Split-proof probabilistic scheduling

Hervé Moulin
Rice University

March 2005

Abstract

If shortest jobs are served first, splitting a long job into smaller jobs reported under different aliases can reduce the actual wait until completion. If longest jobs are served first, the dual maneuver of merging several jobs under a single reported identity is profitable. Both manipulations can be avoided if the scheduling order is random, and users care only about the expected wait until completion of their job.

The Proportional rule stands out among rules immune to splitting and merging. It draws the job served last with probabilities proportional to size, then repeats among the remaining jobs. Among split-proof scheduling rules constructed in this *recursive* way, it is characterized by either one of the three following properties: an agent with a longer job incurs a longer delay; total expected delay is at most twice optimal delay; the worst expected delay of any single job is at most twice the smallest feasible worst delay. A similar result holds within the natural family of *separable rules.*

Key words: probabilistic scheduling, merging, splitting, proportional rule.

Acknowledgements: Thanks to Hatice Ozsoy, Harris Schlesinger and Jay Sethuraman for useful comments on an earlier version. I am grateful for their hospitality to the University of Auckland, New Zealand, the University of Namur, Belgium, and the New School of Economics in Moscow. This work is supported by the NSF under grant SES- 0414543.

1 Introduction

Processing jobs of different lengths that share a single server raises multiple issues of incentive-compatibility and fairness when agents are impatient. Here we focus on two related maneuvers by which the users may be able to "game" the scheduling discipline, namely splitting one's job into several smaller jobs requested under different aliases (so that the server believes they come from different users), or merging the jobs of several users into one large job presented to the server by a single agent.

Such maneuvers are feasible when the server cannot monitor the identity of the real beneficiaries of the jobs it processes. In large networks such as the Internet, aliases are easy to generate and not easy to track, raising concerns about the vulnerability of Peer-To-Peer systems (Douceur [2002]). In such context, preventing merging and splitting is simply not feasible. Alternatively, the unability to detect the real identity of users may be a design constraint of the system, meant to protect the users' privacy because the jobs involve confidential information. Think of users sharing a medical or financial data base.

In line with most of the scheduling literature (e.g., Lawler et al. [1993]), we assume that partially completed jobs are useless, so that an efficient server processes jobs whole. Splitting is a *non cooperative* manipulation where agent A who needs a job of size x, creates several aliases A',A",.. and requests on their behalf jobs of sizes $a, a', a", ..$ such that $a + a' + a" + \cdot\cdot = x$. If the job completed last among A,A',A",.., is done strictly earlier than job A in the initial problem, the maneuver is profitable. Merging is the *cooperative* move where agents A,B,C,.., who need jobs of sizes $a, b, c, ..$, choose one of them to request a single job of size $x = a + b + c + \cdot\cdot$. The merged agents schedule their true jobs as they please during the time where job x is processed, and can enjoy the benefit of their own job as soon as it is completed. Merging is profitable if the waiting time of at least one merged agent can thus be reduced, without increasing that of any other.

Many *deterministic* service disciplines are either "merge-proof" or "split-proof" but not both. The familiar *Shortest Job First* is by far the most popular scheduling rule, because it minimizes total waiting time accross all users It is merge-proof because a merged job is never served earlier than any of its component jobs, but is badly vulnerable to splitting. Similarly, *Longest Job First*s (that *maximizes* total waiting time) is clearly "split-proof" but not merge-proof. We prove in Section 3 that no deterministic scheduling rule is both merge-proof and split-proof.

Randomizing the service ordering of jobs is the easiest way to restore fairness when efficiency compels to process them whole. Here we submit another advantage of randomization: if we assume that each participant seeks to minimize the expected wait until the completion of his or her job, we can construct probabilistic scheduling rules that are simultaneously split-proof and merge-proof. Our punchline is that among such rules, one rule dubbed the *Proportional* rule stands out uniquely for its fairness and efficiency properties. On the way to a formal statement of this result, we find that Split-proofness eliminates many more natural rules than Merge-proofness.

In order to explain the appeal of the Proportional rule, it is useful to contrast it with the simplest split-proof and merge-proof rule, namely the *Uniform* scheduling rule choosing each ordering of the n participants with equal probability $1/n!$, thus ignoring all differences in job lengths[1]. Given n jobs of sizes $x_1, .., x_n$, the simplest definition of the Proportional rule is recursive: choose first the job scheduled last, with the probability of job i being $\frac{x_i}{x_1+\cdots+x_n}$; then select in the same way the job scheduled next to last among the remaining jobs, and so on. It is easy to see that for any two jobs i, j the probability that j is scheduled before i is $\frac{x_i}{x_i+x_j}$, hence the name of this rule[2].

We now compare these two rules from the point of view of one equity test (Ranking) and two performance indices (the excess delay and the liability), and find that Proportional systematically dominates Uniform.

The longer the job, the larger the externality it creates upon other participants. *Ranking* requires that, as a result, the longer the job the longer its expected wait: : if $x_i < x_j$, then $y_i - x_i \leq y_j - x_j$. This kind of test is familiar to the fair division and cost sharing literature (e.g., Moulin and Sprumont [2003]). The Proportional rule meets Ranking, whereas the Uniform rule imposes a *longer* expected delay to a smaller job ($x_i < x_j \Rightarrow y_i - x_i > y_j - x_j$).

[1] The expected wait of job i with size x_i is $y_i = x_i + \frac{1}{2} \sum_{j \neq i} x_j$, because there is a 50% chance that any other job j precedes job i. When a coalition of users other than i merge their jobs, or when user $j, j \neq i$, splits her job, the sum $\sum_{j \neq i} x_j$ is unaffected, and so is user i's expected wait. By Pareto optimality it follows that neither the splitting agent nor the merging coalition can benefit. See Ozsoy [2005].

[2] When user i splits his job into several smaller "subjobs", one checks that the expected delay until completion of all subjobs is precisely the same as that of the initial job (see comment 1 in Section 10). Merge-proofness of the Proportional rule requires more work: Proposition 2 in Section 6.

Our two performance indices are inspired by two conflicting normative goals playing a central role in the management of queues: to minimize the total expected waiting time of users, which leads to the Shortest Job First discipline; and to equalize *slowdown* accross users, namely the ratio of waiting time to job size. The former is a utilitarian concern for efficiency, whereas the latter is the idea of equalizing the (relative) delay externality experienced by each user. See Demers et al. [1990], Bender et al. [1998], Friedman and henderson [2003], Wierman and Harchol-Balter [2003], and references in Moulin [2005].

We define in Section 7 the (relative) *excess delay* of a rule as the worst ratio of actual to optimal expected delay, where the maximizatio bears on all scheduling problems. The excess delay of the Proportional rule is 2, whereas it is unbounded for the Uniform rule. We define the *liability* of a rule as the worst expected slowdown of that job, when other jobs are arbitrarily large. The liability of the Uniform rule is, again, unbounded, because the expected wait of a job goes to infinity when other jobs are infinitely larger. Under the Proportional rule, the liability of any job is n, the number of other jobs in the queue. This is the same liability as under Shortest Job First, and about twice the optimal liability.

Our two main results, Theorem 1 and 2 in Sections 8 and 9, assume one of two invariance properties satisfied by the Proportional and many other scheduling rules. A rule is *separable* if the relative ranking of any subset of jobs is independent of other jobs' sizes. It is *recursive* if it is generated by n independent successive draws, choosing first the agent served *last*, then the agent served next to last in the reduced problem with $n-1$ agents, and so on. The Proportional and Uniform rules are both separable and recursive, and the combination of these two properties produces the rich class of *quasi-proportional* rules (Section 5 and Proposition 3).

Among split-proof scheduling rules that are either separable (Theorem 1), or recursive (Theorem 2), the Proportional one is characterized by any one of the following three properties: Ranking; the excess delay is at most 2; the liability of any job is at most n, the number of users.

2 Related literature

This paper is inspired by three essentially independent streams of micro-economic literature: the first one applies the mechanism design approach to queuing and scheduling, the second one discusses splitting and merging maneuvers in various fair division problems, and the third one studies the random assignment of homogeneous indivisible commodities.

We start with the research on scheduling and queuing (under the assumption that only completed jobs matter) when monetary transfers are feasible. That literature typically assumes quasi-linear preferences in money, and linear waiting costs. One idea is to propose fair monetary compensations by applying solutions concepts from cooperative game theory: the Shapley value is increasingly popular for the scheduling model (Curiel et al. [1989],[2002], Klijn and Sanchez [2002], Maniquet [2003], Chun [2004 a,b]); in the queuing problem, Haviv and Ritov [1998], Haviv [2001] apply the Aumann-Shapley pricing rule to various service disciplines. A different, and more developed, theme is to design cash transfers ensuring the truthful revelation of waiting costs. This leads to interesting (in particular, budget-balanced) Vickrey-Clarke-Groves mechanisms, or generalizations thereof: Dolan [1978], and more recently Suijs [1996], Mitra [2001,2002]. Closer to our model, Hain and Mitra [2001], Kittsteiner and Moldovanu[2003 a,b] consider the truthful revelation of job sizes, assuming that agents cannot be punished ex post for misreporting. In our model the system manager can punish users who minimize ex ante the length of their job, and they have no incentive to exagerate the size of their job either.

The companion paper Moulin [2004] discusses merging and splitting maneuvers in the quasi-linear scheduling model just discussed. The results there are less encouraging than in the probabilistic model of this paper, in the sense that no efficient mechanism can be both split-proof and merge-proof. Just like here, Merge-proofness proves much easier to meet than Split-proofness. Unlike here, the normative consequences of split-proofness are dire: it is incompatible with either Ranking, a finite individual liability, or the monotonicity of the net waiting cost in one's job size.

The follow up paper Ozsoy [2005] explores the robustness of our results to partial transfers of jobs and to a coalitional version of the splitting maneuver. More on this and the companion paper in Section 10 (comment 4).

We turn to the discussion of splitting and merging in the fair division literature. The earliest contribution deal with the rationing problem (Banker [1981], Moulin [1987], De Frutos [1999], Ju [2003]) where the proportional rule is the only one immune against such maneuvers. In the quasi-linear social choice problem, the same property leads to the egalitarian division of surplus (Moulin [1985], Chun[2000]), and in the cost sharing problem with variable demands, to the Aumann-Shapley rule (Sprumont [2004]). Most of these results are surveyed in Ju et al. [2003].

Finally the problem of randomly assigning homogeneous indivisible units (Moulin [2002], Moulin and Stong [2002,2003]) can be interpreted as the couterpart of the current model under processor sharing. Each user requests a job of size x_i ,where x_i is an integer, and the server processes successive unit jobs, typically alternating between the different users (instead of serving jobs whole). This is not inefficient if partially completed jobs bring positive utility. Another Proportional rule emerges in that model, in which x_i balls of color i are thrown in an urn, and the server empties the urn without replacement to determine the order in which the $\sum x_i$ unit jobs are processed. In Section 5 we show that the Proportional rule here can also be generated by throwing x_i balls of color i in an urn, then serving the jobs (whole) in the order in which each color "vanishes" from the urn. The key axioms in the processor sharing model are different, yet the invulnerability to transfers plays a role (see Theorem 1 in Moulin and Stong [2003]).

