2$ if agent 1 stays and $c_1 = 1$ if agent 1 balks; and in general an agent is offered service with a wait $c_q = q$, if exactly q agents among those who drew a lower number decided to stay. This is precisely the RP mechanism for these commons with linearly increasing marginal costs.

The scheduling model above is probably the simplest example of a commons with increasing marginal costs; the RP and AC mechanisms are especially natural in this model, and easy to implement. Our results throw some light on the choice between these two protocols.

1.2   Contents of the paper

We define the continuous version of our model in Section 2. We have a continuum of agents who each want at most one unit of service, and the profile of utilities (willingness to pay for service) is described by a familiar demand function $p \longrightarrow d(p) \equiv$ the number of agents willing to pay p or more. The continuous version raises some technical difficulties

---

[3]There are examples of unorganized queues among real people and organized mechanical queues: think of a packed bar where the bartender cannot keep track of 'who is next', and of these customer assistance lines where a machine announces the expected waiting time until the next available agent.

[4]Once an agent finds it worthwhile to stay for the first period —as $u_i \geq (q + 1)/2$— he will have no incentive to leave later because the expected wait decreases after each period

(for instance, the noncooperative game has a continuum of players), but on the other hand it yields a simple description of the equilibrium outcomes of AC and RP, and a simple formula for measuring their inefficiency, both in terms of overproduction and of surplus loss: Proposition 1.

Section 3 is devoted to the important special case where the demand and marginal cost functions are both linear. The only relevant parameter in this case is the crowding factor: it determines entirely the overproduction and surplus loss —relative to the efficient benchmark— of our two mechanisms. Beyond the crowding factor 2.4, we find that RP strongly outperforms AC; beneath this critical level, AC mildly outperforms RP: Proposition 2 gives a precise meaning to our opening statement.

In Section 4 we go back to the general case of the continuous model and state our most general results about the distributive consequences of the two mechanisms. A lower tail of the demand always prefers RP to AC whereas the agents preferring AC to RP form a —possibly empty— upper tail of the demand: Proposition 3. In particular, if the demand is flat (identical agents), the RP outcome is fully efficient whereas the AC outcome entirely dissipates the surplus. Some convexity properties of the demand give an edge to one mechanism over the other, irrespective of the cost function: to AC over RP if $1/d$ is concave (Proposition 5) and to RP over AC if $1/d$ is convex (Proposition 6). Finally, Proposition 8 explains that RP asymptotically outperforms AC when the crowding factor goes to infinity as a result of replicating the demand function while keeping the technology fixed.

In the last two sections, we turn to the discrete version of our model, with a finite set of agents demanding at most one unit of service. All definitions and results introduced for the continuous model have a counterpart in the discrete model: Section 5. The interest of the discrete model is twofold. Firstly it provides a rigourous foundation for the continuous model, by viewing the latter as the limit of a natural sequence of discrete models: this important technical fact is explained in Appendix 1. Secondly, the discrete model allows us to make a statistical argument in favor of RP over AC, explained in Section 6. We assume quadratic costs and parametrize the demand space in a natural way. We show that with many agents, for almost all choices of the demand function, the pattern of relative overproductions and surplus losses under AC and RP is precisely the same as in the linear demand case of Section 3: Propositions 9 and 10. This in turn gives a broader support to our finding that 2.4 is the critical value of the crowding factor beyond which we expect RP to strongly outperforms AC.

All proofs are gathered in the appendices.

2 The model with a continuum of agents

An economy consists of a marginal cost and demand function, denoted mc and d. The function mc is defined for all x, $x \geq 0$, strictly increasing and continuous. We assume $mc(0) \geq 0$ and $mc(\infty) = \infty$. We denote by C the cost function, and ac the average cost function. The demand function $d(p)$ is defined for all nonnegative price p; d is nonincreasing and continuous; moreover, for the sake of simplicity, we assume there is a price P above which nobody wants to buy the good: $d(P) = 0$. Denote $X_0 = d(mc(0))$ the size of the potential demand; and set $mc(X_1) = P$. If $mc(0)$ is positive, the part of the demand functions for $0 \leq p < mc(0)$ (the agents willing to pay less than $mc(0)$ for service) plays no role in either the AC or RP mechanisms: these agents never buy service. Therefore we can assume $mc(0) = 0$ without loss of generality.

The crowding factor of the economy (d, mc) is $\gamma = X_0/X_1$. Its role appears in the next sections.
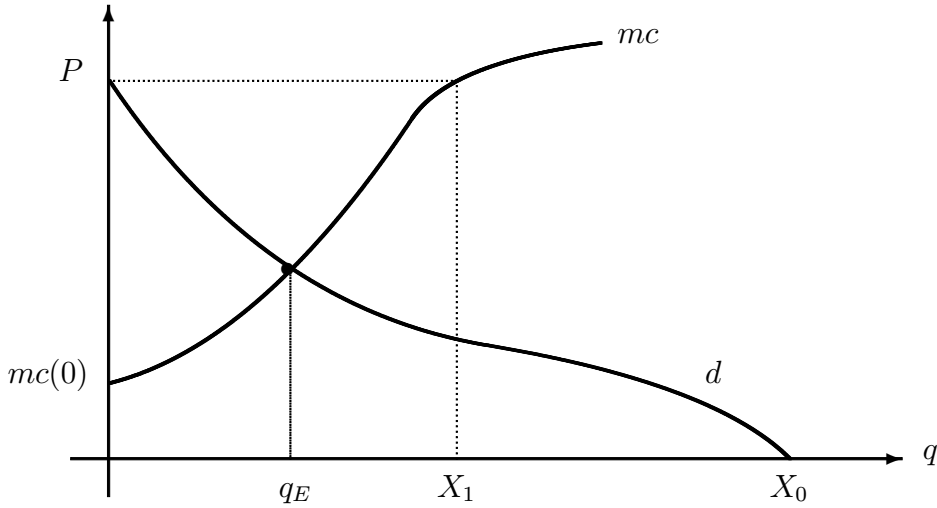

Figure 1

The efficient quantity q_E is at the intersection of the marginal cost and inverse demand curves, i.e., it is the unique solution of $d(mc(q_E)) = q_E$ (see Figure 1). The efficient surplus σ_E is the sum of the consumer surplus at the 'competitive' price $mc(q_E)$ and of the competitive profit:

$$\sigma_E = D(mc(q_E)) + q_E \cdot mc(q_E) - C(q_E) , \tag{1}$$

where $D(p)$ is the consumer surplus at price p:

$$D(p) = \int_p^P d(y) \, dy = - \int_p^P (y - p) \, \dot{d}(y) \, dy .$$

The *Average Cost mechanism* is the noncooperative game (with a continuum of players) where each agent chooses to buy service or not, $x_i = 0$ or 1. An agent who does not buy pays nothing. All agents who do buy pay the average cost $ac(\int x)$, where $\int x$ is the total number of agents who buy.

In equilibrium the quantity q_{AC} produced is where the average cost and inverse demand curves intersect, i.e., it solves the equation:

$$d(ac(q_{AC})) = q_{AC} .$$

(Our assumptions on d and mc guarantee a unique solution.) An agent with valuation (willingness to pay for service) above q_{AC} buys service and pays $ac(q_{AC})$, one with valuation below q_{AC} does not buy. The AC equilibrium surplus is simply the consumer surplus at price $ac(q_{AC})$:

$$\sigma_{AC} = D(ac(q_{AC})) . \tag{2}$$

The *Random Priority mechanism* works as follows. Nature draws an ordering of all agents, with uniform probability on all orderings; agent are successively offered service at an increasing price (defined inductively) and can accept or refuse, $x_i = 0$ or 1; the price offered to agent i is $mc(\int x)$ where the sum bears over all agents preceding i in the ordering in question.

The inductive definition of the above integral is a technical issue that we choose to ignore at this point. The discussion of the RP equilibrium outcome below is similarly heuristic. A more rigourous approach is proposed in Section 5 and Appendix 1, where the RP and AC games with a continuum of agents are viewed as the limit of a sequence of discretized games with finitely many agents.

