GRADING IN GAMES OF STATUS:
MARKING EXAMS AND SETTING WAGES

By

Pradeep Dubey and John Geanakoplos

January 2006

COWLES FOUNDATION DISCUSSION PAPER NO. 1544R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Grading in Games of Status: Marking Exams and Setting Wages*

Pradeep Dubey† and John Geanakoplos‡

December, 2005

Abstract

We introduce *grading* into *games of status*. Each player chooses effort, producing a stochastic output or score. Utilities depend on the ranking of all the scores. By clustering scores into grades, the ranking is coarsened, and the incentives to work are changed.

We first apply games of status to grading exams. Our main conclusion is that if students care primarily about their status (relative rank) in class, they are often best motivated to work *not* by revealing their exact numerical exam scores $(100, 99, ..., 1)$, but instead by clumping them into coarse categories (A,B,C).

When student abilities are *disparate*, the optimal grading scheme is always coarse. Furthermore, it awards fewer A's than there are alpha-quality students, creating small elites. When students are homogeneous, we characterize optimal grading schemes in terms of the stochastic dominance between student performances (when they shirk or work) on subintervals of scores, showing again why coarse grading may be advantageous.

In both the disparate case and the homogeneous case, we prove that absolute grading is better than grading on a curve, provided student scores are independent.

We next bring *games of money and status* to bear on the optimal wage schedule: workers can be motivated not merely by the purchasing power of wages, but also by the status higher wages confer. How should the employer combine both incentive devices to generate an optimal pay schedule?

When workers' abilities are disparate, the optimal wage schedule creates different grades than we found with status incentives alone. The very top type should be motivated solely by money, with enormous salaries going to a tiny elite. Furthermore, if the population of workers diminishes as we go up the ability ladder and their disutility for work does not fall as fast, then the optimal wage schedule exhibits increasing wage differentials, despite the linearity in production.

When workers are homogeneous, the same status grades are optimal as we found with status incentives alone. A bonus is paid only to scores in the top status grade.

Keywords: Status, Grading, Incentives, Education, Exams, Wages
JEL Classification: C70, I20, I30, I33

*This is a revision, with a slightly altered title, of Dubey–Geanakoplos (2004).

†Center for Game Theory in Economics, SUNY, Stony Brook and Cowles Foundation, Yale University

‡Cowles Foundation, Yale University

1 Introduction

Examiners typically record scores on a precise scale 100, 99, ..., 1. Yet when they report final grades, many of them nowadays tend to clump students together in broad categories A, B, C, discarding information that is at hand. Why?

Many explanations come to mind. Less precision in grading may reflect the noisiness of performance: a 95 may be statistically insignificantly better than a 94. Alternatively, the professor may require less effort in dividing students among three categories rather than a hundred. Finally, it may be that lenient grading is a device by which professors lure students into their class; unable to call an exam with 70% correct answers a 95, they call it an A instead.

We call attention to a different explanation. Suppose that the professor judges each student's performance exactly, though the performance itself may depend on random factors, in addition to ability and effort. Suppose also that the professor is motivated solely by the desire to induce his students to work hard. Third, and most importantly, suppose that the students care about their relative rank in the class, that is about their *status*. We show that, in this scenario, coarse grading often motivates students to work harder.

Status is a great motivator.[1] For many people, honors conferring status, but little remuneration now or in the future, often bring forth the greatest effort.[2] Ranks and titles are ubiquitous, in academia, in the armed forces, in corporations, and in public bureaucracies. They define a hierarchy which, even when its original purpose might have been organizational (say to signal lines of authority), always creates incentives for people to exert effort in order to obtain higher status.

One might think that finer hierarchies generate more incentives. But this is often not the case. Coarse hierarchies can paradoxically create more competition for status, and thus provide better incentives for work.

To analyze the incentive effects of status, Section 2 introduces games of "pure" status, i.e., games in which the utilities are *solely* in terms of the relative ranking of the players. The players choose effort levels, which then jointly yield (possibly

[1]Veblen (1899) famously introduced conspicuous consumption, i.e., the idea that people strive to consume more than others partly for the sake of higher status. A large empirical literature, starting from Easterlin (1974), has shown that happiness indeed depends not just on absolute, but also on *relative*, consumption.

The modeling of status has taken two forms. The cardinal approach makes utility depend on the difference between an individual's consumption and others' consumption (see, e.g., Duesenberry (1949), Pollak (1976), and Fehr–Schmidt (1999)). The ordinal approach makes utility depend on the individual's rank in the distribution of consumption (see, e.g., Frank (1985), Robson (1992), Direr (2001), and Hopkins–Kornienko (2004)). Our model of status is in the ordinal tradition.

[2]This should be contrasted with the purely instrumental role status might play, for instance when higher consumption signals higher wealth and hence eligibility as a marriage partner (see e.g., Cole–Mailath–Postlewaite (1992, 1995, 1998) and Corneo–Jeanne (1998)). Like the authors in the previous footnote, we take seriously the value of status to people, in and of itself, even if it never leads to any other benefit. The historical origins of feelings of status are now lost. It may even be that in the distant past status was purely instrumental, but gradually became internalized as a value in itself. In the *Genealogy of Morals*, Nietzche claims that conscience arose in a similar way: people who broke promises were severely punished, leading to the birth of guilt.

random) scores for each player. For simplicity, we focus on additive status, in which a player gains one utile for each opponent he outranks and loses one utile for each opponent who outranks him.

The designer defines a different game according to how he clumps scores into grades, coarsening the ranking. There are many possible grading schemes, and we look for those that elicit maximal effort.

The advantage of coarse grading can most succinctly be seen with two students α and β who have *disparate* abilities, so that α achieves a random but uniformly higher score even when he shirks and β works.[3] Suppose, for example, that β scores between 40 and 50 if he shirks, and between 50 and 60 if he works, while α scores between 70 and 80 if he shirks and uniformly between 80 and 90 if he works. With perfectly fine grading, α will come ahead of β, regardless of their effort levels. Since they care only about rank, *both* will shirk.

But, by assigning a grade A to scores above 85, B to scores between 50 and 85, and C to scores below 50, the professor can inspire β to work, for then β stands a chance to acquire the same status B as α, even when α is working. This in turn generates the competition which in fact spurs α to work, so that with luck he can get an A and distinguish himself from β. Notice that very coarse grading (giving everyone an A) would not elicit effort since then nobody has anything to gain by improving his score. Optimal grading must be coarse, but not too coarse.

Coarse grading is also useful when students are homogeneous (*ex ante* identical). For example, suppose each student scores according to the normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ if he works, and according to $N(\hat{\mu}, \hat{\sigma})$ if he shirks, where $\mu > \hat{\mu}$ and $\sigma < \hat{\sigma}$. It is intuitively evident that an extraordinarily high score is more likely to come from a lucky shirker than from a worker. We show that the optimal grading scheme gives the same grade A to all scores above some threshold x_A, and is perfectly fine for scores less than x_A.

Coarse grading no doubt reduces the screening content delivered by schools. But our analysis reveals that if the schools sought to convey more information about the quality of their students, they would produce students of lower quality![4]

Our analysis presumes that each student knows his own ability and that the students and the professor all know the *distribution* of abilities in the class. (They do not necessarily know which student has which ability.) By virtue of repeated meetings of the class, or similar classes held over many years, it is not unreasonable to suppose that this distribution can be fairly well estimated by the professor and the students alike.[5]

---

[3]The hypothesis of disparate abilities is strong, but not as strong as it seems, and can be plausibly interpreted. For example, one might imagine that students have many effort levels, and that when the alpha students exert their second best effort they will come ahead of the beta students, no matter how hard the betas work or how lucky they get. If the professor wants to motivate each student to do his *very* best, then our analysis still applies.

[4]We take the "quality" of a graduating student to depend on both his (innate) ability and on how hard he studied.

[5]Moldovanu, Sela and Shi (2005) take our model and reconsider our results, replacing our hypothesis that the *distribution* of abilities in the actual class is known with the incomplete information hypothesis that student abilities are independently drawn from that distribution, so that the dis-

It should be emphasized that coarse grading does not involve what are commonly called handicaps. Handicaps discriminate between contestants by bestowing an advantage on the weak. Handicaps thus presume knowledge of individual contestants' abilities, as well as the "legality" of the discrimination. The grading we describe in this paper is, in contrast, required to be completely anonymous in that grades depend only on the exam scores of the students and not on their names. It is also required to be monotonic in the scores: if a student gets a better score than another, he is awarded at least as good a grade. On either count, handicaps are ruled out since they would necessarily entail an artificial boost to the score/grade of the weak student.