3 The model

The infinite set $\mathcal{N}$ of potential users is fixed throughout. A *scheduling problem* is a pair (N, x) where N is a finite subset of $\mathcal{N}$, and x is a profile of non-negative job sizes, $x_i \in \mathbb{R}_+$ for all $i \in N$. The set of orderings of N is $\Phi(N)$, with generic element $\sigma : \sigma(i) < \sigma(j)$ means that job i is served/scheduled before job j.

We prove first our earlier claim that without randomization, all rules are vulnerable to at least one of merging and splitting. The simple argument does not require a formal definition of these properties (given in Section 6). A *deterministic* scheduling rule selects an ordering $\sigma \in \Phi(N)$ for any problem (N, x). Assume without loss of generality that when agents A,B,C each have a job of size 2, the rule orders them alphabetically. If now A,B show up with jobs $(a_1, b_1) = (2, 4)$, A must be served first, otherwise B and C both reduce their wait by merging in the initial three person problem. If B,C show up with $(b_2, c_2) = (4, 2)$, B must be served first otherwise B shortens his wait by splitting into $(a', b') = (2, 2)$. Next consider the problem $(a, b, c) = (1, 4, 1)$ .If B is served last, then at (b_2, c_2), C splits advantageously to $(a", c") = (1, 1)$. If B is served second at (a, b, c), and A,C merge into $a_1 = 2$, the one served first benefits strictly whereas the other suffers no harm; if B is served first at

(a,b,c), both A,C benefit by merging. This contradiction proves our claim.

From now on, we allow the server to schedule jobs randomly. A random ordering is a probability distribution p on $\Phi(N)$; the set of such distributions is denoted $\Delta[\Phi(N)]$. We assume that each user i only cares to minimize the expected completion time y_i of her own job. Given the random ordering p and a problem (N,x), this "disutility" is

$$y_i = x_i + \sum_{j \in N \diagdown i} prob\{\sigma(j) < \sigma(i) | x\} \cdot x_j \tag{1}$$

where $prob\{\sigma(j) < \sigma(i)|x\} = \sum_{\sigma:\sigma(j)<\sigma(i)} p_\sigma(N,x)$.

Conversely, our first result describes those profiles $y \in \mathbb{R}_+^N$ *feasible* at a given problem (N,x), namely such that (1) holds for some lottery $p \in \Delta[\Phi(N)]$. Define for all $x \in \mathbb{R}_+^N$ and all $S \subseteq N$, the function $v(S,x) = \sum_S x_i^2 + \sum_{S(2)} x_i \cdot x_j$, where $S(2)$ is the set (with cardinality $\frac{|S|\cdot(|S|-1)}{2}$) of non ordered pairs from S. Note that v is supermodular with respect to S.

Lemma 1

i) The profile $y \in \mathbb{R}_+^N$ is feasible at (N,x), $x \in \mathbb{R}_+^N$, if and only if for all i $\{x_i = 0 \Longrightarrow 0 \leq y_i \leq \sum_N x_j\}$, and moreover $x \cdot y$ belongs to the core of the game $(N, v(\cdot,x))$, i.e.,

$$\sum_N x_i \cdot y_i = v(N,x) \text{ and } \sum_S x_i \cdot y_i \geq v(S,x) \text{ for all } S \subseteq N.$$

ii) The profile $y \in \mathbb{R}_+^N$ is efficient at (N,x), $x \in \mathbb{R}_+^N$, if and only if for all i $\{x_i = 0 \Longrightarrow y_i = 0\}$, and moreover $x \cdot y$ belongs to the core of the game $(N, v(\cdot,x))$. We denote by $F(N,x)$ the set of efficient profiles y at (N,x).

Statement i) is proven in Queyrane [1993]. For the sake of completeness we provide a (different) proof in the Appendix.

Lemma 1 implies that if all jobs are positive, *all* random orderings are efficient, because the weighted sum $\sum_N x_i \cdot y_i$ is independent of the choice of p. When some jobs are null, efficiency only requires to schedule all null jobs before any non-null job.

From now on we write $x \geq 0$ for $x \in \mathbb{R}_+^N$, and $x \gg 0$ whenever all coordinates are positive.

Definition 1

An efficient scheduling rule is a mapping ρ associating to each problem $(N,x), x \gg 0$, a random ordering $\rho(N,x) = p \in \Delta[\Phi(N)]$.

An efficient scheduling method is a mapping μ associating to each problem $(N,x), x \geq 0$, an efficient profile $\mu(N,x) = y \in F(N,x)$.

We only need to define efficient scheduling *rules* over strictly positive profiles x, because the relative ordering of null jobs is irrelevant, as long as they are served before all non null jobs. On the other hand, an efficient scheduling *method* is defined for all profiles, because the wait of null jobs is unambiguous; moreover it will prove convenient in Section 6 to define a method for all profiles in $\mathbb{R}_+^N$.

In the sequel when we speak of a method or a rule, we always mean that it is efficient.

4 Split-proofness and Separability

We start by a three users example. In the problem (N,x) where $N = \{1,2,3\}$, assume that user 3 reports as two agents $3,4$ and *splits* her (true) job of size x_3 into two jobs of sizes x_{*3}, x_{*4}, with $x_{*3} + x_{*4} = x_3$. After splitting, user 3 must wait until both "sub-jobs" are completed hence the delay she experiences from job $i, i = 1,2$ is determined by the probability that this job is scheduled before at least one of the sub-jobs. Setting $x_* = (x_1, x_2, x_{*3}, x_{*4})$ and using the notations of equation (1), we see that the split benefits agent 3 iff

$$\sum_{i=1,2} prob\{\sigma(i) < \max(\sigma(3),\sigma(4))|x_*\} \cdot x_i < \sum_{i=1,2} prob\{\sigma(i) < \sigma(3)|x\} \cdot x_i$$

The following notations are used in the general definition of Split-proofness. Given a problem $(N,x), x \geq 0$, a coalition T such that $T \cap N = \varnothing$, and an agent $i_* \in N$, the splitting of i_*into $i_* \cup T$ creates a new problem $(N_*, x_*), x_* \geq 0$, where $N_* = N \cup T$, $x_{i_*} = \sum_{i_* \cup T} (x_*)_i$, $(x_*)_j = x_j$ for $j \in N \diagdown i_*$. Note that we can always assume that the initial job is positive, $x_{i_*} > 0$, because a null job has no incentive to deviate; on the other hand we allow some coordinates of x_* to be null, because by splitting x_{i_*} into x_* such that $(x_*)_{i_*} = 0, (x_*)_j = x_{i_*}$, agent i_* effectively assumes a new identity as agent j, and such maneuvers are both realistic and important in the proof of Lemma 2 below.

Definition 2

Fix a scheduling rule ρ, a problem $(N,x), x \geq 0$, and T, i_* as above. We say that ρ is split-proof at (N,x) with respect to T, i_* if

$$y_{i_*}(N,x) \leq x_{i_*} + \sum_{j \in N \diagdown i_*} prob\{\sigma(j) < \max_{i_* \cup T} \sigma(i) | x_*\} \cdot x_j \tag{2}$$

We say that ρ is split-proof if it is split-proof for all $(N,x), T, i_*$.

The right-hand side of the above inequality is the (true) expected wait of agent i_* after the split.

We introduce next a key invariance property of scheduling rules, under which Split-proofness takes a much simpler form. Moreover all rules discussed in the paper meet this property. Given $N, S, S \subset N$, and $\sigma \in \Phi(N)$, we write $\sigma[S] \in \Phi(S)$ for the restriction of σ to S.

Definition 3

The scheduling rule ρ is ***separable*** *if for all $N, S, S \subset N$, $x \gg 0$, the (random) ordering of the jobs in S is independent of the jobs outside S:*

$$\text{for all } \sigma^* \in \Phi(S) : \sum_{\sigma \in \Phi(N): \sigma[S] = \sigma^*} p_\sigma(N,x) \text{ is independent of } N \diagdown S \text{ and of } x_{[N \diagdown S]}$$

For a separable rule ρ, we speak of the probability that a job of size x_j for user j precedes a job of size x_i for user i, without specifying either the rest of the participants or the size of their jobs: we write this probability $\theta^{i,j}(x_i, x_j)$, so that the method μ defined by ρ takes the form

$$y_i(N,x) = x_i + \sum_{j \in N \diagdown i} \theta^{i,j}(x_i, x_j) \cdot x_j \tag{3}$$

Lemma 2

The separable scheduling rule ρ is split-proof ***if and only if*** *the corresponding method is anonymous ($\theta^{i,j} = \theta$ is independent of i, j), and moreover for all $S, i, i \notin S$, and all positive numbers $b, a_j, j \in S$,*

$$\theta(b, \sum_S a_j) \geq prob\{\max_{j \in S} \sigma(j) < \sigma(i) | (b, a_j)\} \tag{4}$$

Proof

Statement "if". Fix (N,x) and T, i_* as in the premises of Definition 2. By assumption $y_{i_*}(N,x) - x_{i_*} = \sum_{N \diagdown i_*} \theta(x_{i_*}, x_j) \cdot x_j$. Next the term $prob\{\sigma(j) < \max_{i_* \cup T} \sigma(i) | x_*\}$ depends only upon $(x_*)_{[j \cup i_* \cup T]}$ by Separability of ρ. Thus inequality (2) follows from $\theta(x_{i_*}, x_j) \leq prob\{\sigma(j) < \max_{i_* \cup T} \sigma(i) | x_*\}$, which is precisely (4) for $S = i_* \cup T$.

Statement "only if". We show first that $\theta^{i,j}$ is independent of i,j. Fix $a,b>0$ and consider in problem $N=\{1,2\}, x=(a,b)$, the split of agent 1 into agents $1,3$ with $(x_*)_1=0,(x_*)_3=a$. In the split problem, agent 1 is scheduled first by efficiency, so the probability that agent 2 is last is just $\theta^{2,3}(b,a)$. Thus split-proofness implies $\theta^{2,1}(b,a)\geq\theta^{2,3}(b,a)$. Exchanging the roles of $1,3$ gives $\theta^{2,1}=\theta^{2,3}$. By $\theta^{i,j}(b,a)+\theta^{j,i}(a,b)=1$ we get $\theta^{1,2}=\theta^{3,2}$. Anonymity follows.

Next fix $S,i,i\notin S$,and b,a_j as in the premises of (4) and choose an agent $1\in S$. In the problem $N=\{1,i\}, x=(a_S,b)$ a split by 1 to S with $(x_*)_j=a_j$ for all $j\in S$ cannot benefit agent 1: this gives inequality (4), and completes the proof.■

Examples of scheduling rules illustrating Definitions 1,2 and Lemma 2 are the subject of the next Section.