Think of the RP mechanism as allocating infinitesimally small units δ of the good. After a total of q units have been allocated, the demand above $p = mc(q)$ has 'shrunk' by a factor $\lambda(q)$, $0 \leq \lambda(q) \leq 1$, because some of these agents are already served. The shrinking factor is uniform because all agents above p are treated symmetrically with respect to the allocation of the q units. The shrunk demand is thus

$$d_q(p') = \lambda(q) \cdot d(p') \ \text{ for all } \ p' \geq p = mc(q) .$$

The evolution of λ follows a differential equation. Denote $\epsilon = \dot{mc}(q) \cdot \delta$, where $\dot{mc}$ is the derivative of mc, then

$$d_{q+\delta}(p + \epsilon) = d_q(p + \epsilon) - \delta ,$$

where we neglect a second order term due to the fact that some of δ is allocated to the agents in $[p, p + \epsilon]$. Combining the two equations above:

$$\lambda(q + \delta) \cdot d(p + \epsilon) = \lambda(q) \cdot d(p + \epsilon) - \delta \implies \lambda(q + \delta) - \lambda(q) = -\frac{\delta}{d(p + \epsilon)} = -\frac{\delta}{d(mc(q))}$$

where in the right-hand side equality we again neglect a second order term. Define

$$\phi(q) = \int_0^q \frac{dt}{d(mc(t))} \quad \text{with the convention } \phi(q) = \infty \text{ whenever } q \geq mc^{-1}(P) \,.$$

Now the differential equation governing λ and the initial condition $\lambda(0) = 1$ yield

$$\lambda(q) = 1 - \phi(q) \quad \text{as long as} \quad \phi(q) \leq 1 \,.$$

Two cases may arise. We may reach a level q_{RP} such that $\phi(q_{RP}) = 1$, in which case all agents in $d(mc(q_{RP}))$ have already been served when q_{RP} units have been allocated. In this case the RP mechanism stops because no agent will accept an offer any longer. If, on the other hand we reach $q_{RP} = mc^{-1}(P)$ with $\phi(q_{RP}) \leq 1$, then there are no agent left in $d(mc(q_{RP}))$ and the mechanism stops as well. In both cases we have $d_{q_{RP}}(p) \equiv 0$ whenever $p \geq mc(q_{RP})$ so that q_{RP} is the quantity produced in equilibrium.

Going back to the definition of λ, we consider an agent i with willingness to pay p. For any q such that $mc(q) \leq p$, agent 'p' buys the good at price $mc(q)$ or less with probability $1 - \lambda(q)$. Hence his total probability of service is

$$x_p = \begin{cases} 1 & \text{if} \quad mc^{-1}(p) > q_{RP} \\ \phi(mc^{-1}(p)) & \text{if} \quad mc^{-1}(p) \leq q_{RP} \end{cases}$$

(Note that if $q_{RP} = mc^{-1}(P)$, the upper case is vacuous.) And his expected cost share is

$$y_p = \int_0^b \frac{mc(t)}{d(mc(t))} dt \quad \text{where} \quad b = \inf\{mc^{-1}(p), q_{RP}\} \,.$$

Finally we can compute the surplus collected by RP in equilibrium:

$$\begin{aligned} \sigma_{RP} &= \int_0^\infty (x_p \cdot p - y_p) \, |\dot{d}(p)| \, dp = -\int_0^\infty \left[\int_0^b \frac{p - mc(t)}{d(mc(t))} dt \right] \dot{d}(p) \, dp \\ &= -\int_0^{q_{RP}} \left[\int_{mc(t)}^\infty (p - mc(t)) \, \dot{d}(p) \, dp \right] \frac{dt}{d(mc(t))} \,. \end{aligned}$$

Recall that $D(p)$ is the consumer surplus at p, now we get the formula

$$\sigma_{RP} = \int_0^{q_{RP}} \frac{D(mc(t))}{d(mc(t))} dt \tag{3}$$

We summarize the above discussion as:

Proposition 1 *At the RP equilibrium outcome the quantity q_{RP} is produced with probability one, where*

$$
\begin{cases}
q_{RP} = X_1 & \text{if } \int_0^{X_1} \dfrac{dt}{d(mc(t))} \leq 1 \\[4mm]
q_{RP} \text{ solves } \int_0^{q_{RP}} \dfrac{dt}{d(mc(t))} = 1 & \text{otherwise.}
\end{cases}
\tag{4}
$$

An agent willing to pay $p = mc(q)$ buys service with probability $\int_0^{\bar{q}} \dfrac{dt}{d(mc(t))}$, where $\bar{q} = \min\{q, q_{RP}\}$ and his expected payment is $\int_0^{\bar{q}} \dfrac{mc(t)}{d(mc(t))}$. Total surplus collected at the RP equilibrium is given by (3).

Our goal in the next two sections is to compare q_E, q_{AC} and q_{RP} as well as σ_E, σ_{AC} and σ_{RP}. We start by the canonical example where both the demand and the marginal cost functions are linear.

3   The case of quadratic costs and linear demands

In this section we assume that marginal costs increase linearly, $mc(x) = a \cdot x$ for all $x > 0$ where a is a positive constant.
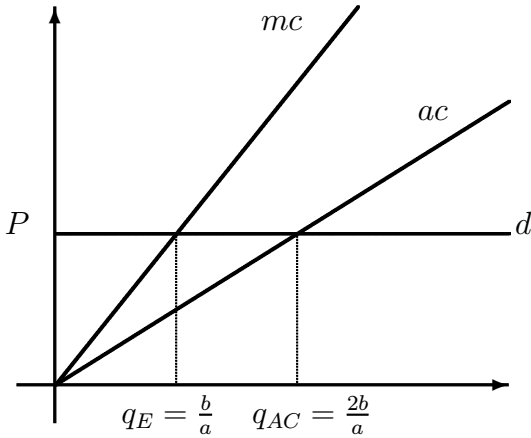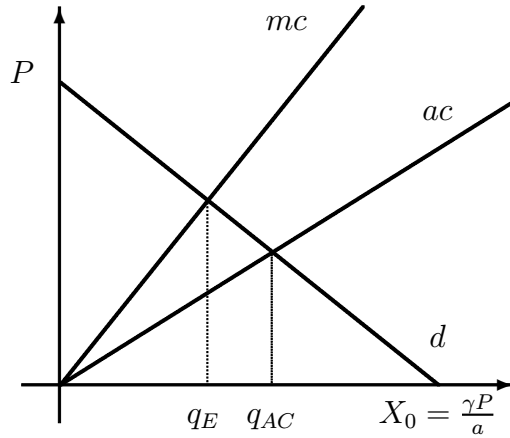

Figure 2

Figure 3

Consider first the case of a flat —infinitely elastic— demand function: $d(p) = X_0$ for $p < b$, $= 0$ for $p > b$, where X_0 is 'large', namely $X_0 > 2b/a$. The demand is perfectly homogeneous, all agents are willing to pay b for service. Under the AC mechanism, we have

a full 'tragedy of the commons', namely inefficient overproduction, $q_{AC} = 2b/a = 2q_E$, up to the point where the average cost equals the common valuation b and so the AC equilibrium dissipates all surplus: $\sigma_{AC} = 0$. See Figure 2.

We can use Proposition 1 to compute the RP equilibrium, provided we account for the discontinuity in d and D. But the direct argument is obvious: the first q_E agents in the random ordering buy the service, the subsequent agents decline. Thus the RP equilibrium is first best efficient: $q_E = q_{RP}$ and $\sigma_E = \sigma_{RP}$. The RP outcome is preferred by every agent to the AC outcome: everyone gets a fair share of the efficient surplus under RP and no surplus at all under AC[5].

We turn to the case of a linearly decreasing demand function, namely

$$
d(p) = \begin{cases} \dfrac{\gamma}{a} \cdot (P - p) & \text{for} \quad 0 \le p \le P \\ 0 & \text{for} \quad P \le p \end{cases}
\tag{5}
$$

In this economy, depicted in Figure 3, we have $X_0 = \gamma \cdot P/a$ and $X_1 = P/a$ so that γ is the crowding factor defined in the preceding section. Applying Proposition 1 delivers easily:

Proposition 2 *In the economy with linear marginal cost $mc(x) = a \cdot x$ and linear demand (5), we have*

$$
q_E = \frac{P}{a} \cdot \frac{\gamma}{\gamma + 1} \; ; \qquad q_{RP} = \frac{P}{a} \cdot (1 - e^{-\gamma}) \; ; \qquad q_{AC} = \frac{P}{a} \cdot \frac{2\gamma}{\gamma + 2} \; ,
$$

and

$$
\frac{\sigma_{RP}}{\sigma_E} = \frac{\gamma + 1}{2\gamma}(1 - e^{-2\gamma}) \; ; \qquad \frac{\sigma_{AC}}{\sigma_E} = \frac{4(\gamma + 1)}{(\gamma + 2)^2} \; .
$$

Thus the relative overproductions q_{AC}/q_E and q_{RP}/q_E, as well as the relative surpluses σ_{AC}/σ_E and σ_{RP}/σ_E only depend upon the crowding factor γ. Figures 4 and 5 depict the overproductions and surpluses when γ varies between 0 and $+\infty$.