In Section 3 we characterize the optimal grading scheme for an arbitrary number of students of disparate abilities. Our first and most important conclusion is that in order to create the largest incentives to work, the professor should always use coarse grading. Our second conclusion is that optimal grading creates small elites, excluding many from membership who have equal abilities and have also worked hard but have been unlucky in the scores realized. In a population made up of equal numbers of students of three disparate abilities, say alpha and beta and gamma, fewer A grades will be given than B's, and fewer B's will be given than C's. In particular, though they all work hard, only some alphas get A and only some betas get B. If less able students have higher costs from studying hard, as Spence (1974) suggested, then the pyramiding becomes still more extreme.

In Section 4 we provide criteria for an optimal grading scheme when students are homogeneous. The key analytical concepts in this analysis are stochastic dominance and uniform stochastic dominance. We show that if a partition of scores into cells (each cell signifying a distinct grade) is optimal, then the shirker's performance (stochastically) dominates the worker's inside each cell; while across cells the worker's uniformly dominates the shirker's. Using this condition, we precisely characterize the optimal grading scheme for generic score densities. We find that fine partitions are typically not optimal, though under certain circumstances they could be.

In moving from scores to grades, professors can grade on an absolute scale (say 85 to 100 is an A) or "on a curve" (say the top 10% get an A). Given that the students only care about their relative rank, which kind of grading is better? We show in Section 5 that if the students are disparate or homogeneous, then absolute grading is always better than grading on a curve.[6] (For instance, in the example of two disparate students α and β, grading on a curve provides no incentives whatsoever.)

The inferiority of grading on a curve is surprising, especially since it is so commonly used in practice. One explanation is that professors fear damaging their reputation if their grade profile differs too much from the school norm. Another possibility is that our theorem is no longer valid if the professor is significantly uncertain about

tribution of abilities in the actual class may be different. With a continuum of students, which we sometimes assume, the two hypotheses are the same. Moreover, with absolute grading and additive status, which we concentrate on, our analysis covers the incomplete information case as well (as explained in Section 3.4). Only when the student population is small, and the professor grades on a curve, will there be a difference between complete and incomplete information.

[6] This principle may be valid with heterogeneous students, but we leave its exploration for future research.

the distribution of students' abilities, or if their scores are correlated, so that they might all do well or all do badly depending on how well designed the exam is. We leave these cases for future work.

Status pertains to situations far more general than grading exams. In Section 6 we analyze grading in games of money and status and apply it to the classical problem of designing the optimal wage schedule. We show that an employer who recognizes that his workers regard superior wages as an indication of status, over and above the direct utility from consuming wages, will be able to get them to work harder and pay them less money. To minimize his cost, the employer must combine both the status incentives and the consumption incentives of wages. Doing so changes the classical wage schedule, and also produces wage-grades that differ from the pure status grades used by the professor.

If the workers are of many disparate abilities, the classical employer would not need to choose wages that rise faster than the disutility of effort. But with status in the picture, he will always pay an exorbitant salary to a tiny elite who perform the best.

The combination of wages and status thus provides a new explanation for the astronomical pay we often see at the top of some corporate hierarchies.

In general, the highest and lowest ability worker will be motivated more by money, and the intermediate worker relatively more by status. The fine structure of the optimal wage schedule in between depends on the distribution of abilities. If the distribution is bell-shaped then, as we go up the ability ladder, wage differentials first diminish and later increase. If it is downward sloping (i.e., the density of workers falls as ability rises), then wage differentials increase. This is so even though productivity is linear in total effort.[7]

We also characterize the optimal wage schedule when workers are homogeneous. If risk neutral workers have just two possible effort levels, the classical employer would pay a lump sum to output above some suitable threshold and nothing below. The status conscious employer would in addition distinguish the lower outputs by slight wage differentials, creating wage grades identical to those which arise from pure status. By bringing status into play, he would lower his total wage bill.

2 Games of Status

In this section we precisely define what we mean by games of status, and the freedom the principal has to create grades.

Imagine a set N of students who are taking an exam. Depending on their effort levels $(e_n)_{n\in N}$, they will get exam scores, $(x_n)_{n\in N}$, which might also depend on random events, such as whether they were lucky enough to have studied the material precisely relevant to the questions, or how they felt that day, or how accurately the

[7] These conclusions depend on the assumption that the disutility of effort does not vary much across workers. Linearity in production rules out the standard explanations of increasing wage differentials based on diminishing marginal productivity in the inputs of different kinds of labor, and scarcity of workers at the top. The phenomenon is sustained here by status alone.

professor corrected the exams. It is natural to assume that a student's score does not depend on others' efforts, but actually none of our mathematics requires it.[8] Given the exam scores $x = (x_n)_{n \in N}$, the professor must assign grades $\gamma(x)$. Students are assumed to care only about status (and not about the education they are getting). We capture this by assuming that they obtain 1 utile for each student whose *grade* is strictly lower, and they lose 1 utile for each student whose grade is strictly higher.[9]

We suppose that the students are told in advance how the professor converts scores to grades, i.e., they know γ. Absolute grading is achieved by specifying intervals of scores corresponding to each grade, say $[85, 100]$ gives A, $[70, 85)$ gives B, and so on. Grading on a curve is based in contrast on relative performance alone, for example, that the top 10% of students get A, the next 20% get B's, and so on. Absolute and relative grading are quite different, though both are widely used.

What grading scheme γ *should* a professor use, if he wants to incentivize (whenever feasible) *all* his students to put in maximal effort?[10] No matter what scheme he chooses, and no matter what efforts the students put in, total utility awarded via grades will be zero, since for every utile gained by a higher-ranked student, there is a utile lost by a lower-ranked student. Indeed when all students work hard, their total net utility is minimized (since work inflicts disutility). Status seeking is the ultimate rat race!

Nevertheless, by the right choice of γ, the professor can often motivate his status-conscious students into working hard, and thus willy-nilly becoming educated.

2.1 The Performance Map

The *strategy set* $E_n \subset \mathbb{R}_+$ of each student $n \in N$ consists of a set of effort levels that are w.l.o.g. identified with the disutility they inflict on n. Efforts lead to (random) performance scores. For $x \in \mathbb{R}^N$, the nth-component x_n of x represents the score (output) obtained by n. Let $E \equiv \times_{n \in N} E_n$ and let $\Delta(\mathbb{R}^N)$ be the set of probability distributions on $\mathbb{R}^N$. The *performance map*

$$\pi : E \to \Delta(\mathbb{R}^N)$$

associates stochastic scores with effort levels. Here $\pi(e)$ gives the probability distribution of score vectors when the students put in effort $e \in E$.[11] We allow for the possibility that n's effort e_n might affect the score x_m of other students $m \neq n$.

[8] When the score x_n of one player depends (perhaps negatively) on the effort e_m, $m \neq n$ of another player, we can reinterpret our model as a parlor game.

[9] This is to keep matters simple. A "harmonic" utility might give $1/n$ utiles to a student who alone has rank n, and $(1/n + \cdots + 1/(n+m-1)) \cdot (1/m)$ utiles to each of m students who rank nth. Coming first instead of second provides a much bigger gain in utility than moving from 27th to 26th. Both additive and harmonic utilities are instances of "positional" status, that reward a player solely on the basis of his own position in the hierarchy.

[10] We could have considered other goals, like what grading scheme would give the highest expected total score, even when it is not feasible to induce all students to exert full effort. The results would have a similar flavor, but we leave them for future research.

[11] In the natural case (see our examples), higher effort levels tend to improve scores in the sense of first-order stochastic dominance.

2.2 Grading

Let $\mathcal{R}$ denote all possible orderings of N with ties allowed. There is a *grading map*

$$\gamma : \mathbb{R}^N \to \mathcal{R}$$

which ranks students according to $\gamma(x)$ when the scores obtained are $x \in \mathbb{R}^N$. Each rank corresponds to a grade. Coarse grading pools different scores into the same rank. We consider, in principle, only maps γ that are anonymous and monotonic: the grades depend on the scores, not on the names, and a higher score implies at least as high a grade.[12] Our focus will be on two particular ways of generating γ.