5 Parametric rules

We construct a large family of separable scheduling rules by choosing for each $a>0$ a cumulative distribution function F_a on $[0,+\infty[$ with no mass at 0. Thus F_a is any non decreasing and right-continuous function on $[0,+\infty[$ such that $F_a(0)=0$ and $F_a(\infty)=1$.

Definition 4

Given a scheduling problem $(N,x), x\gg 0$, the ***parametric*** *rule associated with the family $\{F_a, a>0\}$ picks $|N|$ independent random variables Z_{x_i}, one for each $i\in N$ with cdf F_{x_i}, and orders jobs according to the realization of these variables, breaking ties randomly with uniform probability.*

Separability is clear. Parametric rules are also *anonymous*, namely for all N, the mapping $x\rightarrow p(N,x)$ is symmetric in all variables for $x\gg 0$. In particular $\theta^{i,j}=\theta$ is independent of i,j:

$$\theta(a,b)=prob\{Z_a<Z_b\}+\frac{1}{2}prob\{Z_a=Z_b\} \text{ for all } a,b. \tag{5}$$

Three benchmark parametric rules are: Shortest Jobs First, for which F_a is concentrated at a so that $\theta(a,b)=1$ if $b<a$; Longest Jobs First, where F_a is concentrated at $\frac{1}{a}$ and $\theta(a,b)=1$ if $a<b$; and the Uniform rule, for which F_a does not depend on a, so that $\theta(a,b)=\frac{1}{2}$ for all a,b.

Next we define a rich subset of parametric rules playing a key role below.

Definition 5

Let w be a function on $\mathbb{R}_+$, such that $w(0) = 0$ and $a > 0 \Rightarrow w(a) > 0$. Given a scheduling problem $(N, x), x \gg 0$, the $w-$quasi-proportional rule is the parametric rule associated with $F_a(z) = \min\{z^{w(a)}, 1\}$ for all $z \geq 0$.

The Uniform rule is quasi-proportional with $w(a) = 1$ for all $a > 0$. The Shortest (resp. Longest) Job First rule is the limit of the quasi-proportional rules $w(a) = a^\alpha$, when α goes to $+\infty$ (resp. to $-\infty$). The Proportional rule, to which our main results are devoted, corresponds to $w(a) = a$. Another remarkable quasi-proportional rule is the *Quadratic* one, for which $w(a) = a^2$ (see Lemma 5 in Section 7.2). Our terminology becomes clear when we compute the method corresponding to a quasi-proportional rule.

Lemma 3

Fix N, an ordering $\sigma \in \Theta(N)$, and a profile $x \gg 0$. Set $w_k = w(x_{\sigma^{-1}(k)})$, so w_1 (resp. w_n) is the weight of the job scheduled first (resp. last). The probability of σ under the $w-$quasi-proportional rule is

$$p_\sigma = \frac{w_n}{\sum_1^n w_k} \cdot \frac{w_{n-1}}{\sum_1^{n-1} w_k} \cdot \ldots \cdot \frac{w_2}{w_1 + w_2} \tag{6}$$

In particular $\theta(a, b) = \frac{w(a)}{w(b)+w(a)}$, and

$$y_i = x_i + \sum_{N \setminus i} \frac{w(x_i)}{w(x_j) + w(x_i)} \cdot x_j \text{ for all } i \in N$$

Proof.

Because the distribution of each variable Z_i is non atomic, and these variables are independent, the probability of a tie $Z_i = Z_j$ is null. Hence, with the notations above, the probability that ordering σ is selected is that of the event $\{Z_1 \leq Z_2 \leq \cdot\cdot \leq Z_n\}$, where Z_i follows a beta distribution with parameter w_i. This is precisely (6), as follows from a simple computation that we reproduce for the case $n = 3$:

$$\int_0^1 w_3 \cdot z^{w_3-1}\{\int_0^z w_2 \cdot t^{w_2-1}\{\int_0^t w_1 \cdot s^{w_1-1} ds\}dt\}dz =$$

$$\int_0^1 w_3 \cdot z^{w_3-1}\{\int_0^z w_2 \cdot t^{w_1+w_2-1} dt\}dz = \frac{w_2}{w_1 + w_2} \int_0^1 w_3 \cdot z^{w_1+w_2+w_3} dz$$

■

There is an alternative, perhaps more intuitive definition of quasi-proportional rules, for those problems (N, x) where the numbers $w(x_i)$ are all rationals, namely $w(x_i) = \frac{b_i}{d}$ for some integers b_i, d. For each agent i put b_i balls of color i in an urn and empty the urn by successive draws with uniform probability and without replacement; schedule the jobs in the order in which each color vanishes in the urn (when the last ball of this colour is drawn). This generates precisely the random ordering in Lemma 3[3].

Applying Lemma 2 to parametric rules yields a simple sufficient condition for Split-proofness in that family. Fix a parametric rule $\{F_a, a > 0\}$ and assume for a moment that all distributions F_a are atomless (i.e., F_a is continuous). As ties occur with probability zero, inequality (4) amounts to

$$\begin{aligned} prob\{Z_{a_S} &\leq Z_b\} \geq prob\{\max_S Z_{a_j} \leq Z_b\} \Longleftrightarrow \\ \int_0^\infty F_{a_S}(z) \cdot dF_b(z) &\geq \int_0^\infty \Pi_{j\in S} F_{a_j}(z) \cdot dF_b(z) \end{aligned}$$

which holds true if $F_{a_S}(z) \geq \Pi_{j\in S} F_{a_j}(z)$. The latter inequality is in fact sufficient for Split-proofness, whether or not the corresponding distributions have atoms.

Proposition 1

If the cdfs $\{F_a, a > 0\}$ satisfy $F_a \cdot F_b \leq F_{a+b}$ for all $a, b > 0$, the corresponding parametric rule is split-proof.

Corollary

The $w-$quasi-proportional rule is split-proof if and only if w is subadditive.

The proof of Proposition 1 for distributions with atoms is in the Appendix.

Proof of the Corollary

For the $w-$quasi-proportional rule, the cdfs F_a are atomless. Inequality $F_a(z) \cdot F_b(z) \leq F_{a+b}(z)$ is always true if $z \geq 1$, and amounts to $w(a) + w(b) \geq w(a+b)$ if $z < 1$. This proves the "if" statement. Conversely, suppose the $w-$rule is split-proof. By Separability, inequality (4) must be true. As $\Pi_{j\in S} F_{a_j}(z) = z^{\sum_S w(a_j)}$, the right-hand side in (4) equals $\frac{w(b)}{w(b)+\sum_S w(a_j)}$, therefore this inequality reads

$$\theta(b, \sum_S a_j) = \frac{w(b)}{w(b) + w(\sum_S a_j)} \geq \frac{w(b)}{w(b) + \sum_S w(a_j)}$$

[3] I am grateful to R.J. Aumann for pointing out this interpretation.

implying the subadditivity of w.■

A natural choice of the weight function w is the power function $w(a) = a^\alpha$ for some real number α[4]. This includes all the rules discussed after Definition 5. For these rules, the Corollary says that split-proofness is equivalent to $\alpha \leq 1$, hence includes the Uniform, Proportional, and Longest Job First rules.

6 Merge-proofness and demand monotonicity

We start with a 5 agents example. In the problem (N, x), where $N = \{1, 2, 3, 4, 5\}$, asssume that the coalition $\{1, 2, 3\}$ *merges* its jobs. This means that they report as a single agent of the coalition, say 1^*, a job of length $x^*_{1^*} = x_1 + x_2 + x_3$, and choose freely a scheduling order, possibly a randomized one, of their "true" jobs in the time interval chosen by the server to process the merged job. Write $y^*_{1^*}$ for the expected wait of the merged job of size $x^*_{1^*}$, so that $y^*_{1^*} - x^*_{1^*}$ is the expected delay of job $x^*_{1^*}$ generated by jobs 4 and 5. This delay is borne by all three agents upon merging. The other part of their true wait comes from the processing order of jobs $1, 2$, and 3: Lemma 1 describes the range $F(\{1, 2, 3\}, (x_1, x_2, x_3))$ of the corresponding profiles of expected waits. Summing up we see that the profile $\widetilde{y} = (\widetilde{y}_1, \widetilde{y}_2, \widetilde{y}_3)$ of expected waits is feasible for $\{1, 2, 3\}$ after merging if and only if $\widetilde{y} \in (y^*_{1^*} - x^*_{1^*}) \cdot (1, 1, 1) + F(\{1, 2, 3\}, (x_1, x_2, x_3))$. If y is the profile of waits before merging, this move is profitable if and only if there exists $\widetilde{y}$ such that $\widetilde{y} \leq y$ with at least one strict inequality.

Notice that the description of successful merges only uses the scheduling *method*, not the actual scheduling *rule* from which it is derived. This simplifies greatly the analysis of Merge-proofness. Now some notations. Given a problem (N, x), a proper subset (coalition) T of N, and an agent $i^* \in T$, the merger of T into i^* creates a new problem (N^*, x^*), where $N^* = (N \diagdown T) \cup \{i^*\}$, $x^*_{i^*} = \sum_T x_i$, $x^*_j = x_j$ for $j \in N \diagdown T$. We write e for the vector in $\mathbb{R}^N$ with all coordinates equal to 1. Finally the set $F(T, x_{[T]})$ (Lemma 1) consists of the profiles of expected waits feasible for coalition T when it is scheduled before $N \diagdown T$.

[4]These rules are *Scale Invariant*, namely $\rho(N, x) = \rho(N, a \cdot x)$ for all $a > 0$, and this property is characterictic within the quasi-proportional family. See Ozsoy [2005].

Definition 6
Given a scheduling method μ, a problem $(N, x), x \geq 0$, and T, i^* as above, we set $y = \mu(N, x)$ (before merging) and $y^* = \mu(N^*, x^*)$ for the (reported) waits after merging. We say that μ is vulnerable to merging by T, i^* at problem (N, x) if there exists a vector $\widetilde{y}_{[T]} \in \mathbb{R}^T$ such that

$$\widetilde{y}_{[T]} \in (y^*_{i^*} - \sum_T x_i) \cdot e_{[T]} + F(T, x_{[T]}); \, \widetilde{y}_{[T]} \leq y_{[T]} \text{ and } \widetilde{y}_{[T]} \neq y_{[T]} \tag{7}$$

*We say that the method μ is **merge-proof** if it is not vulnerable to merging at any problem by any coalition. We say that the scheduling rule ρ is merge-proof if the associated method is.*

Merge-proofness is easy to achieve because it is implied by the combination of a separability property less demanding than Definition 3, and of a mild monotonicity property stating that the expected delay of a job is non decreasing in its size.