Insert Figures 4 and 5 here.

The first observation is that $q_{RP} \le q_{AC}$ at all levels of γ: the RP equilibrium overproduces less than AC. On the other hand AC collects more surplus than RP when $\gamma \le 2.3994$ and vice versa:

$$
\gamma < \gamma^* \implies \sigma_{AC} > \sigma_{RP}
$$
$$
\gamma > \gamma^* \implies \sigma_{AC} < \sigma_{RP}
$$

---

[5]Note that this argument is valid for any marginal cost function.

where $\gamma^* = 2.3994$ is the solution of $e^\gamma = \dfrac{\gamma + 2}{\gamma - 2}$.

Figures 4 and 5 also make clear that the relative advantage of RP and AC is monotonic in γ: both ratios q_{AC}/q_{RP} and σ_{RP}/σ_{AC} increase with γ.

Next we note that when $\gamma < 2.4$, AC collects only a bit more relative surplus than RP: σ_{AC}/σ_{RP} never exceeds 1.03 for any $\gamma \geq 0$. On the other hand, σ_{AC}/σ_{RP} goes to zero when γ goes to infinity, at the rate of $8/\gamma$.

Finally, we note that as γ goes to infinity, the RP equilibrium collects only one half of the efficient surplus: thus we cannot think of the flat demand case —discussed at the beginning of this section— as the limit of the γ-linear economies. Another difference is that, for any γ, the agents with valuation close enough to P prefer AC to RP. At the AC equilibrium, they get service at the following price:

$$p_{AC} = ac(q_{AC}) = P \cdot \frac{\gamma}{\gamma + 2} \ .$$

At the RP equilibrium, the price they pay is given by Proposition 1:

$$p_{RP} = \int_0^{q_{RP}} \frac{a \cdot t}{d(a \cdot t)} dt = P \cdot \frac{1 + \gamma - e^{-\gamma}}{\gamma} \ .$$

One checks easily that $p_{RP} \geq p_{AC}$ for all γ. Recall that if the demand is flat, all agents strictly prefer RP to AC.

4   Welfare comparisons in the general case

We go back to the general model of Section 2 and compare first the distributive consequences of our two mechanisms. In the AC equilibrium outcome, all the 'efficient' agents (i.e., the q_E agents with the largest willingness to pay) are served with probability one. The RP equilibrium allocation, on the other hand, guarantees service only to those agents willing to pay $mc(q_{RP})$ or more, and this does not include all efficient agents because, typically, q_{RP} is strictly larger than q_E.

The AC allocation offers no service at all (and zero surplus) to agents with low utility (i.e., below $ac(q_{AC})$), whereas the RP allocation gives a positive probability of service to any agent with valuation greater than $mc(0) = 0$. This is the source of inefficiency in the equilibrium of RP: efficiency —interpreted as surplus maximization— forbids to serve with any positive probability an agent with valuation smaller than $mc(q_E)$.

The pattern 'low valuation agents prefer RP to AC' and 'high valuation agents prefer AC to RP' is general, with one important qualification: there may be no one preferring AC.

Proposition 3 *Consider an economy (d, mc) with a positive surplus: $\sigma_E > 0 \Longleftrightarrow P > 0$;*

a) there is a valuation $\tilde{p}$, $0 < \tilde{p} \leq P$, such that every agent with valuation in $]0, \tilde{p}[$ strictly prefers the RP equilibrium to the AC one;

b) if there is at least one agent that does not strictly prefer the RP equilibrium to the AC one (resp. that strictly prefers AC over RP), the set of those agents forms an interval $[\tilde{p}, P]$ (resp. an interval $]\tilde{p}, P]$);

c) there are economies where every agent strictly prefers the RP equilibrium to the AC one.

Note the statement c is established by the case of an infinitely elastic demand discussed above.

The next four propositions compare the performance of AC and RP in terms of their relative overproduction and surplus loss, just like in Proposition 2.

Proposition 4 *a) Both the RP and AC equilibrium outcomes inefficiently overproduce: $q_E \leq q_{AC}, q_{RP}$. The overproduction under RP is at most 100%: $q_{RP} \leq 2q_E$.*

b) If RP overproduces more than AC, it also collects less surplus:

$$q_{AC} \leq q_{RP} \implies \sigma_{AC} \geq \sigma_{RP}$$
$$q_{AC} < q_{RP} \implies \sigma_{AC} > \sigma_{RP}$$

In the linear economies of Section 3, the overproduction of the AC equilibrium is at most 100% (it is at most 30% at the RP equilibrium: see Figure 3). However, the cap of 100% on the overproduction under RP (statement a) is independent of both the demand and marginal cost functions. It is easy to construct examples where the overproduction under AC is arbitrarily large[6].

The linear economies of Section 3 provide examples where $q_{AC} > q_{RP}$ and σ_{AC} may be larger or smaller than σ_{RP}. Statement b tells us that the only other possible configuration is when AC overproduces less <u>and</u> collects more surplus. Our next result provides a family of examples:

Proposition 5 *If $1/d$ is a concave (resp. strictly concave) function on the interval $[0, mc(q_{AC})]$ then $q_{AC} \leq q_{RP}$ (resp. $q_{AC} < q_{RP}$).*

Note that the concavity of $1/d$ implies (but is not implied by) the convexity of d: the density of agents increases when the willingness to pay decreases.

---

[6]Take mc sharply concave and d flat or nearly flat. We omit the details.

Given our assumption that the demand vanishes at a finite price P, the function $1/d$ cannot be concave over the entire interval $]0, P[$. Nevertheless it is easy to construct an example illustrating Proposition 5. Fix λ, $0 < \lambda < 1$, and $a, b > 0$, the demand function $d(p) = \dfrac{a}{(p+b)^{\lambda}}$ truncated for p high enough to be above $mc(q_{AC})$, is such an example.

Proposition 6 *If $1/d$ is a convex (resp. strictly convex) function on $[0, P[$ then $q_{AC} \geq q_{RP}$ (resp. $q_{AC} > q_{RP}$).*

Unlike in the case of Proposition 5, we may apply Proposition 6 to any mc function: we do not need to worry about the relative position of $mc(q_{AC})$ and P (i.e. of q_{AC} and X_1), whereas $mc(q_{AC}) \leq P$ is necessary to apply Proposition 5.

Unlike in the case of Proposition 5, it is very easy to come up with simple functional forms of the demand for which $1/d$ is convex. Recall that a function is Log concave if $\text{Log} f$ is concave. Then we have: d concave $\implies d$ Log concave $\implies 1/d$ convex. Consider a demand function of the form $d(p) = f(\alpha - \beta p)$, $0 \leq p \leq \alpha/\beta$, where f is increasing and $f(0) = 0$. Then $1/d$ is convex if $1/f$ is convex. If this holds true, and $mc(q_{AC}) \leq P = \alpha/\beta$, we deduce from Proposition 6 that $q_{AC} \geq q_{RP}$ (and $q_{AC} > q_{RP}$ if $1/f$ is strictly convex). An example is the demand $d(p) = (\alpha - \beta p)^{\lambda}$ for any positive exponent λ, because the function $f(x) = x^{\lambda}$ is Logconcave. Our next result offers a family of demand functions for which we guarantee the pattern $q_{AC} \geq q_{RP}$ and $\sigma_{AC} \geq \sigma_{RP}$.

Proposition 7 *If the function D/d is affine on the interval $[0, mc(q_{AC})]$ then $1/d$ is convex and we have $q_{AC} \geq q_{RP}$ and $\sigma_{AC} \geq \sigma_{RP}$.*

Consider the demand function $d(p) = (\alpha - \beta p)^{\lambda}$, with $\lambda > 0$. The function D/d is affine on the interval $[0, P]$ and only there. Therefore Proposition 7 only applies to a marginal cost function small enough such that $mc(q_{AC}) \leq P$. For instance in the case of a linear demand ($\lambda = 1$) and a linear mc function as in Section 3, this inequality holds if and only if the crowding factor is at most 2.

The last results shows that for infinitely crowded commons, the RP equilibrium systematically outperforms the AC one. We fix an arbitrary economy $\mathcal{E}(1) = (d, mc)$ and we replicate the demand to $\lambda \cdot d$, for some $\lambda \geq 1$, without replicating the marginal cost function: the crowding factor of the economy $\mathcal{E}(\lambda) = (\lambda \cdot d, mc)$ is thus $\gamma(\lambda) = \lambda \cdot \gamma(1)$. We compare the two mechanisms as λ goes to infinity.