2.2.1 Absolute Grading

Let $\mathcal{P}$ be a partition of $\mathbb{R}$ into consecutive intervals, each of which has nonempty interior and some of which are designated "fine." When an interval[13] $[a, b)$ is designated fine, it is taken to represent the partition $\{\{x\} : x \in [a, b)\}$ consisting of singleton *cells*. An interval $[a, b)$ not so designated will signify the standard unbroken interval, and will also be called a *cell* in the partition $\mathcal{P}$.[14]

Fix a partition $\mathcal{P}$ as above. Then for any $a, b \in \mathbb{R}$ we define $a \succ_{\mathcal{P}} b$ iff the cell in $\mathcal{P}$ containing a lies strictly above the cell in $\mathcal{P}$ containing b. This leads to an absolute grading $\gamma_{\mathcal{P}} : \mathbb{R}^N \to \mathcal{R}$ where $i \succ_{\gamma_{\mathcal{P}}(x)} j$ iff $x_i \succ_{\mathcal{P}} x_j$. Thus $\gamma_{\mathcal{P}}(x)$ coarsens the information in x, creating ties between players whose scores lie in the same cell of $\mathcal{P}$.

2.2.2 Grading on a Curve

Given scores $x = (x_n)_{n \in N} \in \mathbb{R}^N$, define the class rank $\rho_n(x)$ of each student n by

$$\rho_n(x) = \#\{j \in N : x_j > x_n\} + 1.$$

Several students may have the same class rank. We define the "grading curve" Q to be a consecutive partition of class ranks $\{1, 2, ..., |N|\}$. For any $x \in \mathbb{R}^N$, let $\gamma_Q(x)$ be given by

$$i >_{\gamma_Q(x)} j \text{ iff } \rho_i(x) <_Q \rho_j(x).$$

This defines the grading map $\gamma_Q : \mathbb{R}^N \to \mathcal{R}$.

In words, a grading curve is defined by the number n_A of students getting A, the number n_B getting B, and so on. The grades are obtained by ranking student exam scores, and taking the top n_A scores and giving all the students who got them A. If $k > n_A$ students tie with the top score, then all must get A, and the number of B's is diminished by the excess A's, and so on.

[12]Furthermore, if $y_i = x_i$ for all $i \in N \backslash \{j\}$, and $y_j \geq x_j$, then the relative rank of j versus any $i \neq j$ in $\gamma(y)$ is no worse than in $\gamma(x)$.

[13]We use $[a, b)$ as a proxy for $[a, b)$, $(a, b]$, (a, b) or $[a, b]$. Our analysis works equally in all cases.

[14]Recall that students care only about their relative grade in the class. The professor could *ex ante* fix a different letter grade for each cell. Equivalently, he could wait until the realization of exam scores, and *ex post* assign the letter grade A to the highest cell that includes at least one student's score, a B to the next highest cell that includes at least one score, and so on. That way some student always gets an A, and the number of grades never exceeds the number of students in the class.

2.3 Utilities

The *exam payoff* to a student n from being ranked according to $R \in \mathcal{R}$ is

$$\#\{j \in N : n >_R j\} - \#\{j \in N : j >_R n\}$$

reflecting the fact that n gets a utile for each student he beats, and loses a utile for each student who beats him. He cares about (ordinal) status.

Note that it is not necessarily the case that a higher expected score means a higher exam payoff. Coming behind by a lot with probability .49 and coming ahead by a little with probability .51 yields positive exam payoff to the student with the lower expected score.

A student n who exerts effort $e_n \in E_n$ and obtains ranking $R \in \mathcal{R}$ gets net utility:

$$\#\{j \in N : n >_R j\} - \#\{j \in N : j >_R n\} - e_n$$

Notice again that the student is indifferent to learning. Had he put value on it, our task of incentivizing him to work would have been much simpler.

2.4 The Game Γ_γ

Fix a grading function $\gamma : \mathbb{R}^N \to \mathcal{R}$. Then, given effort levels $e \equiv (e_k)_{k \in N} \in E$, the *payoff* to $n \in N$ is his expected net utility:

$$\text{Exp}_{\pi(e)}[\#\{j \in N : n >_{\gamma(x)} j\} - \#\{j \in N : j >_{\gamma(x)} n\}] - e_n \equiv u^n_\gamma(e) - e_n.$$

Here $\text{Exp}_{\pi(e)}$ denotes expectation w.r.t. the distribution $\pi(e)$ over scores $x \in \mathbb{R}^N$ and $u^n_\gamma(e)$ denotes the *expected exam payoff* to n.

For $e \equiv (e_k)_{k \in N} \in E$ and $n \in N$, denote $e_{-n} \equiv (e_k)_{k \in N \setminus \{n\}}$. Recall that e is a *Nash equilibrium* (NE) of Γ_γ if the payoff each student n gets under e is at least as good as the payoff under (e'_n, e_{-n}) for all $e'_n \in E_n$.

Let $\tilde{e} \equiv (\tilde{e}_n)_{n \in N}$ be the strategy profile of *maximal effort*:

$$\tilde{e}_n = \max\{e'_n : e'_n \in E_n\}.$$

The key concern of our analysis is to design γ so as to ensure that $\tilde{e}$ is an NE — hopefully the unique NE — of the game Γ_γ; or, even more, an NE in weakly dominant strategies.

2.5 Optimal Grading

We shall concentrate on the case of two effort levels: high (work) H_n and low (shirk) L_n, for each agent n. Let $d_n = H_n - L_n$. We shall say that γ is *efficient* (within a given class of grading schemes) if it supports work as a Nash equilibrium when students have disutility $d = (d_1, ..., d_n)$, and if there is no other grading scheme γ' (in that class) that can support work as a NE with disutilities $d' \gneqq d$. Since we are especially interested in the case where disutilities are unobservable, we shall focus on *maxmin* grading schemes, i.e., those that satisfy the requirements of efficiency when d and d' are restricted to be symmetric vectors (of the form $(\lambda, ..., \lambda)$). We call maxmin schemes *optimal* if they are also (genuinely) efficient. Our analysis centers on the class of absolute grading schemes, so both efficiency and optimality will be understood to pertain to this class, unless otherwise stated.

2.6 Injecting Randomization

We could also introduce randomness in γ without violating monotonicity or anonymity of the grading scheme. For example, the professor could announce that he will flip a coin just before grading the exam: if heads he will take the interval $[86, 100]$ to be an A, while if tails he will count any score in the interval $[84, 100]$ as an A. We will not investigate random grading because we shall assume that student performances already contain noise. Were the performance map π deterministic, random grading would be needed to induce maximal effort, as shall become evident.

Adding noise to scores does make γ random, but it violates monotonicity.[15]

3 Disparate Students

3.1 Coarsening

We begin with the simplest example, illustrating the benefits of coarse grading.

First suppose $N = \{\alpha, \beta\}$, i.e., there are just two students. Student n obtains marks uniformly distributed on the interval J_H^n when he works and J_L^n when he shirks. His score depends only on chance and on his own effort. We assume the students have *disparate* abilities: $J_L^\beta < J_H^\beta < J_L^\alpha < J_H^\alpha$, i.e., α is so much more able than β, that he always comes out ahead even when he shirks and his rival works. (See Figure 1.) Thus if the professor were to grade them finely, neither would work, since status could not be affected by effort. More precisely, (L_α, L_β) is the unique NE of the game $\Gamma_{\gamma_{\widetilde{\mathcal{P}}}}$ where $\widetilde{\mathcal{P}} \equiv \{\{x\} : x \in \mathbb{R}\}$ denotes the finest partition — even more, it is an NE in strictly dominant strategies.

The professor can do better with a judiciously chosen coarse partition $\mathcal{P}$. Indeed consider the partition $\mathcal{P}(p) \equiv \{A, B, C\}$ shown in Figure 1. Anything below J_H^β gets grade C (including all scores in J_L^β obtained when the beta type shirks). All scores in J_H^β and J_L^α get B, as well as the bottom $(1 - p)$ fraction of the scores in J_H^α. The partition is completely characterized by the single parameter $0 \leq p \leq 1$, specifying the fraction of J_H^α that counts for the grade A (so that we may abbreviate $\gamma_{\mathcal{P}(p)} \equiv p$, without confusion).[16]

[15] In Dubey–Wu (2001) and Dubey–Haimanko (2003), noise was introduced by varying the sample size on the stream of outputs produced by agents, which maintains monotonicity conditional on the observed sample, but not with respect to the raw outputs.