Definition 7
*The scheduling method μ is **demand monotonic** if for all N, and $i \in N$, $y_i(N, x) - x_i$ is non decreasing in x_i. The scheduling rule ρ is demand monotonic if the associated method is.*

Increasing the size of my job augments the delay externality, and it is only fair that my share of this externality should not decrease.

It is easy to check that the $w-$quasi-proportional method is demand monotonic if and only if w is non decreasing. Longest Job First is not demand monotonic, but the Uniform, Proportional and Shortest Job first methods are. More generally, the parametric rule $\{F_a, a > 0\}$ is demand monotonic if $a \rightarrow Z_a$ is stochastic-dominance monotonic , i.e., $\{a \leq b\} \Rightarrow \{F_b(z) \leq F_a(z)$ for all $z\}$. The proof of this fact is in the Appendix, as Step 2 in the proof of Proposition 1.

Our next result applies to all scheduling methods derived from separable scheduling rules, and more generally to the following class of methods.

Definition 8
*The scheduling method μ is **separable** if there exists for all $i, j \in \mathcal{N}$* a function $\theta^{i,j}$ *from $\mathbb{R}^2_+$ into* $[0, 1]$, *such that for all $a, b \geq 0$*

$$\theta^{i,j}(a, b) + \theta^{j,i}(b, a) = 1, \text{ and } a > 0 \Rightarrow \theta^{i,j}(a, 0) = 1. \tag{8}$$

and such that equation (3) holds for all $N, x, x \geq 0$.

It follows from Lemma 1 that for any choice of the functions $\theta^{i,j}$ meeting (8), equation (3) defines a feasible profile of expected waits. Indeed $\theta^{i,j} \geq 0 \Rightarrow \sum_S x_i \cdot y_i = v(S,x) + \sum_{i \in S, j \in N \diagdown S} \theta^{i,j}(x_i, x_j) \cdot x_i \cdot x_j \geq v(S,x)$. Yet this does not imply the existence of a *separable* scheduling rule ρ implementing the method $\{\theta^{i,j}\}$[5].

Proposition 2

An anonymous, demand monotonic, and separable scheduling method is merge-proof. For instance the $w-$quasi-proportional rule is merge-proof if w is non decreasing.

Proof

We show first a preliminary result. The scheduling method μ is merge-proof if for every $(N,x), x \geq 0, T, i^*$ as in Definition 6, with $y = \mu(N,x)$ and $y^* = \mu(N^*, x^*)$, we have

$$\sum_T x_i \cdot y_i + \sum_{T(2)} x_i \cdot x_j \leq (\sum_T x_i) \cdot y^*_{i^*} \tag{9}$$

(recall that $T(2)$ is the set of non ordered pairs in T). Consider a vector $\widetilde{y}_{[T]}$ in the set defined in property (7). Using the notation $\sum_T x_i = x_T$, we compute from Lemma 1

$$\sum_T x_i \cdot \widetilde{y}_i = x_T \cdot (y^*_{i^*} - x_T) + v(T,x) = x_T \cdot y^*_{i^*} - \sum_{T(2)} x_i \cdot x_j$$

Thus (9) is equivalent to $\sum_T x_i \cdot y_i \leq \sum_T x_i \cdot \widetilde{y}_i$, and implies that $\{\widetilde{y}_{[T]} \leq y_{[T]}$ and $\widetilde{y}_{[T]} \neq y_{[T]}\}$ is impossible.

We pick now an anonymous, demand monotonic and separable method defined by the function θ. Demand Monotonicity implies that $\theta(a,b)$ is non decreasing in a, because $y_i(\{i,j\}, x) - x_i = \theta^{i,j}(x_i, x_j) \cdot x_j$ is non decreasing in x_i. We prove (9) for any N, x, T, i^* by developing this inequality with the help of (3). The term $\sum_T x_i \cdot y_i$ in the LHS is

$$\sum_T x_i \cdot y_i = \sum_T x_i \cdot (x_i + \sum_{N \diagdown i} \theta(x_i, x_j) \cdot x_j) = \sum_T x_i^2 + \sum_{T(2)} x_i \cdot x_j + \sum_{i \in T}^{j \in N \diagdown T} \theta(x_i, x_j) \cdot x_i \cdot x_j$$

---

[5] For a given matrix $[\theta^{i,j}(x_i, x_j)] = [t_{i,j}]$ with non negative entries such that $t_{i,j} + t_{j,i} = 1$, the existence of a lottery $p \in \Delta(\Phi(N))$ such that $prob\{\sigma(j) < \sigma(i)\} = t_{i,j}$ for all i, j is not guaranteed. On this difficult existence problem, see Fishburn [1992].

Thus the LHS of (9) is $x_T^2 + \sum_{i \in T}^{j \in N \setminus T} \theta(x_i, x_j) \cdot x_i \cdot x_j$. The RHS is $x_T \cdot [x_T + \sum_{N \setminus T} \theta(x_T, x_j) \cdot x_j]$, therefore (9) amounts to

$$\sum_{j \in N \setminus T} x_j \cdot [\sum_{i \in T} \theta(x_i, x_j) \cdot x_i] \leq \sum_{j \in N \setminus T} x_j \cdot [\theta(x_T, x_j) \cdot x_T].$$

This holds if $a \to \theta(a,b) \cdot a$ is superadditive for $b > 0$ (the inequality above always holds if $x_j = 0$). The latter is true if $\theta(a,b)$ is non decreasing in a and $\theta \geq 0$.■

Remark 1 *A plausible statement is that an anonymous separable method is merge-proof if and only if it is demand monotonic. This is not true, however. We saw in the above proof that the anonymous separable method defined by* θ *is merge-proof if* $a \to \theta(a,b) \cdot a$ *is superadditive. Hence the* $w-$*quasi-proportional rule is merge-proof if* $a \to a \cdot w(a)$ *is superadditive*[6]*. We can choose such a function* w *that is not everywhere increasing, and the corresponding quasi-proportional rule is not demand monotonic.*

Combining Propositions 1 and 2, we see that the $w-$quasi-proportional rule is split-proof and merge-proof if w is non decreasing and subadditive. For instance among the weight functions $w(a) = a^\alpha$, these two properties hold in the interval $0 \leq \alpha \leq 1$, the two extreme points of which are the Uniform and Proportional rules.

But we stress that there are many other merge-proof and split-proof parametric scheduling rules. It suffices that the cdfs satisfy $F_a \cdot F_b \leq F_{a+b} \leq F_a$ for all $a, b > 0$. Two examples are the cdfs $F_a(z) = \min\{\frac{(1+a)z}{a+z}, 1\}$, and $F_a(z) = e^{-\frac{a^2}{z(a+z)}}$. Moreover, Merge-proofness and Split-proofness are stable by convex combinations of (lotteries over) rules. As the parametric rules are *not* stable by convex combinations, we can thus generate many more rules both merge-proof and split-proof.

7 Ranking, Excess Delay and Liability

7.1 The Ranking axiom

Definition 9

[6]Indeed this implies for all a, a' : $w(a + a') \geq \frac{a}{a+a'} w(a) + \frac{a'}{a+a'} w(a')$; applying the concave and increasing function $f(z) = \frac{z}{z+w(b)}$ to both sides of this inequality gives the superadditivity of $a \cdot \theta(a,b) = \frac{a \cdot w(a)}{w(a)+w(b)}$ as claimed.

The scheduling method μ meets Ranking if for all $N, x, x \geq 0$, and $i, j \in N$

$$x_i \leq x_j \Rightarrow y_i(N,x) - x_i \leq y_j(N,x) - x_j$$

The interpretation of Ranking is similar to that of Demand Monotonicity, with the difference that it makes interpersonal comparison of delay shares. Another similarity is that both axioms are easy to express for anonymous separable methods, hence for all parametric rules.

Lemma 4

The anonymous separable method defined by θ ($\theta^{i,j} = \theta$ does not depend on i, j) meets Ranking if and only if $\theta(a,b)$ is non decreasing in a for all $b > 0$ and

$$\text{for all } i,j, \text{ all } a,b > 0 : a \leq b \Longrightarrow \theta(a,b) \leq \frac{a}{a+b}.$$

Proof

The "if" statement follows by developing the inequality $y_i - x_i \leq y_j - x_j$ with the help of (3). To prove "only if", apply first Ranking in the two person problem $N = \{1,2\}, x = (a,b)$, to get the upper bound on $\theta(a,b)$. Next show by contradiction that $\theta(a,b)$ is non decreasing in a. Suppose for some $a, a', b > 0$, we have $a < a'$ and $\theta(a,b) > \theta(a',b)$. Consider the problem $N = \{1,2,..,n\}, x = (a,a',b,..,b)$: for n large enough we get $y_1 - x_1 > y_2 - x_2$, in contradiction of Ranking.■

Thus Ranking is strictly more demanding than Demand Monotonicity for separable methods. For instance the $w-$quasi-proportional method is demand monotonic iff w is non decreasing, whereas Ranking requires that $\frac{w(a)}{a}$ be non decreasing. The latter property implies that w is superadditive. Compare with the Corollary to Proposition 1: split-proofness is equivalent to subadditivity of w. If $w(a) = a^\alpha$, Ranking amounts to $\alpha \geq 1$ and Split-proofness to $\alpha \leq 1$. The results of the next two Sections show that this tension occurs in the much larger classes of separable or recursive rules.

7.2 Excess delay and limited liability

We refer to the Introduction for the motivation of our two measures of welfare performance of a scheduling rule.

We write $\nabla(N,x)$ for the smallest feasible total delay at problem (N,x), obtained by scheduling shortest jobs first. Thus $\nabla(N,x) = \sum_{N(2)} \min\{x_i, x_j\}$.

Definition 10

a) The (relative) ***excess delay*** *of a scheduling method μ is $\delta(n) = \sup_x \frac{y_N(x) - x_N}{\nabla(N,x)}$, where n is an integer and the supremum is over all subsets N of $\mathcal{N}$ with n agents and all profiles $x \geq 0$.*
b) The ***liability*** *of a scheduling method μ is the function $\lambda(n) = \sup_{i,x} \frac{y_i(x)}{x_i}$, where n is an integer and the supremum is over all subsets N of $\mathcal{N}$ with n agents, all $i \in N$, and all profiles $x \geq 0$.*

The liability of Shortest Job First is $\lambda(n) = n$, because the worst that could happen to job i is that all other jobs are barely shorter than x_i. But the optimal (smallest feasible) liability is about twice smaller.