It is easy to check on the formulas of Section 2 (e.g., (4)) that $q_E(\lambda)$, $q_{AC}(\lambda)$ and $q_{RP}(\lambda)$ are increasing in λ. Moreover when λ goes to infinity, the proportions of the efficient surplus generated by RP and AC respectively converge. The latter converges to

zero: for infinitely crowded commons, AC collects almost nothing of the efficient surplus. The former converges toward a strictly positive ratio (of course smaller than 1), which means that RP always collects a significant proportion of the efficient surplus, even for infinitely crowded commons. Formally,

Proposition 8

$$\lim_{\lambda \to \infty} q_{AC}(\lambda) = \lim_{\lambda \to \infty} q_{RP}(\lambda) = \lim_{\lambda \to \infty} q_E(\lambda) = X_1$$

and

$$\lim_{\lambda \to \infty} \frac{\sigma_{AC}}{\sigma_E}(\lambda) = 0 \quad and \quad \lim_{\lambda \to \infty} \frac{\sigma_{RP}}{\sigma_E}(\lambda) = \frac{\int_0^{X_1} \frac{D(mc(t))}{d(mc(t))} dt}{P - ac(X_1)} \ .$$

5   The model with a finite set of agents

We turn to the version of our model with a finite set of agents. This makes the notational burden substantially heavier, but on the other hand it allows for a rigorous definition of the strategic games AC and RP. All results of Sections 3 and 4 have a counterpart in the discrete model. Moreover the continuous model can be viewed as the limit of the discrete model when the size n of the population tends to infinity: the straightforward limiting argument is explained in Appendix 1. Thus we interpret the results reviewed in this section as the mathematical justification of those of the previous sections. The less technical reader may go directly to Section 6 presenting a statistical argument genuine to the discrete model.

An economy is a triple $\mathcal{E} = (\mathcal{N}, u, c)$ where $\mathcal{N}$ is the finite set of agents, $i \longrightarrow u_i$ for $i \in \mathcal{N}$ is the utility profile (that is, u_i is agent i's willingness to pay for the service), and $q \longrightarrow c_q$ for $q = 1, 2, \ldots$, is the strictly increasing marginal cost function. We write total cost and average cost as C and ac respectively: $C(q) = \sum_1^q c_k$, $ac(q) = C(q)/q$.

Definition 1 *An agent is said to be of <u>type</u> q if his utility u_i is such that $c_q < u_i \leq c_{q+1}$. The <u>profile of types</u> is the sequence $(n_0, n_1, n_2, \ldots, n_Q)$ such that $\sum_{q=0}^{Q} n_q = n$, where n_q is the number of agents of type q. Let m_q denote the number of agents of type q or more: $m_q = \sum_{k=q}^{Q} n_k$.*

Definition 2 *The crowding factor of the economy $\mathcal{E} = (\mathcal{N}, u, c)$ is $\gamma = n/Q$, where n is the number of agents and Q is the largest integer q such that $u_i \geq c_q$ for some $i \in \mathcal{N}$.*

14

We define successively the RP and AC mechanisms and compute their equilibrium outcome. When the consumption of agent i is a random variable, we interpret u_i as his von Neumann-Morgenstern utility, and assume risk neutrality with respect to monetary payments.

The RP mechanism is *probabilistic*: the allocation it determines is a random variable. Because we assume risk neutrality, all we need to know about agent i's allocation is the probability x_i that he will be served, and his expected payment y_i.

Definition 3 *: **The Random Priority mechanism.** Nature draws agents successively without replacement and with uniform probability (equivalently, an ordering of the agents is drawn at random, with equal probability on all orderings). The agent drawn in the first stage is offered the good at price c_1 and chooses between taking the offer or declining it (in both cases, this agent leaves the game). The second agent in line is offered the good at price c_2 if the first agent did buy at c_1, or price c_1 if the first agent declined. An so on: the agent drawn at stage q is offered the price $c_{q'+1}$, where q' is the number of agents drawn before him who did buy.*

The strategic analysis of this game is transparent. It is a dominant strategy for an agent to 'buy truthfully' (i.e., buy if and only if $u_i > c_{q'}$); barring indifference, the dominant strategy equilibrium is the unique Nash equilibrium, and is also a strong equilibrium (i.e., it resists coalitional deviations). Even if indifferences are allowed, the above equilibrium remains the essentially unique strong equilibrium of the game. It is also Pareto superior to any other Nash equilibrium. Note that these strategic properties are independent of agents' preferences toward risk.

The canonical equilibrium is determined by the *altruistic* tie-breaking rule: whenever indifferent between buying or not, an agent does not buy. We shall maintain this assumption throughout the paper; removing it would complicate the analysis at the margin without bringing any new insight. As not confusion may arise, we simply call this equilibrium 'the' RP equilibrium.

The description of the RP equilibrium allocation only depends upon the partition $(\mathcal{N}_0, \mathcal{N}_1, \ldots, \mathcal{N}_Q)$ where $\mathcal{N}_q$ is the (possibly empty) set of agents of type q (namely, $c_q < u_i \leq c_{q+1}$) and Q is the largest nonempty type. Two agents of the same type behave exactly in the same way in the RP game and receive the same allocation denoted (x^q, y^q). Clearly, this allocation only depends upon the profile of types $n_q = \sharp \mathcal{N}_q$, $0 \leq q \leq Q$ (Definition 1).

The key to compute the RP allocation is the probability β_q that the q-th unit be sold, namely the probability that at least q agents be served. Let $\mathcal{M}_q = \bigcup_{k \geq q} \mathcal{N}_k$ denote the set of agents willing to buy at c_q (with cardinality $m_q = \sum_{k \geq q} n_k$). As long as the mechanism offers the good at a price not larger than c_q, all agents in $\mathcal{M}_q$ behave in exactly the same way. Therefore β_q / m_q is the probability that a given agent in $\mathcal{M}_q$ pays c_q for the good. Now we describe the RP allocation. Clearly $x_i = y_i = 0$ whenever $i \in \mathcal{N}_0$. Next, for all q, $1 \leq q \leq Q$, all $i \in \mathcal{N}_q$:

$$x_i = x^q = \sum_{k=1}^{q} \frac{\beta_k}{m_k} \qquad y_i = y^q = \sum_{k=1}^{q} \frac{\beta_k}{m_k} \cdot c_k \qquad (6)$$

The computation of the sequence β_q is easy for the small values of q but hard afterwards.

Denote by q_E the 'efficient' quantity, i.e., the largest integer q such that $q \leq m_q$. In the RP equilibrium, the first q_E units are sold for sure, therefore $\beta_q = 1$ for $1 \leq q \leq q_E$. Beyond q_E, the number β_q must be computed by a recursive algorithm given in Crès and Moulin (1998) (Proposition 1).

An important measure of the performance of RP is the expected quantity it produces, $q_{RP} = \sum_{k=1}^{Q} \beta_q$. The argument just given shows $q_{RP} \geq q_E$ (with probability 1) namely RP never produces less than the efficient quantity.

We now compute the surplus generated by RP. If the q-th unit is distributed, then it is distributed with equal probability to every agent of type at least q (i.e., by construction, agents labeled $n - m_q + 1$ to n), and generates the stochastic surplus:

$$\sigma_q = -c_q + \frac{1}{m_q} \sum_{i=n-m_q+1}^{n} u_i \ .$$

Thus the expected surplus generated by the RP allocation is computed as follows:

$$\sigma_{RP} = \sum_{q=1}^{Q} \beta_q \, \sigma_q \ . \qquad (7)$$

Definition 4 : **The Average Cost mechanism.** *Each agent chooses to buy the good or not: $x_i = 0$ or 1. An agent who does not buy pays nothing. An agent who does pays average cost. Hence*

$$x_i = 0 \implies y_i = 0 \ ,$$
$$x_i = 0 \implies y_i = ac\left(\sum_{i \in \mathcal{N}} x_i \right) \ .$$

16

The strategic analysis of this mechanism is straightforward and well known. Denote by d the (set valued) demand function: $d(p) = \{i \in \mathcal{N} \mid p < u_i\}$ for all $p \geq 0$. Let q_{AC} be the largest integer such that $q \leq \sharp\, d(ac(q))$. Consider the outcome where exactly q_{AC} agents in an upper tail of $d(ac(q_{AC}))$ are served (that is no agents in $d(ac(q_{AC}))$ is willing to pay more for the good than any agent in the upper tail). This outcome is a Nash equilibrium. There may be other equilibria but the above one is a surplus maximizing equilibrium (verification of these claims is immediate).