[16] The randomness in scores if β works or shirks, or if α shirks, is irrelevant to this example. All that matters is that the distribution of scores if α works be continuous. In this example, and in our treatment of the general disparate case, if the assumed randomness were deterministic instead, we could still achieve the same grading by randomizing the grading, e.g., in this example, the cutoff to get an A.

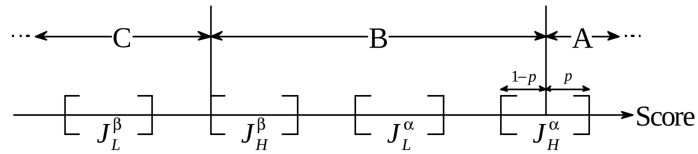

Figure 1: The Partition $\mathcal{P}(p)$

The (grade) incentive $I^n(p)$ to switch from effort level L_n to H_n for any student n (assuming that his rival is working hard) is given by:

$$\begin{aligned} I^\alpha(p) &= u_p^\alpha(H_\alpha, H_\beta) - u_p^\alpha(L_\alpha, H_\beta) \\ I^\beta(p) &= u_p^\beta(H_\alpha, H_\beta) - u_p^\beta(H_\alpha, L_\beta). \end{aligned}$$

It is easy to compute that

$$I^\beta(p) = -p - (-1) = -p + 1$$

and

$$I^\alpha(p) = p - 0 = p.$$

Denote $d_n = H_n - L_n$, i.e., d_n is n's disutility for switching from shirk to work. Then (H_α, H_β) is a Nash equilibrium if and only if $I^\beta(p) = -p+1 \geq d_\beta$ and $I^\alpha(p) = p \geq d_\alpha$.

Fix the disutilities $d = (d_\alpha, d_\beta)$. The efficient p is defined by

$$p^* = \arg\max_{0\leq p\leq 1}\{\lambda : I^\alpha(p) \geq \lambda d_\alpha, I^\beta(p) \geq \lambda d_\beta\} = d_\alpha/(d_\alpha + d_\beta).$$

If the students have the same disutilities, i.e., $d_\alpha = d_\beta = d$, the optimal $p^* = 1/2$ is given by

$$p^* = \arg\max_{0\leq p\leq 1} \min\{I^\alpha(p), I^\beta(p)\} = 1/2.$$

Note that $I^n(1/2) = 1/2$ for both students n.

Multiple Effort Levels and Less Disparateness The hypothesis of disparate students is not as strong as it seems. One may imagine that each student has several effort levels and that J_L^n is the performance interval for $n \in N$ when n exerts his *second-highest* effort. Now the two students are not as heterogeneous as before: all we are postulating is that α is sufficiently more able than β so that his second-highest effort leads to uniformly better scores than β's highest effort. (The term $d_n = H_n - L_n$ must be interpreted as the extra disutility incurred when n switches from his second-highest to his highest effort.) In this setting, it is harder to sustain maximal effort as an NE (more conditions will have to be met), and our analysis gives only *necessary* conditions. It shows that any partition that induces both agents to work their hardest must pool part of J_H^α with part of J_H^β.

3.2 Pyramiding

Notice that the optimal grading partition, given by $p^* = 1/2$, implies:

$$\begin{aligned}\text{Expected \# of students getting } A &= p^* = \frac{1}{2} \\ \text{Expected \# of students getting } B &= 1 + (1 - p^*) = \frac{3}{2}.\end{aligned}$$

In other words, optimal grading creates a pyramid with fewer expected A's than B's even though there are equal numbers of strong and weak students in the class.

Spence (1974) postulated that typically the weak student incurs more disutility from effort than the strong, i.e.,

$$d_\beta > d_\alpha.$$

It is evident that the Spence condition has the effect of accentuating the pyramid, since $p^* = d_\alpha/(d_\alpha + d_\beta)$ falls as d_β rises, diminishing the expected number of A's and increasing the B's.

3.2.1 Multiple Students of Each Type

Now we show that coarsening and pyramiding persist with many students of each type. Suppose there are N β-type students of low ability and K α-type students of high ability. The reader can check that the incentive functions become:

$$\begin{aligned}I^\beta(p) &= -pK - (-(N + K - 1)) \\ &= (-\mu^H p + 1)\delta\end{aligned}$$

where $\mu^H \equiv K/(N+K-1)$ gives the fraction of high ability in the population, when a single low-ability student stands aside; and $\delta \equiv N + K - 1 \equiv$ utiles to a student when he beats all the others.[17] Similarly, one can compute

$$I^\alpha(p) = \delta p.$$

Assuming $d_\alpha = d_\beta$, the optimal (maxmin) $p = 1/(1 + \mu^H)$ is obtained by solving $-\mu^H p+1 = p$. When K and N are large and equal, μ^H is nearly 1/2, and the optimal p converges to 2/3. The pyramid remains. Indeed, the pyramid becomes more visible since expected # of students getting A $\approx$ actual # of students getting A, etc., by the law of large numbers.[18]

[17] The status incentive grows *pro forma* with the population. One could rescale status, dividing by δ, without affecting our analysis, when comparing different populations, so that rank is given in terms of population percentile. Is there 100 times more status in being the President of India's 1 billion than of Greece's 10 million?

[18] Observe that if the population changes to include more α-type students, this will *lower* the fraction of the α-type students who get A. (Recall that all the β-type get B.) This is so since $p = 1/(1 + \mu_H)$ is decreasing in μ_H. It is also interesting to observe that so long as there is at least one β student, i.e., $N \geq 1$, the proportion of A's in the whole population is always less than 1/2, since $p\mu_H = \mu_H/(1 + \mu_H) < 1/2$.

Thus we have found a coarse partition producing an absolute grading scheme that gives incentive of $(N+K-1)(1/(1+\mu_H)) = (N+K-1)^2/(N+2K-1)$ to each agent. We shall now prove that no other monotonic and anonymous scheme could do better. In particular, grading on a curve cannot do better than the best absolute grading scheme.

Theorem 1: *Suppose there are N β-type students of low ability and K α-type students of high ability. Let $p^* = (N+K-1)/(N+2K-1)$. The absolute grading partition $\mathcal{P}(p^*)$ is optimal in the class of all anonymous, monotonic grading schemes.*

Proof: Consider an arbitrary monotonic and anonymous grading scheme. Let the expected (exam) payoff to each α student, if all work, be a. (By anonymity each α student must have the same expected payoff.) The expected payoff to each β student must be $-(K/N)a$, since the total status payoff is always zero. Since the α students come ahead of the β students, monotonicity implies that $a \geq 0$.

The incentive to work for a β student is at most $-(K/N)a - (-(N+K-1)) = (N+K-1) - (K/N)a$. By monotonicity, the incentive to work for an α student is at most $a - ((K-1)/K)(-(K/N)a) = a + ((K-1)/N)a = ((N+K-1)/N)a$. The reason is that even when the α student shirks, his expected payoff against the β students is at least zero. His expected score against the other $K-1$ α students is at worst $((K-1)/K)$ multiplied by the score a β student got when all K alpha students were working.

Thus the maxmin incentive is at best $\max_a \min\{N+K-1-(K/N)a, ((N+K-1)/N)a\}$, which is achieved for $a = N(N+K-1)/(N+2K-1)$, giving incentive $(N+K-1)^2/(N+2K-1)$, which is achieved via $\mathcal{P}(p^*)$. ■

3.3 Many Disparate Student-Types

When there are ℓ types, the optimal absolute grading partition will entail $\ell+1$ letter grades (i.e., will divide the numerical score line into $\ell+1$ consecutive cells). Each type i will have a positive probability $0 < p_i \leq 1$ of obtaining grade i if he works; but will lapse into the lower grade $i-1$ with certainty if he shirks.

We illustrate the case of three disparate types: 1 (low ability), 2 (middle ability), 3 (high ability) in Figure 2.