Lemma 5
i) The Quadratic method ($w(z) = z^2$) has the liability $\lambda^*(n) = \frac{n+1}{2}$.
ii) The liability of ***any*** *method μ is not smaller than λ^*: $\lambda(n) \geq \lambda^*(n)$. Thus λ^* is the optimal liability.*

Proof
For an arbitrary method μ, consider a problem (N, x) where $x_i = a$ for all $i \in N$. Feasibility of $y = \mu(N, x)$ (Lemma 1) implies $a \cdot (\sum_N y_i) = v(N, x) = \frac{n(n+1)}{2} \cdot a^2$. Therefore $\max_i y_i \geq \frac{(n+1)}{2} \cdot a$, implying $\frac{y_i(a, x_{-i})}{a} \geq \frac{n+1}{2}$ and statement *ii*. For statement *i*, consider the anonymous separable method μ associated with θ. For any N, i, and $a > 0$, compute i's worst expected wait from equation (3):

$$\sup_{x_{-i}} y_i(a, x_{-i}) = a + \sum_{j \in N \setminus i} \sup_{x_j} \theta(a, x_j) \cdot x_j = a + (n-1) \sup_{b>0} \theta(a, b) \cdot b$$

Thus μ has liability λ^* if and only if $\sup_{b>0} \theta(a, b) \cdot b = \frac{a}{2}$. The latter equality checks easily for the quadratic method.■

Note that there are other methods with minimal liability λ^*, for instance the Serial method defined in Section 9. But among quasi-proportional rules, only the Quadratic one delivers λ^*[7].

The above proof gives the simple form of the liability λ for anonymous separable methods:

$$\lambda(n) = 1 + (n-1) \sup_{a,b>0} \theta(a, b) \cdot \frac{b}{a} \tag{10}$$

[7] The $w-$method guarantees λ^* iff $\frac{w(a)}{w(a)+w(b)} \leq \frac{a}{2b} \Leftrightarrow \frac{w(a)}{w(b)} + 1 \geq 2\frac{a}{b}$ for all $a, b \geq 0$. Letting a go to b, then b go to a in this inequality shows that w is differentiable and $\dot{w}(a) = 2\frac{w(a)}{a}$.

For such methods, the index of excess delay δ takes also a similar simple form.

Lemma 6

Consider an anonymous separable method μ defined by θ. *Then* $\delta(n) = \delta$ *is independent of* n, *and*

$$\delta = \sup_{b \geq a > 0} \theta(a,b) \cdot \frac{b-a}{a} \tag{11}$$

Proof

Observe first that $\delta(n)$ is non decreasing in n. Indeed if we add to any problem (N,x) a new agent 0 with a null job, neither the optimal delay $\nabla(N,x)$ nor the expected delay $y_N(x) - x_N$ under μ are affected. Next we set $\beta = \sup_{b \geq a > 0} \theta(a,b) \cdot \frac{b-a}{a}$, fix an arbitrary problem (N,x) and compute

$$\begin{aligned} y_N(x) - x_N &= \sum_{N(2)} (\theta(x_i,x_j) \cdot x_j + \theta(x_j,x_i) \cdot x_i) \qquad (12) \\ &= \sum_{\{i,j\} \in N(2): x_i \leq x_j} x_i \cdot (1 + \theta(x_i,x_j) \cdot \frac{x_j - x_i}{x_i}) \leq (1+\beta) \cdot \nabla(N,x) \end{aligned}$$

Thus $\delta(n) \leq 1 + \beta$ for all n. When $N = \{1,2\}$ and $x_1 \leq x_2$, equation (12) reduces to

$$y_N(x) - x_N = x_1 \cdot (1 + \theta(x_1,x_2) \cdot \tfrac{x_2 - x_1}{x_1}) = \nabla(N,x) \cdot (1 + \theta(x_1,x_2) \cdot \tfrac{x_2 - x_1}{x_1})$$

implying $\delta(2) = 1 + \beta$.■

An easy consequence of equations (10) and (11) is that δ is infinite if and only if $\lambda(n)$ is infinite for all n.

Lemma 6 applies to all parametric methods, in particular to the quasi-proportional ones. We conclude this section by computing our two indices for some of the latter methods. Start with a $w-$method where $r(z) = \frac{w(z)}{z}$is non increasing. One checks easily

$$\sup_{a,b>0} \theta(a,b) \cdot \frac{b}{a} = \sup_{a,b>0} \frac{w(a)}{w(a)+w(b)} \cdot \frac{b}{a} = \frac{r(0)}{r(\infty)}$$

Therefore the excess delay and liability are both finite if and only if r is bounded away from 0 and from $+\infty$. For instance $w(z) = z^{\alpha}$ with $\alpha < 1$ has infinite excess delay and liability.

Next consider a $w-$method where $r(z)$ is non decreasing. For any a, b we have $\frac{a}{b} + \frac{r(a)}{r(b)} \geq 1$, by considering the cases $a \geq b$ and $a \leq b$ separately; hence $\frac{w(a)}{w(a)+w(b)} \cdot \frac{b}{a} \leq 1$ so liability and excess delay are both finite. Straightforward computations for $w(z) = z^{\alpha}$ with $\alpha \geq 1$ give

$$\delta = 1 + \sup_{x \geq 1} \frac{x-1}{x^{\alpha}+1};\ \lambda(n) = 1 + (n-1)\frac{(\alpha-1)^{1-\frac{1}{\alpha}}}{\alpha} \text{ for } 1 \leq \alpha \leq +\infty$$

Liability decreases with α, from $\lambda(n) = n$ for $\alpha = 1$ (Proportional method) until $\lambda(n) = \frac{n+1}{2}$ for $\alpha = 2$ (Quadratic method); it increases with α beyond 2, and its limit is $\lambda(n) = n$ at infinity (Shortest Job First).

The excess delay decreases for all α, from $\delta = 2$ for the Proportional method, to $\delta = \frac{1+\sqrt{2}}{2} \simeq 1.21$ for the Quadratic rule, and $\delta = 1$ at infinity.

8 Characterization of the proportional method

Our first characterization result contains two statements explaining a general trade-off between Ranking and Split-proofness for any separable rule. The third statement says that the proportional rule lies on the cusp of this trade-off. Recall from Lemma 2 that the method associated with a separable and split-proof rule is anonymous.

Theorem 1

i)If the separable rule ρ is split-proof, then $\lambda(n) \geq n$ for all n;
ii)If the separable method μ meets Ranking, then $\lambda(n) \leq n$ for all n.
iii)If the separable rule ρ is split-proof and meets at least one of

$$\textit{Ranking,}\text{or } \{\delta \leq 2\}, \text{ or}\{\lambda(n) \leq n \textit{ for some } n\}$$

then its method is Proportional.

The proof of Theorem 1, combined with that of Theorem 2, is in the Appendix.

Remark 2 *For $|N| = 4$ it is easy to construct a separable and split-proof rule that is not the Proportional one. I conjecture that such a construction is possible for $|N| = \infty$ as well. This suggests that Theorem 1 cannot be improved into a characterization of the Proportional rule.*

9 Recursivity and characterization of the proportional rule

For our second main result, we introduce another invariance property of scheduling rule satisfied by quasi-proportional rules and many other rules as well. Consider equation (6) in Lemma 3: p_σ is computed recursively by finding first the probability $\frac{w_n}{\sum_1^n w_k}$ that user $\sigma^{-1}(n)$ be ranked last in problem (N, x), multiplying it by the probability that user $\sigma^{-1}(n-1)$ be ranked last in the reduced problem $(N \diagdown \sigma^{-1}(n), x_{[N \diagdown \sigma^{-1}(n)]})$, and so on. Here the notation $x_{[S]}$ is the projection of $x \in \mathbb{R}_+^N$ on $\mathbb{R}_+^S$ .In other words the rule is entirely determined by the probability distribution of the job served last.

Definition 11

*The scheduling rule ρ is **recursive** if there exists for all N and all $x \gg 0$ a probability distribution $\pi(N, x) \in \Delta(N)$, such that for all $\sigma \in \Phi(N)$ with $\sigma^{-1}(n) = i$*

$$p_\sigma(N, x) = \pi_i(N, x) \cdot p_{\sigma[N \diagdown i]}(N \diagdown i, x_{[N \diagdown i]})$$

We extend the definition of $\pi(N, \cdot)$ to $\mathbb{R}_+^N \diagdown 0$ by setting $\pi_i(N, x) = 0$ whenever $x_i = 0$ and $\pi_i(N, x) = \pi_i(S, x_{[S]})$, where S is the set of positive coordinates in x.

The quasi-proportional rules are recursive, as well as separable. It turns out that the combination of these two properties essentially characterizes the quasi-proportional family. See Subsection 11.3 in the Appendix for a statement and proof.

Not all parametric rules are recursive. Define the **Serial** rule as the parametric rule such that the cdf F_a is uniform on $[0, a]$, i.e., $F_a(z) = \min\{\frac{z}{a}, 1\}$. Here (5) gives $\theta(a, b) = \frac{a}{2b}$ if $a \leq b$, $= 1 - \frac{b}{2a}$ if $b \leq a$. Thus for a profile x such that $0 < x_1 \leq x_2 \leq \cdots \leq x_n$, we have

$$y_i = (1 + \frac{n-i}{2}) \cdot x_i + \sum_{j=1}^{i-1} (1 - \frac{x_j}{2x_i}) \cdot x_j$$

This formula explains the serial terminology. Indeed user i's expected wait does not depend on the sizes of jobs larger that his own: $y_i(x) = y_i(x')$ if x and x' only differ in coordinate j, $j \neq i$, and $x_i \leq x_j < x'_j$, then $y_i(x) = y_i(x')$. This is the *serial principle* discussed by Sprumont [1998] and others. Moreover, the Serial scheduling *method* is clearly the only mapping $x \rightarrow y$

meeting the three properties: equal treatment of equals ($x_i = x_j \Rightarrow y_i = y_j$), $\sum_N x_i \cdot y_i = v(N, x)$ (Lemma 1), and the serial principle.

It is easy to check that in a typical 3 person problem the Serial rule is not recursive. Moreover it is not split-proof. On the other hand it achieves the optimal liability because $\sup_{b>0} \theta(a, b) \cdot b = \frac{a}{2}$.

Theorem 2
The Proportional rule is the only recursive and split-proof rule that meets at least one of

$$\textit{Ranking}, \text{ or } \{\delta(2) \leq 2\}, \text{ or } \{\lambda(2) \leq 2\}$$

Remark 3 *Parallell to Lemma 2, we have a sufficient condition for Split-proofness among recursive rules. The recursive scheduling rule ρ is split-proof if for all $N, i \in N, j \notin N$, and all $x \in \mathbb{R}^N_+, \widetilde{x} \in \mathbb{R}^{N \cup j}_+$*

$$\{x_{[N \setminus i]} = \widetilde{x}_{[N \setminus i]}, x_i = \widetilde{x}_i + \widetilde{x}_j\} \Longrightarrow \{\pi_k(N, x) \geq \pi_k(N \cup j, \widetilde{x}) \text{ for all } k \in N \setminus i$$

This says that when agent i splits his job in two (or more) pieces, the probability that another job is scheduled last does not become larger. Loosely speaking, each function $\pi_k(N, x)$ is subadditive upon splitting[8]

Remark 4 *An dual definition of Recursivity works by drawing the user scheduled first in the given problem, then repeating in the reduced problem and so on. For instance if we schedule i first with probability $\frac{1}{w(x_i)}$, the resulting rule is separable and the probability that job j precedes job i is $\frac{w(x_i)}{w(x_j)+w(x_i)}$, just like with the rule of Definition 2. These two rules are different with three or more users, but they implement the same method. However this dual definition of Recursivity does not yield interesting Split-proof rules. For instance the dual Proportional rule (scheduling i first with probability $\frac{1}{x_i}$) is not Split-proof.*

10 Concluding comments and Open problems

1 Split- and Merge-invariance.
The Uniform method meets a stronger property than merge-proofness, namely it is *Merge-invariant.* When a coalition merges, it can achieve after merging precisely the same profile of expected wait as before merging, and no better. Moreover users outside the coalition are not affected by the merge.