In Appendix 1 we explain in what sense the RP and AC equilibrium outcomes of the continuous economy are the limit of the RP and AC outcomes just described for the discrete economy. This convergence argument is intuitive in the case of the AC mechanism: it is standard fare in partial equilibrium analysis. Less so in the case of the RP outcome: the convergence property relies on the nontrivial asymptotic result borrowed from Crès and Moulin (1998) and explained in Appendix 1.

We give now some examples where the RP outcome dominates the AC outcome and vice versa. All examples involve quadratic costs, $c_q = q$, that can be interpreted as a scheduling problem where agent i's utility is $u_i - q$ for receiving service after waiting q periods, and $-q$ for waiting q periods without being served.

Assume first $n = 25$ and a flat demand curve, $u_i = 10.3$ for all $i = 1, \ldots, 25$. The efficient production is $q_E = 10$, and $\sigma_E = 48$. In the AC equilibrium, 19 agents stay in the queue (because $ac(q) = (q + 1)/2$) and $\sigma_{AC} = 5.7$ or about 12% of σ_E. By contrast the RP equilibrium is fully efficient: the first 10 agents in the priority ordering stay in the queue and all others leave[7].

Our next example is one where the AC equilibrium is first best efficient but the RP equilibrium is not. We have $n = 25$ (and quadratic costs). The utility profile is

$$u_i = 1.5 \text{ for } i = 1, \ldots, 10; \ u_i = 2.5 \text{ for } i = 11, \ldots, 20; \ u_i = 5.5 \text{ for } i = 21, \ldots, 25 \ .$$

The AC equilibrium is the efficient outcome where the five high valuation agents, and only them, are served (each paying $ac(5) = 3$) for a total surplus $\sigma_E = \sigma_{AC} = 12.5$. On the other hand, Random Priority inefficiently serves with positive probability each one of the 20 'inefficient' users. It results from the formulas (6) and (7) that the RP outcome collects 7.7 units of surplus, or only 62% of the efficient surplus.

Our last example is a discrete analog of the linear demand/quadratic costs continuous economies of Section 3. The two equilibrium allocations collect nearby the same surplus

---

[7]More generally, in a discrete economy with quadratic costs where all agents are of the same type Q and $n > Q$, one checks that the ratio σ_{RP}/σ_{AC} is of order of $2Q$.

but have different distributive consequences. We have $n = 4$ and a 'uniform' profile of types (Definition 1). The utilities of our agents are spread as follows:

$$1 < u_1 \le 2 < u_2 \le 3 < u_3 \le 4 < u_4 \le 5 \quad \text{with} \quad c_q = q \quad \text{for} \quad q = 1, \ldots, 5 . \qquad (8)$$

The efficient quantity is $q_E = 2$ and the total surplus is $\sigma_E = u_3 + u_4 - 3$. To compute the RP allocation, we need to determine the probability β_q that the q-th unit is sold:

$$\beta_1 = \beta_2 = 1 \quad \text{and} \quad \beta_3 = \frac{3}{4} \quad \beta_4 = \frac{1}{24} .$$

The AC equilibrium allocation is unique with agents 2, 3 and 4 active and agent 1 left out. Comparing the two allocations we find that agent 1 always prefer RP to AC, and agents 3,4 always prefer AC to RP. As for agent 2, he prefers RP if and only if $u_2 < 2.6$. Finally, total surplus collected may be larger in either one of the two allocations, but the ratio remains close to one. When $(u_i)_{i=1}^4$ varies within the limits given in (8), the bounds of this ratio are: $0.865 \le \dfrac{\sigma_{RP}}{\sigma_{AC}} \le 1.042$.

The entire discussion of Section 4 remains valid in the discrete economy. One can prove the analog of Proposition 3 about which agents prefer RP over AC and vice versa, and of Proposition 4 about the property $q_{AC} \le q_{RP} \Longrightarrow \sigma_{AC} \ge \sigma_{RP}$. The same holds true for Propositions 5, 6, 7 and 8, all of which have exactly the same interpretations as in the continuous model.

We turn to a statistical comparison of RP and AC in discrete economies with quadratic costs and a large number of agents.

6 Statistical comparison of RP and AC

In this section, we compare the asymptotic behavior of AC and RP from a statistical angle. We look at a discrete economy with quadratic cost $c_q = q$ and a large number of agents, and we *randomize* over all profiles of types.

Here we consider a population where all n agents are potentially active in the distribution process (i.e., $\mathcal{N}_0 = \emptyset$), and have a highest possible utility for service strictly smaller than c_{Q+1}. The number of overall possible profiles is the number of ways one can distribute n balls into Q urns: $\begin{pmatrix} Q + n - 1 \\ Q - 1 \end{pmatrix}$. We call *Bose-Einstein* statistics (BE-statistics in short), the uniform distribution over all these possible profiles of types.

The asymptotic BE-statistics is the limit of this probability distribution when the number n of agents, as well as the number $Q = n/\gamma$ of units distributed (where γ is

fixed), goes to infinity. One then obtains the following convergence result: the variance of the random variable $x_M^{(n)} = q_M^{(n)}/Q$ (where $M = E$ or RP or AC) converges to zero asymptotically. Hence its limit x_M behaves according to a Dirac measure with probability 1 on the asymptotic limit of the mean of the random variable $q_M^{(n)}/Q$.

Proposition 9 *Asymptotically under the BE-statistics, with probability one, the efficient proportion x_E of the Q units that should be distributed, and the proportion x_{RP} (resp. x_{AC}) of the Q units that is distributed by RP (resp., by AC under the assumption of quadratic costs) are:*

$$x_E = \frac{\gamma}{\gamma + 1}, \qquad x_{RP} = 1 - e^{-\gamma}, \qquad x_{AC} = \frac{2\gamma}{\gamma + 2}.$$

The interpretation is clear. Asymptotically, a profile of types chosen at random among all the possible ones[8] will *almost surely* have the above ratios.

Proposition 10 *Asymptotically under the BE-statistics, with probability one, the ratio of the efficient surplus secured by RP and AC are:*

$$\frac{\sigma_{RP}}{\sigma_E} = \frac{\gamma + 1}{2\gamma}(1 - e^{-2\gamma}), \qquad \frac{\sigma_{AC}}{\sigma_E} = \frac{4(\gamma + 1)}{(\gamma + 2)^2}.$$

Comparing with Proposition 2, we find that the quantity produced and surplus collected under the Bose-Einstein statistics are exactly the same as under the linear demand economy of Section 3! In particular, as shown by Figure 4, AC overproduces more than RP for any crowding factor. In terms of surplus collected, AC mildly outperforms RP if γ is below 2.4 and RP strongly outperforms AC when γ is above 2.4. See Section 3.

The Bose-Einstein statistics over profiles of types results from a simple urn model[9] for dispatching agents within the intervals of marginal costs. Consider an urn containing Q balls, one for each type. Agents line up in front of the urn in a random order, and the first one selects his type by drawing a ball at random in the urn. He then puts back the ball into the urn together with one other of the same type. Hence the urn is modified: the second agent in line chooses his type by drawing a ball at random in the urn —which contains now $Q+1$ balls— and puts it back into the urn with one more of the same type. Consequently, all agents draw and modify the urn which at the end of the process contains $Q+n$ balls. This way one generates the Bose-Einstein statistics over the profiles of types:

---

[8]Recall that n agents are dispatched within Q types at most (some types can be empty, even the last one: Q).

[9]Known as the Polya-Eggenberger urn schemes.

a type which has already been drawn is more likely to be drawn once more. Hence the economy is more homogeneous that if the urn were not changed after one draw. This last case corresponds to the Maxwell-Boltzmann statistics, which obviously generates an almost uniform profile of types if $n >> Q$. Thus Propositions 9 and 10 hold under the MB statistics as well, and in that case they follow at once from Proposition 2.

One basic point is that the intervals of marginal costs are implicitly assumed to be approximately of the same length, so that they can all be represented by one and only one ball inside the urn. This is the case as soon as the marginal cost curve keeps asymptotically a linear direction (with positive, finite slope), the reason why we only used quadratic costs in this section.

7   Concluding comments

Does RP outperform AC as a simple mechanism to manage the commons?

The answer is unambiguously yes when the demand function is uniform, or nearly uniform (see the discussion of Sections 3 and 6), and Propositions 8 establishes that the answer is yes whenever the commons are sufficiently crowded.