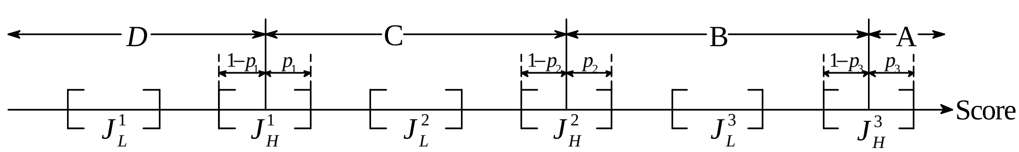

Figure 2. The Partition $\mathcal{P}(p_1, p_2, p_3)$

Suppose there are $N_1, ...N_\ell$ students of type $i = 1, ..., \ell$. Given the grading partition $p = (p_1, ..., p_\ell)$, the incentive to work for the ℓ types is

$$\begin{aligned} I_\ell^1(p) &= p_1[(N_1 - 1) + (1 - p_2)N_2] \\ I_\ell^i(p) &= p_i[(N_i - 1) + p_{i-1}N_{i-1} + (1 - p_{i+1})N_{i+1}] \text{ for } 2 \le i \le \ell - 1 \\ I_\ell^\ell(p) &= p_\ell[(N_\ell - 1) + p_{\ell-1}N_{\ell-1}]. \end{aligned}$$

When working, a student of type $2 \le i \le \ell - 1$ might get unlucky, with probability $1 - p_i$, and find himself no better off than if he shirked. But with probability p_i he will be lucky, beating the fraction p_{i-1} of type $i - 1$ he otherwise would be equal with, and coming equal with the fraction $1 - p_{i+1}$, of type $i + 1$ he would otherwise have lost out against. In addition, he either beats (instead of equalling) or equals (instead of losing to) every student of his own type. This gives the formula $I_\ell^i(p)$ for $2 \le i \le \ell - 1$. Taking $N_0 = N_{\ell+1} = 0$ gives the formulas for $I_\ell^1(p)$ and $I_\ell^\ell(p)$.

When there are vastly more students of some types than others, an optimal partition will not necessarily equalize all the incentives. For example, suppose there are one billion students of the lowest type 1, and just two students of types 2 and 3. An efficient partition will always set $p_1 = 1$, giving an incentive to work of at least one billion (minus one) to type 1 students. A top student (of type 3) is only competing against the students of type 3 and 2, and can therefore never have incentives exceeding three utiles. This also shows that maxmin grading schemes need not be unique, since choosing $p_1 < 1$ (but not too small) will also achieve the maxmin.

Surprisingly, if $1 \le N_1 \le \cdots \le N_\ell$, there will be a unique maxmin partition, and it will indeed equalize all the incentives, and be optimal. Furthermore, it will generate pyramiding. Indeed, each student of type $i > 1$ will have positive probability of getting a grade lower than his type.

Theorem 2: *Let $1 \le N_1 \le \cdots \le N_\ell$. Then $I_\ell \equiv \max_{p \in [0,1]^\ell} \min_{1 \le i \le \ell} I_\ell^i(p)$ is achieved at a unique $\bar{p}$; moreover, $\bar{p}_1 = 1$ and $0 < \bar{p}_i < 1$ for $i = 2, ..., \ell$, and all agents have the same incentive: $I_\ell^i(\bar{p}) = I_\ell \; \forall i = 1, ..., \ell$. Therefore $\mathcal{P}(\bar{p})$ is optimal in the class of all absolute grading schemes. Furthermore, there is a grading pyramid: the ratio of students obtaining the highest grade to the number of top students is equal to $\bar{p}_\ell < 1$, whereas the ratio of students getting the lowest observed grade to the number of bottom students is $\bar{p}_1 + (1 - \bar{p}_2) = 1 + (1 - \bar{p}_2) > 1$.*

Finally, consider an infinite sequence of disparate types with populations $1 \le N_1 \le N_2 \le \cdots$. For each ℓ, let I_ℓ be the maxmin incentive for the status game with types $1, ...\ell$, as above. Then I_ℓ is monotonically increasing in ℓ, converging to $I^* \le N_1 + N_2 - 1$ as $\ell \to \infty$.

Proof: Since each $I_\ell^i(p)$ is continuous in p, $I_\ell(p)$ is also continuous, and so $I_\ell = \max_{p \in [0,1]^\ell} \min_{1 \le i \le \ell} I_\ell^i(p)$ is achieved at some $\bar{p}$. Clearly any maxmin $\bar{p} \gg 0$, for otherwise $I_\ell = I_\ell(\bar{p}) = 0$, which can be bettered by choosing all $p_i = 1$.

Inspection of the formulae immediately reveals that raising $\bar{p}_i$ raises $I_\ell^i(\bar{p})$ and $I_\ell^{i+1}(\bar{p})$, but lowers $I_\ell^{i-1}(\bar{p})$. Furthermore, for any $2 \le i \le \ell$, if $\bar{p}_i = 1$, then from

$N_1 \leq \cdots \leq N_\ell$ we get $I^i_\ell(\bar{p}) \geq N_i - 1 + \bar{p}_{i-1}N_{i-1} \geq \bar{p}_{i-1}[\bar{p}_{i-2}N_{i-2} - 1 + N_{i-1}] = \bar{p}_{i-1}[N_{i-1} - 1 + \bar{p}_{i-2}N_{i-2} + (1-\bar{p}_i)N_i] = I^{i-1}_\ell(\bar{p})$, where the second inequality is strict if $\bar{p}_{i-1} < 1$.

Now we argue that for any maxmin $\bar{p}$, $I^i_\ell(\bar{p}) = I_\ell$ for all $i = 1, ..., \ell$. Take any maxmin $\bar{p}$ with the fewest number of coordinates i with $I^i_\ell(\bar{p}) = I_\ell$. Suppose i is the largest coordinate with $I^i_\ell(\bar{p}) = I_\ell$. If $i < \ell$, then $I^j_\ell(\bar{p}) > I^i_\ell(\bar{p})$ for all $j > i$. Lowering $\bar{p}_{i+1}$, which is possible since $\bar{p} \gg 0$, raises $I^i_\ell(\bar{p})$, and lowers the irrelevant $I^{i+1}_\ell(\bar{p})$ and $I^{i+2}_\ell(\bar{p})$. This either raises I_ℓ or reduces the number of i at which I_ℓ is attained, a contradiction either way. Hence $I^\ell_\ell(\bar{p}) = I_\ell$. Suppose $I^{i-1}_\ell(\bar{p}) > I^i_\ell(\bar{p}) = I_\ell$, for some $i = 2, ..., \ell$. Then from the last line of the last paragraph, $\bar{p}_i < 1$. But then raising $\bar{p}_i$ raises $I^i_\ell(\bar{p})$ and $I^{i+1}_\ell(\bar{p})$, lowering the irrelevant $I^{i-1}_\ell(\bar{p})$. This either raises I_ℓ or reduces the number of i at which I_ℓ is attained, a contradiction either way. Thus $I^i_\ell(\bar{p}) = I_\ell$ for all i and any maxmin $\bar{p}$.

Now we show that I_ℓ is achieved at a unique $\bar{p}$. Observe first that at any maxmin $\bar{p}$, $\bar{p}_1 = 1$, for if $\bar{p}_1 < 1$, increasing $\bar{p}_1$ will increase $I^1_\ell(\bar{p})$ without lowering any other $I^i_\ell(\bar{p})$, contradicting $I^1_\ell(\bar{p}) = I_\ell$ for every maxmin $\bar{p}$. But $\bar{p}_1 = 1$ and $I^1_\ell(\bar{p}) = I_\ell$ uniquely determines $\bar{p}_2$. But then $\bar{p}_1$, $\bar{p}_2$, and $I^2_\ell(\bar{p}) = I_\ell$ uniquely determines $\bar{p}_3$, and so on.

Observe that if $\bar{p}_1 = \bar{p}_2 = 1$, then obviously $I^1_\ell(\bar{p}) < I^2_\ell(\bar{p})$, contradicting all $I^i_\ell(\bar{p}) = I_\ell$. This shows $\bar{p}_2 < 1$. We showed earlier that for any $3 \leq i \leq \ell$, if $\bar{p}_{i-1} < 1$ and $\bar{p}_i = 1$, then $I^i_\ell(\bar{p}) > I^{i-1}_\ell(\bar{p})$, contradicting their equality. Thus we have shown that $\bar{p}_i < 1$ for all $i = 2, ..., \ell$.