[8] The proof is available upon request from the author.

Formally Merge-invariance requires that for all N, x, T, i^* as in the premises of Definition 6,

$$y_{[T]} \in (y^*_{i^*} - \sum_T x_i) \cdot e_{[T]} + F(\widetilde{T}, x_{[T]}). \tag{13}$$

It turns out that the Uniform method is the *only* merge-invariant scheduling method[9]. Thus merge-invariance is not compatible with even a modicum of responsiveness, or a finite cap on individual liability or excess delay.

On the other hand, the Proportional rule is *Split-invariant*, namely splitting one's job is a matter of indifference. To check this, observe that the cdfs $F_a(z) = z^a$ satisfy $F_a \cdot F_b = F_{a+b}$ for all $a, b > 0$, therefore inequality (4) is actually an equality for the Proportional rule, and the same holds true for inequality (2), expressing that the expected wait of agent i_* is the same before and after the split.

While Merge- invariance single-handedly characterizes the Uniform rule, the Proportional rule is not the only split-invariant rule. Yet it is characterized by Split invariance within our two benchmark families:

i) If the separable scheduling rule ρ is split-invariant, then its method is Proportional: $y_i(N, x) = x_i + \sum_{N \setminus i} \frac{x_i \cdot x_j}{x_i + x_j}$ for all $N, x, x \geq 0$, and i.

i) The Proportional is the only recursive and split-invariant scheduling rule[10].

2 Queuing problems

Parametric rules have a natural extension to queuing problems, where jobs are born at arbitrary dates. Each time a new job i of size x_i is born, we draw the variable Z_i with cdf F_{x_i} independently of the draws of all Z_j corresponding to jobs born earlier, and we process jobs in the preemptive priority defined by these realizations. This definition preserves the property of merge-proofness provided the reported merged job is born not earlier than the youngest of the component jobs; it preserves split-proofness as well, if the split jobs are born not earlier than the true job. Thus the extended Proportional rule remains equally appealing because of these two properties, plus Ranking and the same liability $n \cdot a$, where n is the number of jobs not yet completed or not yet born when job a appears. Whether or not the characterizations of the Proportional extend to the queuing context is left for future research.

3 An open question.

[9] The proof is available upon request from the author.

[10] The proof is similar to that of Theorems 1 and 2. It is available upon request from the author.

Is Recursivity a necessary assumption in Theorem 2? I.e., can we find split-proof rules meeting at least one of Ranking, $\{\delta(2) \leq 2\}$, or $\{\lambda(2) \leq 2\}$, and different from the Proportional one? I conjecture that such rules exist.

We can replace Recursivity in Theorem 2 by the assumption that the rule is parametric. Recall that many parametric rules are not recursive (e.g., the Serial rule); clearly many recursive rules are not separable, hence not parametric. If a rule is split-invariant, or if it is split-proof and meets one of Ranking, $\{\delta(2) \leq 2\}$ and $\{\lambda(2) \leq 2$ for all $a\}$, the argument in Step 3 of the proof of Theorem 2 shows that for all N, x, i, the probability that i is scheduled last is proportional to x_i. I have shown that this property characterizes the Proportional rule among *continuous* (atomless) parametric rules[11], and conjecture that the same holds among all parametric rules.

4 Transfers and coordinated splitting.

A strategic maneuver related to merging and splitting is the partial transfer of jobs among users. Two (or more) agents choose to reallocate their jobs of sizes x_1 and x_2 as x_1' and x_2', with $x_1 + x_2 = x_1' + x_2'$. The transfer of jobs encompasses splitting, by taking $x_2 = 0$, and merging, by taking $x_2' = 0$.

Ozsoy [2005] shows that the Proportional rule is not vulnerable to *pairwise* transfers, i.e., involving only two agents. The same is true for all quasi-proportional rules with $w(a) = a^\alpha$ and $0 \leq \alpha \leq 1$. However, transfers among three agents can be profitable under these scheduling rules, except for the Uniform rule. The latter is invulnerable to transfers among any number of agents.

Another strategic maneuver threatening the Proportional rule is the cooperative version of splitting: several agents split each job into smaller pieces, and decide on the allocation of these slots conditional upon the actual service ordering. Again, the Uniform rule stands out as the only simple method immune to such coordinated splitting tactics (Ozsoy [2005]).

In the quasi-linear model of the companion paper Moulin [2004], no efficient rules is invulnerable to transfers involving three or more agents, but pairwise transfer-proofness points to a one-dimensional line of rules borne by two appealing rules, one split-proof and the other merge-proof. These solutions are not built on any proportionality idea, but apply instead the Shapley value to an appropriate cooperative game. They can also be derived from invariance or consistency properties (Maniquet [2003], Chun [2004 a,b]).

[11] The proof is available upon request.

References

[1] Banker, R. 1981. Equity consideration in traditional full-cost allocation practices: an axiomatic perspective, *Joint Cost Allocations*, S. Moriarty, ed. Norman: University of Oklahoma.

[2] Bender, M., Chakrabarti, S., and Muthukrishnan, S. 1998. "Flow and Stretch Metrics for Scheduling Continuous Job Streams," in *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms.*

[3] Chun, Y. , 2000. Agreement, Separability, and other Axioms for Quasi-linear Social Choice Problems, *Social Choice and Welfare,* 17, 507-521.

[4] Chun, Y., 2004. A Note on Maniquet's Characterizations of the Shapley Value in Queuing Problems, mimeo, Seoul National University.

[5] Chun, Y., 2004. Consistency and Monotonicity in Sequencing Problems, mimeo, Seoul National University.

[6] Curiel, I., G. Pederzoli and S. Tijs, 1989. Sequencing Games, *European Journal of Operational Research*, 40, 344-351.

[7] Curiel, Il, H. Hamers and F. Klijn, 2002. Sequencing Games: a Survey, in *Chapters in Game Theory in Honor of Stef Tijs*, P. Borm and H. Peters (eds.), pp. 27-50, Kluwer Academic Publishers, Boston.

[8] De Frutos, M.A., 1999. Coalitional Manipulations in a Bankruptcy Problem, *Review of Economic Design,* 4, 255-272.

[9] Demers A., S. Keshav and S. Shenker 1990. Analysis and simulation of a fair queuing algorithm, *Internetworking: Research and Experience, 1, 3-26.*

[10] Dolan, R., 1978. Incentive Mechanisms for Priority Queueing Problems, *The Bell Journal of Economics* 9, 421-436.

[11] Douceur, J., 2002. The Sybil Attack, *Proceedings of the 1st International Workshop on Peer-to-Peer Systems* (IPTPS 2002), LNCS 2429, pp 251-260.

[12] Fishburn, P., 1992. Induced binary probabilities and the linear ordering polytope: a status report, *Mathematical Social Sciences*, 23, 67-80.

[13] Friedman, E.J. and Henderson, S.G. (2003). "Fairness and Efficiency in Web Server Protocols," in *Proceedings of the 2003 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, 229-237, ACM Press.

[14] Hain R., and M. Mitra, 2001. Simple Sequencing Problems with Interdependent Costs, mimeo, University of Bonn.

[15] Hardy G., J.E. Littlewood, and G. Polya, 1934. *Inequalities*, Cambridge University press.

[16] Haviv, M.,2001. The Aumann-Shapley Price Mechanism for Allocating Congestion Costs, *Operation Research Letters*, 29, 211-215.

[17] Haviv, M. and Y. Ritov, 1998. Externalities, Tangible Externalities, and Queue Disciplines, *Management Science*, 44, 6, 850-858.

[18] Ju, B.G., 2003. Manipulations via Merging and Splitting in claim problems, *Review of Economic Design,* 8, 205-215.

[19] Ju, B.G., E. Miyagawa, and T. Sakai, 2003. Coalitional Manipulations and Generalized Proportional Rules, mimeo, Columbia University.

[20] Kittsteiner, T. and B. Moldovanu, 2003. Auction-Based Queue Disciplines, mimeo, University of Bonn.

[21] Kittsteiner, T. and B. Moldovanu, 2003. Priority Auctions and Queue Disciplines that Depend on Processing Time, mimeo, University of Bonn.

[22] Klijn, F. and E. Sánchez, 2002. Sequencing Games Without a Completely Specific Initial Order, *Reports in Statistics and Operations Research*, Report 02-04, 1-17.

[23] Lawler, E., J.K. Lenstra, Kan A. Rinnooy, and D. Shmoys, 1993. Sequencing and Scheduling: Algorithms and Complexity, Chapter 9 in Handbooks in *Operations Research* and *Management Science*, Vol. 4, Elsevier.

[24] Maniquet, F., 2003. A Characterization of the Shapley Value in Queueing Problems, *Journal of Economic Theory,* 109,1, 90-103.

[25] Mitra, M., 2001. Mechanism Design in Queueing Problems, *Economic Theory,* 17, 277-305.

[26] Mitra, M., 2002. Achieving the First Best in Sequencing Problems, *Review of Economic Design,* 7, 75-91.

[27] Moulin, H., 1985 Egalitarianism and Utilitarianism in Quasi-linear Bargaining, *Econometrica,* 53, 1, 49-67.

[28] Moulin, H., 1987. Equal or Proportional Division of a Surplus, and Other Methods, *International Journal of Game Theory*, 16, 3, 161–186.

[29] Moulin H., 2002. The Proportional Random Allocation of Indivisible Units, *Social Choice and Welfare,* 19,2,381-413.

[30] Moulin, H., 2004. On Scheduling Fees to Prevent Merging, Splitting and Transferring Jobs, mimeo, Rice University.

[31] Moulin, H., 2005. Minimizing the Worst Slowdown: Off-line and On-line, mimeo, Rice University.

[32] Moulin, H. and Y. Sprumont, 2003. On Demand Responsiveness in Additive Cost Sharing, forthcoming in *Journal of Economic Theory.*

[33] Moulin, H. and R. Stong, 2002. Fair Queuing and other Probabilistic Allocation Methods, *Mathematics of Operations Research*, 27, 1, 1-31.