The robustness of these results could be tested in several natural variations of the model. One of them is the dual 'output sharing' game, where each agent contributes 0 or 1 unit of input (e.g., labor) and total output $F(y)$ produced when y agents do work must be shared among all these workers. This second model is standard fare to discuss the exploitation of natural resources, such as fisheries (Gordon (1954), Levahri and Mirman (1975)), forests, oil reserves (Dasgupta and Heal (1979)), etc. Input represents the fishing, logging or pumping effort and output is the total catch. It is easy to adapt our results to the output sharing context.

Another natural variation of our model is to allow for variable demands: agents can buy more than one unit of the good. The RP mechanism generalizes as follows: units of output with increasing marginal costs are successively offered to an agent selected randomly and without bias from the pool of active agents —where an agent is active as long as he did not refuse any such offer. When the pace of discretization (the size of each indivisible unit) goes to zero, this mechanism converges to the serial cost sharing mechanism analyzed by Moulin and Shenker (1992). The welfare comparison of the serial and average cost sharing mechanisms is still a widely open subject: see, however, Moulin and Shenker (1994) and Chapter 6 in Moulin (1995).

References

Bogomolnaia, A. and Moulin, H. (1999). A simple random assignment problem with a unique solution, Mimeo, Duke University.

Crès, H. and Moulin, H. (1998). Scheduling with Opting Out: Improving Upon Random Priority, *Cahiers de recherche, HEC*, CR656/1998, forthcoming *Operations Research*.

Dasgupta, P. and Heal, G. (1979). *Economic Theory and Exhaustible Resources*, Cambridge: Cambridge University Press.

Demers, A., Keshav, S. and Shenker, S. (1990). Analysis and simulation of a fair queuing algorithm, *International Research and Experience*, 1: 3-26.

Feller, W. *An Introduction to Probability Theory and Its Applications,* volume 1, Wiley, New York, NY, 1971.

Gordon, H. (1954). The Economic Theory of a Common-Property Resource: The Fishery, *Journal of Political Economy*, 62: 124-142.

Levahri, D., and Mirman, L. (1980). The Great Fish War: An Example Using a Dynamic Cournot-Nash Solution, *Bell Journal of Economics*, 11: 322-334.

Mendelson, H. (1985). Pricing computer services: Queuing effects, *Communications*, 28: 312-321.

Moulin, H. (1995). *Cooperative Microeconomics: a Game Theoretic Introduction*, Princeton University Press.

Moulin, H., and Shenker, S. (1992). Serial Cost Sharing, *Econometrica*, 50: 1009-1039.

Moulin, H., and Shenker, S. (1994). Average Cost Pricing Versus Serial Cost Sharing : An Axiomatic Comparison, *Journal of Economic theory*, 64: 178-201.

Naor, P. (1969). The regulation of queue size by levying tolls, *Econometrica*, 37: 15-24.

Shapley, L., and Shubik, M. (19697). Pure Competition, Coalitional Power and Fair Division, *International Economic Review*, 10: 337-362.

Shenker, S. (1995). Making greed work in network: A game theoretic analysis of switch service discipline, *IEEE/ACM Transactions on Networking*, 3: 819-831.

Spulber, D. (1994). Pricing and the Incentive to Invest in Pipelines After *Great Lakes*, *Energy Law Journal*, 15: 377-404.

Tovey, C. (1997). Probabilities of Preferences and Cycles with Super-Majority Rules, *Journal of Economic Theory*, 70: 471-479.

Appendix 1

From the continuous to the discrete model

We fix a continuous economy (d, mc) and construct a canonical $\mathcal{E}^{(n)}$ of discrete economies, one for each cardinality of $\mathcal{N}$, approximating more and more finely the continuous economy.

From now on, fix the size, n of $\mathcal{N}$. Consider the discrete economy $\mathcal{E}^{(n)}$ with n agents. We label the potential users of the service: agent i, with $1 \leq i \leq n$. Fix the utility profile in the following way: for all i, $0 \leq i \leq n - 1$, agents i has a utility (reservation value) for the service:

$$u_i = d^{-1}\left(\frac{X}{n}(n - i)\right)$$

where for all $x \geq 0$, $d^{-1}(x) = \sup\{p \geq 0 \mid d(p) \geq x\}$. If the service were proposed for free, each agent would demand one unit, therefore n units of service would be provided. Hence along the quantity axis, the unit measure is $\frac{X}{n}$. Given the utility profile, since the agent with the highest utility is agent labeled n (with a utility P), the maximum number of units of service that the institution can possibly provide is approximately $\frac{\overline{X}}{X}n$. In fact we define it as[10]:

$$Q = \left\lceil \frac{\overline{X}}{X}n \right\rceil .$$

We fix the marginal costs of the successive units of service produced to be, for all q, $1 \leq q \leq Q + 1$[11],

$$c_q = mc\left(\frac{\overline{X}}{Q + 1}q\right) .$$

For future reference, we compute the profile of types of the economy $\mathcal{E}^{(n)}$. Define the function δ, from the unit interval $[0, 1]$ into itself by:

$$\forall t \in [0, 1], \delta(t) = \frac{1}{X} d \circ mc(\overline{X}t) .$$

The function δ is continuous, nonincreasing and $\delta(0) = 1$, $\delta(1) = 0$.

Lemma 1 *For the discrete economy $\mathcal{E}^{(n)}$, for all q, $1 \leq q \leq Q$, the number of agents of type q is:*

$$n_q^{(n)} = \left\lceil n\delta\left(\frac{q}{Q + 1}\right) \right\rceil - \left\lceil n\delta\left(\frac{q + 1}{Q + 1}\right) \right\rceil ,$$

---

[10]Denote $\lceil x \rceil$ the smallest integer higher than or equal to x; and $\lfloor x \rfloor$ the highest integer smaller than or equal to x.

[11]The $(Q + 1)$-th unit is never sold.

and the number of agents of type at least q:

$$m_q^{(n)} = \left\lceil n\delta \left(\frac{q}{Q+1} \right) \right\rceil .$$

Proof: Recall that agent i is of type q if and only if $c_q < u_i \leq c_{q+1}$. Then for all q, $1 \leq q \leq Q$ [12],

$$n_q^{(n)} = \sharp \left\{ i, 0 \leq i \leq n-1 \mid n\delta \left(\frac{q+1}{Q+1} \right) \leq n-i < n\delta \left(\frac{q}{Q+1} \right) \right\} ,$$

and an integer k is strictly smaller (resp. weakly bigger) than x if an only if it is strictly smaller (resp. weakly bigger) than $\lceil x \rceil$. Here we take $k = n - i$, and the desired formula obtains at once. $\square$

Hence the function δ is the distribution function of agents across the intervals of marginal costs.

Fortunately, when n goes to infinity and the discrete economies $(\mathcal{N}, u, c)$ converge toward the continuous economy (d, mc), the limiting behavior of the sequence (β_q) is given by a simple threshold function, which allows us to compute easily the quantity produced and surplus collected at the RP equilibrium in the continuous model. The following theorem is a rigourous proof of Proposition 1.

Theorem 1 *(Crès and Moulin (1998)) In the continuous economy defined by the demand and marginal cost curve d and mc, the RP equilibrium allocation produces the quantity q_{RP} with probability one, where q_{RP} is defined as follows:*

$$\begin{cases} q_{RP} = X_1 & \text{if} & \int_0^{X_1} \frac{dx}{d(mc(x))} \leq 1 \\\\ q_{RP} \in [0, X_1[& solves & \int_0^{q_{RP}} \frac{dx}{d(mc(x))} = 1 & otherwise. \end{cases}$$

An agent of type $p = mc(q)$ receives service with probability $\displaystyle\int_0^{\bar{q}} \frac{dx}{d(mc(x))}$ and his expected payment is $\displaystyle\int_0^{\bar{q}} \frac{mc(x)\,dx}{d(mc(x))}$, where $\bar{q} = \min\{q, q_{RP}\}$.

In other words, the sequence of "curves" $\left((q/Q, \beta_q)_{q=1}^Q \right)_n$ converges pointwise toward the threshold "function":

$$\Phi_c(x) = \begin{cases} 0 & \text{for } x > x_{RP} = q_{RP}/X_1 \\ [0,1] & \text{for } x = x_{RP} \\ 1 & \text{for } x < x_{RP} \end{cases} .$$

As for the argument of convergence of the AC equilibrium in the discrete case toward its continuous version, it is standard and omitted.