Finally, we claim that I_ℓ increases in ℓ, hence converging to some I^*. To verify this, let $I_\ell = I_\ell(\bar{p})$. Define $\hat{p} = (\hat{p}_1, ..., \hat{p}_\ell, \hat{p}_{\ell+1}) = (\bar{p}_1, ..., \bar{p}_\ell, 1)$. Then $I_{\ell+1} \geq I_{\ell+1}(\hat{p})$. But $I^i_{\ell+1}(\hat{p}) = I^i_\ell(\bar{p}) = I_\ell$ for all $i = 1, ..., \ell$. Moreover, $I^{\ell+1}_{\ell+1}(\hat{p}) = \hat{p}_{\ell+1}((N_{\ell+1}-1) + \hat{p}_\ell N_\ell) = 1 \cdot ((N_{\ell+1} - 1) + \bar{p}_\ell N_\ell) \geq \bar{p}_\ell(N_\ell - 1) + \bar{p}_{\ell-1}\bar{p}_\ell N_{\ell-1} = \bar{p}_\ell(N_\ell - 1 + \bar{p}_{\ell-1}N_{\ell-1}) = I_\ell$. But $I_\ell \leq I^1_\ell \leq N_1 + N_2 - 1$ for all ℓ. ■

3.4 A Continuum of Students and Incomplete Information

Fix ℓ and $N_1, N_2, ..., N_\ell$. We can consider optimal partitions for the population profile $(N_1 k, ..., N_\ell k)$, for $k = 1, 2, ...$ In the limit we effectively have a game of status with a continuum of players of type i with measure $\mu_i = N_i/(N_1 + \cdots + N_\ell)$. The entire analysis of Theorem 2 holds, with N_i and $N_i - 1$ replaced by μ_i. Such a game is of special interest because (when grading is absolute, and status is additive) it is equivalent to a game with "incomplete information."

Consider a variant of our games of status, in which there are N students, each of whom is drawn (independently) from a distribution of ℓ disparate abilities, with probabilities $(\mu_1, ..., \mu_\ell)$. Each student is informed of his type, but not of the others'. The number of students N_i that turn out to be type i is now random, and unknown to them and to the professor. How should the exam be graded?

It is easy to see that with additive status and risk neutrality, the finite-player incomplete-information game is equivalent to the continuum-player complete-information game we have already analyzed (after normalizing payoffs by population size). Each student optimizes the same way, whether he faces one student of type α and one of type β, or two students, each of whom has a 50–50 chance of being α or β.

Thus we have already derived the optimal absolute grading scheme for the incomplete information game. In order to compare incentives as $\ell \to \infty$, fix the measures $\mu_1 = \mu_2 = \cdots = \mu_\ell = 1$ of being of any type $1, ..., \ell$. For reasonably large ℓ, such as $\ell = 20$, the incentive $I_\ell \approx I^*$ is about 1.389 for each student. Since $p_1 = 1$, I^* is also the measure of students receiving the lowest grade $i = 1$: $I^1(\bar{p}) = \bar{p}_1(2 - \bar{p}_2) = (2 - \bar{p}_2) = 1 + (1 - \bar{p}_2)$.

In the table below we list the optimal $(p_1, ..., p_{20})$ and the measure of students for each grade $i = (1, ..., 20)$.

TABLE A. Pyramiding

Grade		Partition probabilities	Number of students in grade
Lowest	1	1	1.389726998
	2	0.610273002	0.887761199
	3	0.722511804	1.036040471
	4	0.686471333	0.9887908
	5	0.697680533	1.00352003
	6	0.694160503	0.998897996
	7	0695262507	1.000345096
	8	0.69491741	0.99989156
	9	0.69502585	1.000032891
	10	0.694992959	0.999986499
	11	0.69500646	0.999996886
	12	0.695009574	0.99997551
	13	0.695034064	0.999925241
	14	0.695108824	0.999760849
	15	0.695347975	0.999235638
	16	0.696112336	0.997568329
	17	0.698544007	0.992284376
	18	0.706259632	0.975830841
	19	0.730428791	0.927161102
Highest	20	0.803267689	0.803267689

Observe that there are more C's than B's, and more B's than A's, but for lower grades the number of students stay equal until the very bottom is reached. The bottom grade (aside from the failing grade $i = 0$ that nobody gets) $i = 1$ is the most commonly given.[19]

4 Homogeneous Students

Until now we have concentrated on the case where students differ substantially in their abilities. In that case, coarsening the grading allows the weaker student to compete with the stronger. We turn now to the case where all students have the same ability, and we show that coarsening still has a role to play.[20]

Homogeneity simplifies our task, because the incentives of all the players are aligned. Maxmin and optimal become identical. In Section 5 we shall prove that

[19] It is worth noting that with one (instead of a continuum) of students of each of three types, the optimal $(p_1, p_2, p_3) = (1, 1/2, 1)$ yielding incentive 1/2 to each, so that the expected number of A's = 1, of B's = 1/2 and of C's = 3/2, giving us pyramiding but not in the strongest sense. But even here, if we introduce the Spence condition $d_3 \ll d_2 \ll d_1$ on disutility of effort, the inequalities $I^i(p) \geq d_i$ will (as is obvious) require $p_3 < p_2 < p_1$ by way of a solution, bringing back the full pyramid.

[20] The case of *heterogeneous* students, with distinct but overlapping performances, is more complex to analyze. But the fundamental principle still holds that coarse grading often creates more incentives than perfectly fine grading. We illustrate this, via computer-generated examples, in Dubey–Geanakoplos (2004).

absolute grading gives better incentives than grading on a curve. Hence in this section we shall concentrate on absolute grading schemes, characterizing the optimal scheme within this class, when the densities of scores are regular. It is worth noting that homogeneity and additive status make absolute grading independent of how many students N there are.

4.1 Examples

Imagine an exam with two questions covering the two halves of the course. Suppose that if a student studies hard, he has probability $p = .6$ of getting any question right, independently across questions. If he shirks and studies only half the course, he has probability $q = .8$ of getting the corresponding question right, and zero chance of getting the other question. Thus the probabilities for getting $(0, 1, 2)$ questions right are $(.2, .8, 0)$ for the shirker and $(.16, .48, .36)$ for the worker. Clearly the hard working student will do better most of the time (his score stochastically dominates the shirker's score). How should the professor grade the exams?[21]

Fine grading gives the shirker an expected exam payoff of $(1-p)^2q - [p^2q + p^2(1-q) + 2p(1-p)(1-q)] = -0.328$. If both students work, then by symmetry and the fact that total exam payoff is inevitably zero, the expected utility of each is 0. The incentive to work with fine grading is thus 0.328.

Suppose instead that the professor uses just two grades, an A for a perfect exam, and a B for anything else. Then the expected exam payoff of a shirker is $-[p^2q + p^2(1-q)] = -p^2 = -0.36$. His incentive to work is thus 0.36, since again if they both work, each has an expected exam payoff of zero. Since $0.36 > 0.328$, we see that coarse grading gives higher incentives to work.

Before moving on to another example, observe that with only two students, grading on a curve either gives no incentive to work (when the curve has two A's), or else is identical to fine grading (when the curve has one A, one B). Hence in this example grading on a curve is worse than absolute grading.

Consider another situation in which N identical students take an exam. Suppose that if a student works hard, his score will be uniformly distributed on $[50\%, 100\%]$, that is, his score has density $f(x) = 2$ if $50\% \leq x \leq 100\%$, and 0 otherwise,independent of the others' scores and effort levels. If he shirks, suppose his score has density $g(x) = 2x$ for $0 \leq x \leq 100\%$, and 0 otherwise, again independent of the others.

[21] Suppose the shirker could study either half of the course, so the professor cannot distinguish the two students by attaching higher weight to the second question. We assume his grading depends only on the total number of correct answers of each student.