[34] Moulin, H. and R. Stong, 2003. Filling a Multicolor Urn: an axiomatic analysis, *Games and Economic Behavior*, 45, 242-269.

[35] Ozsoy, H., 2005. Coordinated Splitting in Probabilistic Scheduling, mimeo, Rice University.

[36] Queyranne, M., 1993. Structure of a Simple Scheduling Polyhedron, *Mathematical Programming*, 58, 263-285.

[37] Shapley, L.S., 1971. Core of Convex Games, *International Journal of Game Theory*, 1, 11-26.

[38] Sprumont, Y., 1998. Ordinal cost sharing. *Journal of Economic Theory* **81**, 126-162.

[39] Sprumont, Y., 2003. Aumann-Shapley pricing: a reconsideration of the discrete case, mimeo, Universite de Montreal.

[40] Suijs, J., 1996. On Incentive Compatibility and Budget Balancedness in Public Decision Making, *Economic Design,* 2, 193-209

[41] Wierman, A. and Harchol-Balter, M., 2003. Classifying Scheduling Policies with Respect to Unfairness in an M/GI/1, in *Proceedings of ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems.*

11 Appendix: remaining proofs

11.1 Lemma 1

Statement i. Write $F_0(N,x)$ for the set of feasible profiles of expected waits at (N,x), and $C(N,x)$ for the core of $(N,v(\cdot,x))$. Also use $x\cdot y$ for the coordinatewise multiplication of vectors. Fix (N,x) and define $\widetilde{N}=\{i\in N|x_i>0\}$. Together the three following facts prove the statement:

1.If $N=\widetilde{N}$ then $y\in F_0(N,x)\Longleftrightarrow x\cdot y\in C(N,x)$

2. $y\in F_0(N,x)\Longleftrightarrow\{y_{[\widetilde{N}]}\in F_0(\widetilde{N},x_{[\widetilde{N}]})$ and $y_i\in[0,\sum_N x_j]\}$

3. $x\cdot y\in C(N,x)\Longleftrightarrow x_{[\widetilde{N}]}\cdot y_{[\widetilde{N}]}\in C(\widetilde{N},x_{[\widetilde{N}]})$.

Fact 3 requires no proof and fact 2 is clear: once we order jobs randomly in $\widetilde{N}$ to achieve $y_{[\widetilde{N}]}$, where we schedule null jobs does not matter and we can achieve any $y_{[N\diagdown\widetilde{N}]}$ within the announced bounds.

Now fix (N,x) with $x\gg 0$, and for all $\sigma\in\Phi(N)$ let y^σ be the profile of wait under σ namely $y_i^\sigma=\sum_{P(i;\sigma)}x_j$, where $P(i;\sigma)=\{j\in N|\sigma(j)\leq\sigma(i)\}$. Set $P_-(i;\sigma)=P(i;\sigma)\diagdown i$. Routine computation shows $(x\cdot y^\sigma)=v(P(i;\sigma),x)-v(P_-(i;\sigma),x)$, namely $x\cdot y^\sigma$ is the vector of marginal contributions in the game $(N,v(\cdot,x))$. As $v(\cdot,x)$ is supermodular, a classic result (Shapley [1971]) says that $C(N,x)$ is the convex hull of $x\cdot y^\sigma,\sigma\in\Phi(N)$. On the other hand $F_0(N,x)$ is the convex hull of $y^\sigma,\sigma\in\Phi(N)$. This gives fact 1 because $x\gg 0$.

Statement ii. "Only if" is clear from statement i). "If" follows because such a vector minimizes the sum $\sum_{\widetilde{N}}x_i\cdot y_i+\sum_{N\diagdown\widetilde{N}}y_i$ over $F_0(N,x)$.

11.2 Proposition 1

Step 1 Preliminary result.

We fix an arbitrary parametric rule $\{F_a, a > 0\}$ where the cdfs may have atoms. We start by a reformulation of the probability that a job of size b precedes one of size a, namely $\theta(a,b)$ in (5). Recall that each F_a lives in the space $\mathcal{H}$ of non-decreasing, right-continuous real-valued functions on $\mathbb{R}_+$ with bounded range. For such a function F we write $\partial F(z) = F(z) - F_-(z)$ for the jump of F at z, where $F_-(z)$ is the left limit. We denote by F^+ the integral of these jumps, namely the staircase function $F^+(z) = \sum_{t \le z} \partial F(t)$ (the sum is well defined because F is non-decreasing). Check that $F^0 = F - F^+$ is continuous and non decreasing.

The canonical decomposition $F = F^0 + F^+$ allows us to define the integral $\int_0^\infty F \cdot dG$, for any two $F, G \in \mathcal{H}$. If F, G are both continuous, this is the familiar Stieltjes integral (e.g., Hardy et al [1938]). For arbitrary F, G in $\mathcal{H}$ we write $\mathcal{J}(F), \mathcal{J}(G)$ for the set of their jumps, namely $z \in \mathcal{J}(F) \Leftrightarrow \partial F(z) > 0$; then we define $\int F^+ \cdot dG^0 = \sum_{\mathcal{J}(F)} \partial F(z) \cdot (G(\infty) - G(z)) = (F^+ \cdot G)(\infty) - \sum_{\mathcal{J}(F)} \partial F(z) \cdot G(z)$, and $\int F \cdot dG^+ = \sum_{\mathcal{J}(G)} F(z) \cdot \partial G(z)$. These combine into the following definition

$$\int_0^\infty F \cdot dG = \int_0^\infty F^0 \cdot dG^0 + (F^+ \cdot G^0)(\infty) - \sum_{\mathcal{J}(F)} \partial F(z) \cdot G^0(z) + \sum_{\mathcal{J}(G)} F(z) \cdot \partial G(z).$$

From this it is straightforward to deduce the integration by parts formula:

$$\int_0^\infty F \cdot dG + \int_0^\infty G \cdot dF = (F \cdot G)(\infty) - (F \cdot G)(0) + \sum_{\mathcal{J}(F) \cap \mathcal{J}(G)} \partial F(z) \cdot \partial G(z). \tag{14}$$

Now for two independent random variables Z_a, Z_b with cdfs F_a, F_b, we have

$$prob\{Z_a = Z_b\} = \sum_{\mathcal{J}(a) \cap \mathcal{J}(b)} \partial F_a(z) \cdot \partial F_b(z)$$

Indeed the equality only occurs when both draws are in the atomic part of the two cdfs. Similarly the probability of $\{Z_b \le Z_a\}$ is

$$\int_0^\infty F_b^0 \cdot dF_a^0 + (F_b^+ \cdot F_a^0)(\infty) - \sum_{\mathcal{J}(b)} \partial F_b(z) \cdot F_a^0(z) + \sum_{\mathcal{J}(a)} F_b(z) \cdot \partial F_a(z) = \int_0^\infty F_b \cdot dF_a$$

where the last term in the LHS is the probability of $\{Z_b \leq Z_a\} \cap \{Z_a \in \mathcal{J}(a)\}$ and the sum of the two middle terms is that of $\{Z_b \leq Z_a\} \cap \{Z_b \in \mathcal{J}(b)\}$. Comparing with equation (5), and using integration by parts (14) we conclude

$$\begin{aligned} \theta(a,b) &= \int_0^\infty F_b \cdot dF_a - \frac{1}{2} \sum_{\mathcal{J}(a) \cap \mathcal{J}(b)} \partial F_a(z) \cdot \partial F_b(z) = \\ &= \int_0^\infty (1 - F_a) \cdot dF_b + \frac{1}{2} \sum_{\mathcal{J}(a) \cap \mathcal{J}(b)} \partial F_a(z) \cdot \partial F_b(z) \end{aligned} \qquad (15)$$

Step 2: Demand monotonic parametric rules.
In this Step we show that the parametric rule $\{F_a, a > 0\}$ is demand monotonic if $\{a \leq b\} \Rightarrow \{F_b(z) \leq F_a(z)$ for all $z\}$ (in the sense of stochastic dominance). Let $\{F_a, a > 0\}$ define a parametric rule. We develop $\theta(a,b)$ with the help of (15) and the definition of the integral

$$\theta(a,b) = \int_0^\infty (1 - F_a) \cdot dF_b^0 + \sum_{\mathcal{J}(b)} (1 - F_a)(z) \cdot \partial F_b(z) + \frac{1}{2} \sum_{\mathcal{J}(a) \cap \mathcal{J}(b)} \partial F_a(z) \cdot \partial F_b(z)$$

The first integral term is non decreasing in a, because $F_a(z)$ is non increasing in a. The first sum is also non decreasing, but the variation of $\partial F_a(z)$ (hence of the second sum) in a is ambiguous. However, for any $z \in \mathcal{J}(a) \cap \mathcal{J}(b)$ the corresponding terms in the second and third sums are

$$(1 - F_a(z)) \cdot \partial F_b(z) + \partial F_a(z) \cdot \partial F_b(z) = 1 - \frac{1}{2}(F_a(z) + F_b^-(z)) \cdot \partial F_b(z)$$

and the desired monotonicity follows.

Here is a demand monotonic parametric rule for which $a \to Z_a$ is not stochastic-dominance monotonic. For $a < 1$, let $Z_a = 2$ with probability one, and for $a \geq 1$, let $Z_a = 1$ or 3 with respective probabilities $\frac{1}{4}$ and $\frac{3}{4}$.
Step 3: Proof of Proposition 1. We must show inequality (4) when $F_a \cdot F_b \leq F_{a+b}$ holds for all $a, b > 0$. Fix $S, i, i \notin S$, and $b, a_j, j \in S$ as in the

premises of (4), and set $\overline{a} = \sum_1^m a_j$. From (15) we have $\theta(b, \overline{a}) = \int F_{\overline{a}} \cdot dF_b - \frac{1}{2}\sum_{\mathcal{J}(\overline{a})\cap\mathcal{J}(b)} \partial F_{\overline{a}} \cdot \partial F_b$. Next we compute $\gamma = prob\{\max_S \sigma(j) < \sigma(i)|(b, a_j)\}$ with the help of $\widetilde{Z}_k$, the k-th order statistics of the independent variables Z_{a_j} (thus $\widetilde{Z}_1 = \min_{1,..,m} Z_{a_j}$ and $\widetilde{Z}_m = \max_{1,..,m} Z_{a_j}$). We have

$$\gamma \;=\; prob\{\widetilde{Z}_m < Z_b\} + \frac{1}{2} prob\{\widetilde{Z}_{m-1} < \widetilde{Z}_m = Z_b\}$$

$$+\frac{1}{3} prob\{\widetilde{Z}_{m-2} \;<\; \widetilde{Z}_{m-1} = \widetilde{Z}_m = Z_b\} + \cdot\cdot \leq prob\{\widetilde{Z}_m < Z_b\} + \frac{1}{2} prob\{\widetilde{Z}_m = Z_b\} =$$

$$= \int_0^\infty F_{a_1} \cdot .. \cdot F_{a_m} \cdot dF_b - \tfrac{1}{2}\sum_{\mathcal{J}^*\cap\mathcal{J}(b)} \partial(F_{a_1} \cdot \cdot \cdot F_{a_m})(z) \cdot \partial F_b(z)$$

where $\mathcal{J}^* = \cup_{1,..,m}\mathcal{J}(a_j)$. Mimicking the argument in Step 2 it is easy to show that $T(G) = \int G \cdot dF - \frac{1}{2}\sum \partial G \cdot \partial F$ is monotonic in the sense that $\{G_1(z) \leq G_2(z) \text{ for all } z\} \Longrightarrow T(G_1) \leq T(G_2)$. Applying this to $G_1 = F_{a_1} \cdot .. \cdot F_{a_m}$ and $G_2 = F_{\overline{a}}$ completes the proof of (4).