Appendix 2: Proofs in the continuous model

Proof of Proposition 2: Omitted, straightforward. $\square$

---

[12] The symbol $\sharp Z$ stands for the cardinality of the set Z.

Proof of Proposition 3 and 4: A consequence of the proof of their analog in the discrete case, in Appendix 3, and of the limit argument. □

Proof of Proposition 5, 6 and 7: Fix mc and d. Suppose $1/d$ is concave on $[0, mc(q_{AC})]$ and apply Jensen's inequality to this function and the following integral:

$$\frac{1}{q_{AC}} \int_0^{q_{AC}} \frac{1}{d}[mc(t)]dt \leq \frac{1}{d}\left(\frac{1}{q_{AC}} \int_0^{q_{AC}} mc(t)dt\right) = \frac{1}{d[ac(q_{AC})]} \ .$$

By definition, q_{AC} is the solution of $d[ac(q_{AC})] = q$ so the above inequality gives

$$\int_0^{q_{AC}} \frac{1}{d[mc(t)]}dt \leq 1 \ \text{ implying } \ q_{AC} \leq q_{RP} \ \text{ by Theorem 1 .}$$

Next we suppose that $1/d$ is convex on $[0, P[$ and show $q_{AC} \geq q_{RP}$. From Proposition 1, q_{RP} is not larger than X, therefore if $q_{AC} \geq X_1$ there is nothing to prove. Assume now $q_{AC} < X_1 \Longleftrightarrow mc(q_{AC} < P$. From Proposition 1 and Jensen inequality for the function $1/d$ on the interval $[0, mc(q_{AC})]$, we have:

$$\int_0^{q_{AC}} \frac{1}{d[mc(t)]}dt \geq 1 \geq \int_0^{q_{RP}} \frac{1}{d[mc(t)]}dt \ ,$$

and the desired conclusion $q_{AC} \geq q_{RP}$.

Finally we suppose that D/d is affine. First we check that $1/d$ is convex. Set $f = 1/d$. If f is twice differentiable, its convexity obtains easily by differentiating $D/d = D \cdot f$ twice:

$$\overbrace{D \cdot f}^{\bullet} = -d \cdot f + D \cdot \dot{f} = -1 + D \cdot \dot{f} \text{ and } 0 = \overbrace{D \cdot f}^{\bullet\bullet} = -g \cdot \dot{f} + D \cdot \ddot{f} \Longrightarrow \ddot{f} = \frac{g \cdot \dot{f}}{D} \geq 0 \ .$$

The easy argument in the case where f is not twice differentiable is omitted.

As $1/d$ is convex, Proposition 7 implies $q_{AC} \geq q_{RP}$. We combine this inequality with the fact that an affine function commutes with the integral operation:

$$\frac{1}{q_{AC}} \int_0^{q_{AC}} \frac{D}{d}[mc(t)]dt = \frac{D}{d}\left(\frac{1}{q_{AC}} \int_0^{q_{AC}} mc(t)dt\right) = \frac{D[ac(q_{AC})]}{d[ac(q_{AC})]} = \frac{\sigma_{AC}}{q_{AC}}$$

$$\Longrightarrow \ \sigma_{AC} = \int_0^{q_{AC}} \frac{D}{d}[mc(t)]dt \geq \int_0^{q_{RP}} \frac{D}{d}[mc(t)]dt = \sigma_{RP} \ .$$

Proof of Proposition 8: The first assertion, $\lim q_{RP} = X_1$, follows easily from Theorem 1. Moreover, for λ infinite, $(\lambda d)^{-1}(t) \to P$ for all t; therefore we get straightforwardly $\lim q_E = X_1$. Finally the limit value of q_{AC} when $\lambda \to \infty$ satisfies: $\lim q_{AC} = ac^{-1}(P)$.

As a consequence $\lim_{\lambda \to \infty} D[ac(q_{AC})] = 0$, which implies the fourth equality. Finally the fact that $\int_0^{q_E} [(\lambda d)^{-1}(x) - mc(x)] dx \to P - ac(X_1)$ allows to conclude. □

Appendix 3: Proofs of Section 6

The parameter space of profiles of types: the $(Q-1)$–simplex, and the uniform distribution

A distribution $(n_1, n_2, \ldots, n_Q)$ can be represented by a point in the $(Q-1)$–simplex:

$$\mathcal{S}_{Q-1} = \left\{\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_Q) > 0 \ | \ \sum_{q=1}^{Q} \lambda_q = 1\right\} ,$$

with $\lambda_q = \dfrac{n_q}{n}$.

Consider $Q-1$ random variables: $X_1, X_2, \ldots, X_{Q-1}$, uniformly distributed in $\left\{0, \dfrac{1}{n}, \dfrac{2}{n}, \ldots, \dfrac{n-1}{n}, 1\right\}$; i.e., they all reach each value with proba $\dfrac{1}{n+1}$. Reordering the $Q-1$ points $X_1, X_2, \ldots, X_{Q-1}$ in their natural order, one gets $Q-1$ new random variables denoted $X_{(1)}, X_{(2)}, \ldots, X_{(Q-1)}$. These random variables are called the *order statistics*. For notational convenience, we set $X_{(0)} = 1$, $X_{(Q)} = 1$.

There exists a simple characterization of the uniform distribution on the $(Q-1)$–simplex.

Lemma 2 *(Tovey (1997)) The uniform distribution on the* $\begin{pmatrix} Q+n-1 \\ Q-1 \end{pmatrix}$ *attainable points of the simplex S_{Q-1} is equivalent to taking the random variable $\lambda = (\lambda_q)_{q=1}^{Q}$ where $\lambda_q = X_{(q)} - X_{(q-1)}$.*

We give now some properties of the distribution of the order statistics.

Lemma 3 *For all $q \geq 1$, the distribution function of the q-th order statistics is*

$$\mathbf{P}\left\{X_{(q)} \leq \frac{k}{n}\right\} = \sum_{r=q}^{Q-1} \begin{pmatrix} Q-1 \\ r \end{pmatrix} \left(\frac{k+1}{n+1}\right)^{r} \left(\frac{n-k}{n+1}\right)^{Q-1-r} .$$

Proof: The random variable $X_{(q)}$ is smaller or equal to $\dfrac{k}{n}$ if and only if at least q among the random variables $X_1, X_2, \ldots, X_{Q-1}$ are in $\left\{0, \dfrac{1}{n}, \ldots, \dfrac{k}{n}\right\}$, and the remaining ones are in $\left\{\dfrac{k+1}{n}, \ldots, 1\right\}$, hence the result. $\square$

The uniform distribution over profiles of types

In the framework presented here, by setting, for all q, $1 \leq q \leq Q$,

$$n_q = n(X_{(q)} - X_{(q-1)}) \quad \text{thus} \quad m_q = n(1 - X_{(q-1)}) ,$$

the uniform distribution over all attainable profiles of types is generated. The parameter space of profiles of types being endowed with this statistics, we can study the asymptotic distribution (i.e., for $(n, Q) \longrightarrow \infty$, with $\dfrac{n}{Q} \longrightarrow c$, $0 < c \leq \infty$) of the quantities distributed by each mechanism.

Proof of proposition 9: For a fixed q, the probability that $q_E \geq q$ is exactly the probability that $m_q \geq q$, then is equal, by lemma 3, to

$$\mathbf{P}\left\{X_{(q-1)} \leq 1 - \frac{q}{n}\right\} = \sum_{r=q-1}^{Q-1} \begin{pmatrix} Q-1 \\ r \end{pmatrix} \left(1 - \frac{q}{n+1}\right)^{r} \left(\frac{q}{n+1}\right)^{Q-1-r} .$$

This is exactly the probability that $Q-1$ Bernoulli trials, with probabilities $1 - \dfrac{q}{n+1}$ for success and $\dfrac{q}{n+1}$ for failure, result in *at least* $q-1$ successes.