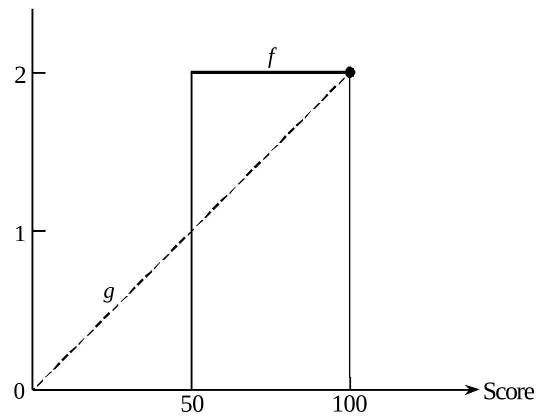

Figure 3: Score Densities

Assume all other students are working hard, and a single student is debating whether to work or shirk. If he works, his expected exam payoff will be 0. If he shirks, his expected exam payoff is equal to the probability he comes ahead of a worker, less the probability he comes behind, all multiplied by $N-1$.

The probability the shirker comes behind a worker is $\int_0^{1/2} 2xdx + \int_{1/2}^{1} 2x2(1-x)dx = 7/12$. The probability the shirker comes ahead of a worker is therefore $1 - 7/12 = 5/12$, and we conclude that shirking gives an expected exam payoff $(N-1)(\frac{5}{12} - \frac{7}{12}) = -\frac{1}{6}(N-1)$. This shows that the incentive to study hard is $\frac{1}{6}(N-1)$, which must be compared to the disutility of effort.

Suppose instead that just two grades are issued, namely A for scoring between 50% and 100%, and B for scoring between 0 and 50%. If a student works, along with all his $N-1$ rivals, then all will receive a score above 50% and therefore all will receive A. Each student will get a payoff of 0. If a single student fails to study, then his expected payoff is $(N-1)$ multiplied by $-1\int_0^{1/2} 2xdx + 0\int_{1/2}^{1} 2xdx = -1/4$ giving an incentive to study of $\frac{1}{4}(N-1)$.Since this is greater than $\frac{1}{6}(N-1)$, we see that giving only two grades creates significantly higher incentives to work than perfectly fine grading.

We will show later that our partition of scores

$$\mathcal{P} = \{[0, 50\%), [50\%, 100\%]\}$$

into just two grades yields the optimal absolute grading partition.

With three students the incentive to work under $\mathcal{P}$ is $(1/4)(3-1) = 1/2$. Now consider grading on a curve. Giving everybody an A provides no incentive at all. Giving three grades is just like the perfectly fine partition with absolute grading and is therefore not as good as the optimal partition $([0, 50), [50, 100))$. (Indeed we computed that fine grading gave incentive $1/3$). With a curve that gives two B's and one A, the incentive to work is $99/324 < 1/2$, while with a curve that gives one B and two A's it is $63/324 < 1/2$. Once again absolute grading is better than grading on a curve.

As a third example, suppose an exam contains K questions, and that a student who studies has (independent) probability p of getting each question right, while if he shirks the probability drops to $q < p$. It will turn out that the optimal grading is perfectly fine.

As a final example, suppose that we are grading the relative performances of two hedge funds. Suppose that a hedge fund that works on research will generate log returns normally distributed with mean μ and standard deviation σ, while a hedge fund that shirks will generate returns distributed according to $\hat{\mu} < \mu$ and $\hat{\sigma} > \sigma$. If the fund managers cared only about their relative grade, would they be motivated to work harder if the grade was simply their return? Intuitively that seems wrong, since an extraordinarily high return is more likely from the high variance shirker, despite his lower expected return. We shall see that the optimal grading scheme indeed does not reward higher returns after some point.

4.2 The General Theory with iid Students

We turn now to the general situation. Observe first that as long as the students are identical, their expected exam scores must be zero if they all work, as we noted in the examples. The incentive to work therefore is precisely the negative of the expected (exam) payoff to a shirker who competes against workers. It follows that there is *no* simplification gained by assuming that each student's performance is independent of the others' effort levels. If in the second example we continued to let (f, g) be the score densities of the (workers, shirker), and introduced $h \neq f$ as the score density when all work, our analysis would remain absolutely unchanged; h would be irrelevant.

On the other hand, the following independence assumption does play an important simplifying role.

Assumption: *Conditional on any choice of effort levels* $(e_1, ..., e_n)$, *students' exam scores are independent.*

We shall maintain this assumption for the rest of the paper.

Our analysis of the optimal absolute grading partition begins by asking whether the shirker's payoff can be lowered below zero by a single cut at θ, creating a two cell partition $\{(-\infty, \theta), [\theta, \infty)\}$. We shall find that if the worker's score distribution f stochastically dominates the shirker's score distribution g, then any cut will help, while under the reverse domination, every cut will hurt.

Next we consider a partition and ask whether cutting a cell in the partition into two cells will further help incentives or set them back. The answer depends on who is the better player, *conditional* on both scores lying inside the same cell. If the shirker is better, we must not reveal this, since we are trying to minimize his score, and keep the cell uncut. For example, the shirker may be very unlikely to get a score above 90. But conditional on both the shirker and worker getting above 90, it may be more likely that the shirker does better. (The shirker may have memorized the answers to last year's exam. In the unlikely event that this year's exam questions are the same he will get 100; otherwise he will get 0.)

The guiding principle in creating optimal partitions is to mask regions of the score space where the shirker is better than the worker, and to ensure that across cells the worker is better, so that the partition reveals the deficiencies of the shirker. This characterization will be used in Section 5 to prove that absolute grading is better than grading on a curve, and then again in Section 6 to derive the optimal wage schedule when workers are homogeneous.

Lemma 1: *Suppose two students H and L take an exam, yielding independent scores x_H and x_L. If the grading partition is $\{(-\infty,\theta),[\theta,\infty)\}$, then the expected exam payoff to L is*

$$P(x_L \in [\theta,\infty)) - P(x_H \in [\theta,\infty)).$$

Similarly, if the grading partition includes cells $[a,\theta)$, $[\theta,b)$, for $a<\theta<b$, then conditional on both x_H and x_L being in $[a,b)$, the expected exam payoff to L is

$$\frac{P(x_L \in [\theta,b))}{P(x_L \in [a,b))} - \frac{P(x_H \in [\theta,b))}{P(x_H \in [a,b))}.$$

Proof: In the first case, the expected exam payoff to L is

$$P(x_L \in [\theta,\infty) \wedge x_H \in (-\infty,\theta)) - P(x_H \in [\theta,\infty) \wedge x_L \in (-\infty,\theta)).$$

With independence, this becomes

$$\begin{aligned} & P(x_L \in [\theta,\infty))P(x_H \in (-\infty,\theta)) - P(x_H \in [\theta,\infty))P(x_L \in (-\infty,\theta)) \\ = \; & P(x_L \in [\theta,\infty))(1-P(x_H \in [\theta,\infty)) - P(x_H \in [\theta,\infty))(1-P(x_L \in [\theta,\infty))) \\ = \; & P(x_L \in [\theta,\infty)) - P(x_H \in [\theta,\infty)). \end{aligned}$$

The second case is analogous. ■

Corollary: *If $\mathcal{P}$ is a partition of scores including the cell $[a,b)$ and if $\mathcal{P}^*$ modifies $\mathcal{P}$ by cutting $[a,b)$ at θ, into $[a,\theta)$ and $[\theta,b)$, leaving all the other cells intact, then the move from $\mathcal{P}$ to $\mathcal{P}^*$ increases the expected exam payoff to L by*

$$P(x_L \in [a,b))P(x_H \in [a,b)) \left[\frac{P(x_L \in [\theta,b)}{P(x_L \in [a,b)} - \frac{P(x_H \in [\theta,b))}{P(x_H \in [a,b))}\right].$$

Proof: This follows from Lemma 1 after observing that if either $x_L \notin [a,b)$ or $x_H \notin [a,b)$, the payoff is the same under $\mathcal{P}$ or $\mathcal{P}^*$. ■

This should remind the reader of stochastic dominance.