11.3 Proof of theorems 1 and 2

Step 1
Let ρ be a split-proof scheduling rule, and let $\theta^{i,j}(a, b) = prob\{\sigma(j) < \sigma(i)|(a, b)\}$ be the probability that job j of size b precedes job i of size a in the $\{i, j\}$ problem. Then $\theta^{i,j} = \theta$ is independent of i, j. Moreover $a \to \theta(a, c-a)$ is subadditive in a. Finally $\sup_b \theta(a, b) \cdot b \geq a$ for all a, namely $\lambda(2) \geq 2$.

Proof. That $\theta^{i,j}$ does not depend on i, j was established in the proof of statement "only if"in Lemma 2 (note that Separability of ρ is not needed for the argument). Next we define for all a, c such that $0 \leq a \leq c$, $f(a, c) = \theta(a, c - a)$. Fix $a_1, a_2, a_3 > 0$ and consider the split of user 2 in problem $\{1, 2\}, (a_1, a_2+a_3)$ to users $2, 3$ in $\{1, 2, 3\}, (a_1, a_2, a_3)$. Split-proofness implies

$$f(a_1, a_1 + a_2 + a_3) = \theta(a_1, a_2 + a_3) \geq prob\{1 \text{ is last in } (a_1, a_2, a_3)\}$$

Exchanging the role of agents gives $f(a_i, a_1 + a_2 + a_3) \geq prob\{i$ is last in $(a_1, a_2, a_3)\}$ for $i = 1, 2, 3$. Summing the three inequalities and using $1 - f(a_3, a_1 + a_2 + a_3) = f(a_1 + a_2, a_1 + a_2 + a_3)$ (because $\theta(a, b) + \theta(b, a) = 1$) gives the desired subadditivity property.

Finally we fix $a > 0$, $m \in \mathbb{N}$, and observe $f(m \cdot a, 2m \cdot a) = \frac{1}{2}$. Superadditivity of f implies $f(a, 2m \cdot a) \geq \frac{1}{m} f(m \cdot a, 2m \cdot a)$, or equivalently $\theta(a, (2m - 1) \cdot a) \geq \frac{1}{2m}$, from which $\sup_b \theta(a, b) \cdot b \geq a$ follows at once.

Step 2: Statements i and ii in Theorem 1.
Statement i follows at once from Step 1 and equation (3) for separable rules. For statement ii), fix a separable method ρ for which $\theta^{i,j}$ may depend on i, j. Apply Ranking to $N = \{i, j\}$ and $x_i = a, x_j = b$ such that $a \leq b$. We get

$$\theta^{i,j}(a,b) \cdot b \leq \theta^{j,i}(b,a) \cdot a \Rightarrow \theta^{i,j}(a,b) \leq \frac{a}{a+b} \leq \frac{a}{b}$$

and $\lambda(n) \leq n$ then follows from the non anonymous version of (3).

Step 3

Let ρ be a split-proof rule, with the corresponding θ for two person problems as in Step 1. If one of Ranking, or $\{\delta(2) \leq 2\}$, or $\{\lambda(2) \leq 2\}$ holds, then $\theta(a,b) = \frac{a}{a+b}$ for all a, b.

Assume Ranking. As in the above step, we have then $\theta(a,b) \leq \frac{a}{a+b} \Leftrightarrow f(a,c) \leq \frac{a}{c}$ for all $a, c, 0 < a < c$. By subadditivity of f in a, we get

$$1 = f(c,c) \leq f(a,c) + f(c-a,c) \leq \frac{a}{c} + \frac{c-a}{c} = 1$$

hence $f(a,c) = \frac{a}{c}$ as claimed.

Assume $\lambda(2) \leq 2$. By (3) we have $\theta(a,b) \leq \frac{a}{b}$ for all a, b. Equivalently, $f(a,c) \leq \frac{a}{c-a}$ for all $a, c, 0 < a < c$. Now apply k times the subadditivity of f in its first variable:

$$f(a,c) \leq 2f(\frac{a}{2},c) \leq .. \leq 2^k f(\frac{a}{2^k},c) \leq \frac{a}{c - \frac{a}{2^k}}$$

implying $f(a,c) \leq \frac{a}{c}$ and the desired conclusion as above.

Assume $\delta(2) \leq 2$. Then, with the notations in the proof of Lemma 6, $\delta(2) = 1 + \beta$ holds even for a non separable rule, provided $\theta^{i,j}$ does not depend on i, j, implied here by Split-proofness and Step 1. Thus

$$\beta = \sup_{b \geq a > 0} \theta(a,b) \cdot \frac{b-a}{a} \leq 1 \Leftrightarrow f(a,c) \leq \frac{a}{c-2a} \text{ for all } a, c, 0 < a < \frac{c}{2}$$

Applying k times the subadditivity of f as above, we get $f(a,c) \leq \frac{a}{c - \frac{a}{2^{k-1}}}$, hence $f(a,c) \leq \frac{a}{c}$ whenever $a < \frac{c}{2}$. As $f(\frac{c}{2},c) = \frac{1}{2}$, f must be linear in a on $]0, \frac{c}{2}[$, and once more subadditivity on $]\frac{c}{2}, c[$ gives $f(a,c) \leq \frac{a}{c}$ everywhere and we are done.

Step 4: End of proof. If the rule ρ is separable and split-proof, it is anonymous (lemma 2), and $\lambda(2) \leq 2$ is equivalent to $\lambda(n) \leq n$ for any n. If ρ

meets the premises of the statement in Step 3, and is separable as well, then the function θ defines its method for any number of jobs, and this method is proportional. This proves statement *iii* in Theorem 1.

Now suppose ρ meets the premises of the statement in Step 3, and is recursive as well. We fix an arbitrary problem (N, x) and consider the split of 2 in problem $\{1,2\}, (x_1, \sum_{N \setminus 1} x_i)$ to $\{2,3,..,n\}$ in problem (N, x). Splitproofness gives $\pi_1(x) \leq \theta(x_1, \sum_{N \setminus 1} x_i) = \frac{x_1}{\sum_N x_i}$. Repeat the argument for all i: as $\pi(x)$ is a probability distribution, it must be the Proportional one, thus by Recursivity ρ is the Proportional rule.■

11.4 A characterization of the quasi-proportional family

Proposition 3
A scheduling rule is quasi-proportional if and only if it is separable, recursive, anonymous and meets the following ***Positivity*** *property*

$$prob\{\sigma(j) < \sigma(i)|x\} > 0 \text{ for all } N, x, x \gg 0, \text{ and all } i, j.$$

Positivity is needed in the above statement: Shortest (or Longest) Jobs First meets the other three properties.

Proof
Only the " if" statement requires a proof. Let ρ be a rule with the four stated properties, and $\pi(N, x)$ be the associated probability distribution of the last scheduled agent. By Anonymity, it takes the form $\pi(n, x)$, $n = |N|$, and is symmetric in all variables x_i. Fix n, x and use the simplified notation $\{1, 2, .., n-1|x\}$ for the event $\{\sigma(1) < \sigma(2) < .. < \sigma(n-1)|x\}$. A simple partition of this event gives the equality

$$prob\{1,2,..,n-1|x\} = prob\{1,2,..,n|x\} + \sum_{1 \leq i \leq n-1} prob\{1,..,i-1,n,i,..,n-1|x\}$$

By Separability and Recursivity, the summation in the RHS is $\pi_{n-1}(x) \cdot prob\{1,2,..,n-2|x_{[N \setminus n,n-1]}\}$. Writing $\pi(n, x)$ simply as $\pi(x)$, we get

$$(1-\pi_n(x)) \cdot prob\{1,2,..,n-1|x_{[N \setminus n]}\} = \pi_{n-1}(x) \cdot prob\{1,2,..,n-2|x_{[N \setminus n,n-1]}\} \quad (16)$$

Apply (16) first for $n = 3$, with the notation $\theta(b,a) = \pi_2((a,b))$ for the probability that a job of size a precedes one of size b. We get $(1 - \pi_3(x)) \cdot \theta(b,a) = \pi_2(x)$ for all $x = (a,b,c)$, $x \gg 0$. By Positivity, $\theta(b,a) > 0$, so $\pi_2(x) = 0$ would imply $\pi_1(x) = 0$. Exchanging the role of agents 1 and 3, we see that $\pi_2(x) = 0$ is impossible. Thus $\pi_i(x) > 0$ for $i = 1,2,3$. We can now rewrite (16) as

$$\theta(b,a) = \frac{\pi_2}{\pi_1 + \pi_2}(a,b,c) \Longleftrightarrow \frac{\pi_1}{\pi_2}(a,b,c) = \frac{1}{\theta(b,a)} - 1$$

namely the ratio $\frac{\pi_1}{\pi_2}$ only depends on a, b. From $\frac{\pi_1}{\pi_2} \cdot \frac{\pi_2}{\pi_3} \cdot \frac{\pi_3}{\pi_1} = 1$, a standard argument gives the existence of three real positive function w_i such that $\frac{\pi_i}{\pi_j}(a,b) = \frac{w_i(a)}{w_j(b)}$ for all i, j in $\{1,2,3\}$. Anonymity gives $w_i = w$ for all i, and we conclude $\pi_i(x) = \frac{w_i(x_i)}{w_1(x_1)+w_2(x_2)+w_3(x_3)}$ for all profiles $x \gg 0$ of dimension 3.

Finally we show $\pi_i(x) = \frac{w_i(x_i)}{\sum w_j(x_j)}$ for all (N, x), by an induction argument on $n = |N|$. Assume this holds up to $n-1$. This implies $prob\{1,2,..,n-2|x_{N\smallsetminus n,n-1}\} \neq 0$, therefore (ab) becomes $(1-\pi_n(x)) \cdot \pi_{n-1}(x_{N\smallsetminus n}) = \pi_{n-1}(x)$. Now $\pi(x) \gg 0$ and the desired claim follow easily.