Let Y_{Q-1} denote the ratio of successes in $Q-1$ Bernoulli trials, with probabilities $1 - \dfrac{q}{n+1}$ for success. The mean and variance of Y_{Q-1} are respectively (see Feller (1957)):

$$\mathrm{E}\,(Y_{Q-1}) = 1 - \frac{q}{n+1} \quad \text{and} \quad \mathrm{Var}\,(Y_{Q-1}) = \frac{1}{Q-1}\left(1 - \frac{q}{n+1}\right)\frac{q}{n+1} .$$

Hence by Chebishev's inequality, for $\dfrac{q-1}{Q-1} > 1 - \dfrac{q}{n+1}$, one has:

$$
\begin{aligned}
\mathbf{P}\left\{X_{(q-1)} \le 1 - \frac{q}{n}\right\} &= \mathbf{P}\left\{Y_{(Q-1)} \ge \frac{q-1}{Q-1}\right\} \\
&\le \left(\frac{q-1}{Q-1} + \frac{q}{n+1} - 1\right)^{-2} \mathrm{Var}\,(Y_{Q-1}) \,.
\end{aligned}
$$

Consider a sequence $(n,Q) \longrightarrow \infty$ in such a way that $\dfrac{n}{Q} \longrightarrow \gamma$, and a sequence $(q^{(n,Q)})_{(n,Q)}$ such that $\dfrac{q^{(n,Q)}}{Q} \longrightarrow x \in [0,1]$. Since $\lim\limits_{(n,Q)\to\infty} Var(Y_{Q-1}) = 0$, the condition

$$
\lim_{(n,Q)\to\infty} \frac{q-1}{Q-1} + \frac{q}{n+1} - 1 \left(= x + \frac{x}{\gamma} - 1\right) > 0 \iff x > \frac{\gamma}{\gamma+1} \,,
$$

entails: $\lim\limits_{(n,Q)\to\infty} \mathbf{P}\{q_E \ge q\} = 0$. Concluding through the same way in the other case, one gets:

$$
\lim_{(n,Q)\to\infty} \mathbf{P}\{q_E \ge q\} = \begin{cases} 0 \ \text{ if } \ x > \dfrac{\gamma}{\gamma+1} \\[2mm] 1 \ \text{ if } \ x < \dfrac{\gamma}{\gamma+1} \end{cases} .
$$

We turn to the asymptotic distribution of q_{AC}. Under the assumption of quadratic costs and the uniformity assumption, for a fixed q, the probability that $q_{AC} \ge q$ is exactly the probability that $m_{\lfloor \frac{q+1}{2}\rfloor} \ge q$, so it is equal, by lemma 3, to $\mathbf{P}\left\{X_{(\lfloor \frac{q-1}{2}\rfloor)} \le 1 - \dfrac{q}{n}\right\}$. By an argument similar to the one used above, this yields:

$$
\lim_{(n,Q)\to\infty} \mathbf{P}\{q_{AC} \ge q\} = \begin{cases} 0 \ \text{ if } \ x > \dfrac{2\gamma}{\gamma+2} \\[2mm] 1 \ \text{ if } \ x < \dfrac{2\gamma}{\gamma+2} \end{cases} .
$$

The last case to be investigated is the one of Random Priority. Define the $Q-1$ new random variables $Z_q = \dfrac{1}{1-X_q}$, $1 \le q \le Q-1$, such that one also has $Z_{(q)} = \dfrac{1}{1-X_{(q)}}$. The random variable Z_q reaches with equal probability the $n+1$ values $\left\{1, \dfrac{n}{n-1}, \dots, \dfrac{n}{n-q}, \dots, n, \infty\right\}$. The same use of Chebishev's inequality yields, for $k > (Q-1)\dfrac{q+1}{n+1}$,

$$
\mathbf{P}\left\{Z_{(k)} \le \frac{1}{1-\frac{q}{n}}\right\} \le \left(\frac{k}{Q-1} - \frac{q+1}{n+1}\right)^{-2} \frac{1}{Q-1}\left(1 - \frac{q+1}{n+1}\right)\frac{q+1}{n+1} \,.
$$

Hence the cumulative distribution function of the random variable $Z_{(k)}$ converges to a threshold function with threshold at $q = \dfrac{n+1}{Q-1}k - 1$ (the function equals zero before q, one after).

Let us now prove by induction that for $\eta > 0$,

$$
\lim_{(n,Q)\to\infty} \mathbf{P}\left\{\sum_{k=0}^{q-1} Z_{(k)} \le \sum_{k=0}^{q-1} \frac{1}{1-\frac{k}{Q}} - q\eta\right\} = 0 \,.
$$

Suppose it is true at rank $q-1$. One has

$$
\mathbf{P}\left\{\sum_{k=0}^{q} Z_{(k)} \le \sum_{k=0}^{q} \frac{1}{1-\frac{k}{Q}} - (q+1)\eta\right\} \le \sum_{h=0}^{n} \mathbf{P}\left\{Z_{(q)} = \frac{1}{1-\frac{h}{Q}} - \eta\right\} \times
$$

26

$$\mathbf{P}\left\{\sum_{k=0}^{q-1} Z_{(k)} \leq \sum_{k=0}^{q-1} \frac{1}{1-\frac{k}{Q}} - q\eta + \frac{1}{1-\frac{q}{Q}} - \frac{1}{1-\frac{h}{Q}}\right\} . \tag{9}$$

Fix $\epsilon > 0$. By the induction assumption and the Chebishev's inequality respectively, there exists (n_1, Q_1) such that if $(n, Q) > (n_1, Q_1)$, then for $h \geq q$[13],

$$\mathbf{P}\left\{\sum_{k=0}^{q-1} Z_{(k)} \leq \sum_{k=0}^{q-1} \frac{1}{1-\frac{k}{Q}} - q\eta + \frac{1}{1-\frac{q}{Q}} - \frac{1}{1-\frac{h}{Q}}\right\} \quad < \quad \frac{\epsilon}{2}$$

$$\text{and } \mathbf{P}\left\{Z_{(q)} \leq \frac{1}{1-\frac{q-1}{Q}} - \eta\right\} < \frac{\epsilon}{2} .$$

Therefore, separating the sum in the righthand side of inequality (9) into $\sum\limits_{h=1}^{q-1}$ and $\sum\limits_{h=q}^{n}$, one gets, for $(n, Q) > (n_1, Q_1)$,

$$\mathbf{P}\left\{\sum_{k=0}^{q} Z_{(k)} \leq \sum_{k=1}^{q} \frac{1}{1-\frac{k}{Q}} - (q+1)\eta\right\} \leq \epsilon ,$$

hence the result. We can then prove that the cumulative distribution function of the random variable $\frac{1}{q}\sum\limits_{k=0}^{q-1} Z_{(k)}$ converges toward a threshold function with threshold at

$$\lim \frac{1}{q}\sum_{k=0}^{q-1} \frac{1}{1-\frac{k}{Q}} = -\frac{1}{x}\mathrm{Log}(1-x) .$$

Finally, given that $\sum\limits_{k=1}^{q} \frac{1}{m_k} = \frac{1}{n}\sum\limits_{k=0}^{q-1} Z_{(k)}$, x_{RP} is such that $\frac{x}{\gamma}\left(-\frac{1}{x}\right)\mathrm{Log}(1-x) = 1$, hence the result. $\square$

Proof of proposition 10: It is a direct corollary of the preceding proposition. The asymptotic threshold for the quantities distributed yields an asymptotic threshold for the surpluses generated. Under the assumption of quadratic costs, the surplus generated by the q-th unit distributed is contained between $\sum\limits_{k=q}^{Q} n_k(k-q)$ and $\sum\limits_{k=q}^{Q} n_k(k-q+1)$. Its normalized mean value (i.e., mean value divided by Q^2) then converges toward $\frac{(1-x)^2}{2}$. The normalized mean value (i.e., mean value multiplied by n) of the ratio $\frac{1}{m_q}$ converges toward $\frac{1}{1-x}$. Hence the normalized mean value of the surplus generated by RP is

$$\sigma_{RP} = \int_0^{x_{RP}} \frac{1}{1-x}\frac{(1-x)^2}{2}dx = \frac{1}{4}\left(1-e^{-2\gamma}\right) .$$

The distribution of surpluses degenerates around this mean value, since its variance can easily be bounded above by use of the variance of the distribution of the quantities distributed, which itself converges toward 0 thanks to the preceding proposition.

The same line of argument gives the normalized mean value of the efficient surplus

$$\sigma_E = \gamma \int_0^{\frac{x_E}{\gamma}} (1-x)dx - \frac{x_E^2}{2} = \frac{\gamma}{2(\gamma+1)} ,$$

---

[13]Recall that $\frac{n}{Q} \longrightarrow c$, and $(q^{(n,Q)})_{(n,Q)}$ is such that $\frac{q^{(n,Q)}}{Q} \longrightarrow x \in [0,1]$.

and the normalized mean value of the surplus generated by AC:

$$\sigma_{AC} = \gamma \int_0^{\frac{x_{AC}}{\gamma}} (1-x)dx - \frac{x_{AC}^2}{2} = \frac{2\gamma}{(\gamma+2)^2} \ .$$

Hence the proposition. □