4.2.1 Stochastic Dominance

Definition: We say that the *random variable* x *(stochastically) dominates the independent random variable* y on the interval $[a, b)$ if, whenever $P(x \in [a, b)) > 0$ and $P(y \in [a, b)) > 0$, we have

$$P(x \in [\theta, b) | x \in [a, b)) - P(y \in [\theta, b) | y \in [a, b)) \geq 0,$$

i.e.,

$$\frac{P(x \in [\theta, b))}{P(x \in [a, b))} \geq \frac{P(y \in [\theta, b))}{P(y \in [a, b))}$$

for all $\theta \in (a, b)$. In this case we write

$$x \succsim y \text{ on } [a, b).$$

If the inequality is strict for all $\theta \in (a, b)$, we write $x \succ y$ on $[a, b)$ and call it strict dominance. If $[a, b) = (-\infty, \infty)$, then we simply write $x \succsim y$ or $x \succ y$.

Stochastic dominance has an extremely important role to play in monotonic grading schemes, including absolute grading.

Lemma 2: *Suppose* $x \succsim y$. *Let the exam scores* x *and* y *be independent of the exam scores of every student* $n = 1, ..., N - 1$. *Let* γ *be any monotonic grading scheme for* N *students. Then the expected exam payoff to the last student* N *is at least as high under an exam score of* x *as it is under* y.

Proof: By independence, the payoffs from exam scores x and y depend only on their distributions. According to Theorem 1.A.1 of Shaked–Shanthikumar, there exist $\hat{x}$ and $\hat{y}$ with the same distributions as x and y respectively, such that $\hat{x} \geq \hat{y}$ with probability one. But then for *any* realization of the other $N - 1$ scores, $\hat{x}$ will clearly get a (weakly) higher payoff than $\hat{y}$. ∎

It follows that if a shirker has exam scores distributed according to x_L while the worker's is distributed according to x_H, and $x_L \gtrsim x_H$, then no monotonic grading scheme can provide any incentive to work. We might as well give all students an A.

In both our examples of this section, the worker scores stochastically dominated the shirker's scores, and indeed this is what gave the student an incentive to work under all the grading schemes.

It will be useful to also consider a strengthened form of domination.

Definition: We say that x *uniformly dominates* y on the interval $[A, B)$ if x dominates y on every subinterval $[a, b) \subset [A, B)$. In this case we write $x \succsim_U y$ on $[A, B)$.

Uniform domination can be characterized in terms of likelihood ratios in a manner that makes it much more handy to work with. But first we must restrict the random variables slightly.

Density Assumption: *Whenever we consider several random variables (exam scores) together, either all will be discrete or else all will be continuous, i.e., have measurable density functions on $(-\infty, \infty)$ with no atoms. In either case we can speak of the density $f(t)$, for $t \in (-\infty, \infty)$ or $t \in \{\alpha_1, \alpha_2, ..., \alpha_k\} \subset (-\infty, \infty)$.*

Lemma 3: *Let x and y be independent on $[A, B)$ with density functions f and g, respectively. Then x uniformly dominates y on $[A, B)$ if and only if the likelihood ratio $f(t)/g(t)$ is increasing almost everywhere on $[A, B)$, where $f(t)/g(t)$ can be defined suitably arbitrarily if $f(t) = g(t) = 0$.*

Proof: This follows from Theorem 1.C.2 in Shaked–Shanthikumar. ■

It is critical in understanding the first two examples of this section to observe that although the worker's scores (stochastically) dominates the shirker's, there are subintervals on which the shirker's score uniformly (stochastically) dominates the worker's. Thus in the first example, $q/(2p(1-p)) = .8/.48 > .2/.16 = (1-q)/(1-p)^2$, so on the cell $\{0, 1\}$ the shirker uniformly dominates the worker. In the second example of this section x_L uniformly dominates x_H on $[50, 100]$.

Another instance of uniform domination occurs in the third example, where an exam has K independent questions, and a student has a probability p of getting any answer correct. If another student independently has probability q of getting each question right, then the likelihood ratio condition reduces to[22]

$$\frac{\binom{K}{k} p^k (1-p)^{K-k}}{\binom{K}{k-1} p^{k-1} (1-p)^{K-k+1}} > \frac{\binom{K}{k} q^k (1-q)^{K-k}}{\binom{K}{k-1} q^{k-1} (1-q)^{K-k+1}}$$

or

$$\frac{p}{1-p} > \frac{q}{1-q}.$$

Thus if $p > q$, the first score uniformly dominates the second score over the range of all scores.

When f and g are differentiable, $f(t)/g(t)$ is increasing if and only if $f'(t)/f(t) \geq g'(t)/g(t)$. Let $N(\mu, \sigma)$ denote the normal distribution with mean μ and standard deviation σ. If $x \sim N(\mu, \sigma)$ with density $f(t)$ and $y \sim N(\tilde{\mu}, \tilde{\sigma})$ with density $g(t)$ then

$$\frac{f'(t)}{f(t)} = \frac{-(t-\mu)}{\sigma^2}; \frac{-(t-\tilde{\mu})}{\tilde{\sigma}^2} = \frac{g'(t)}{g(t)} \quad \forall t \in (-\infty, \infty).$$

If $\mu > \tilde{\mu}$ and $\sigma = \tilde{\sigma}$, then x uniformly dominates y on all of $(-\infty, \infty)$. More generally, x will uniformly dominate y on the interval including all t such that

$$\frac{t}{\sigma^2} - \frac{t}{\tilde{\sigma}^2} < \frac{\mu}{\sigma^2} - \frac{\tilde{\mu}}{\tilde{\sigma}^2}$$

[22] The notion of domination does not rely on independence. For example, suppose that with probability π the two students have chance $p_1 > q_1$ of getting each question, while with probability $1 - \pi$ they have chance $p_2 > q_2$ of getting each question; still the score of the first student would uniformly dominate that of the second. This suggests that much of our analysis can be extended to nonindependent scores, but we have not undertaken this extension here.

and y will uniformly dominate x on the complementary interval. Thus if $\sigma^2 < \tilde{\sigma}^2$, then x uniformly dominates y on the lower tail, and y uniformly dominates x on the upper tail. This is crucial in understanding the fourth example.[23]

Domination and uniform domination can be defined exactly the same way for any totally ordered set, such as a partition $\mathcal{P}$. The likelihood ratio criterion for uniform domination appearing in Lemma 3 also carries over to partitions. Given a density f and a partition $\mathcal{P}$, define the density

$$f_{\mathcal{P}}(x) = \begin{cases} f(x) & \text{if} \quad \{x\} \in \mathcal{P} \\ \frac{1}{b-a} \int_a^b f(t)dt & \text{if} \quad x \in [a,b) \in \mathcal{P} \end{cases}$$

where the integral is understood to be a sum in the discrete case. The analogue of Lemma 3 still holds: x uniformly dominates y on $[A, B)$ with respect to $\mathcal{P}$ if and only if $f_{\mathcal{P}}(t)/g_{\mathcal{P}}(t)$ is increasing on $[A, B)$.

4.2.2 Optimal Partitions

We are now ready to state some theorems about the optimal absolute grading partition when students have conditionally independent scores. It turns out that uniform stochastic dominance plays the central role in determining whether there should be masking (i.e., giving the same grade to different scores). If a worker's score uniformly stochastically dominates a shirker's, then the grading should be perfectly fine (so that doing better always means getting a strictly higher grade). Conversely, if on some subinterval of scores $[A, B)$ the shirker uniformly stochastically dominates the worker, then all scores in $[A, B)$ should be given the same grade. When the score densities are piecewise differentiable or discrete, $(-\infty, \infty)$ can be divided into intervals over which uniform stochastic dominance goes one way or the other, and so the optimal absolute partition can be determined. The non-regular case is more subtle, and we have not completely characterized the optimal grading. But we give necessary conditions for a partition to be optimal and slightly stronger sufficient conditions for it to be optimal, based on whether the shirker score stochastically dominates the worker score inside each partition cell, and whether the worker uniformly stochastically dominates the shirker score across (outside) partition cells.

The phrase "*N iid students who can work or shirk*" means that each student has two effort levels, and that assuming any one student shirks while the others work, he has score x_L with density g while each other student k has an independent score $x_k \sim x_H$, with density f. (Here $\sim$ denotes identical in distribution.) We begin with a simple theorem showing that when $x_H \gtrsim_U x_L$, perfectly fine grading is optimal, even in the wider class of all monotonic and anonymous grading schemes.

Theorem 3 (Fine Grading Can Be Optimal): *Let there be N iid students who can work or shirk. Suppose $x_H \gtrsim_U x_L$. Further, suppose x_H and x_L are either*

[23]The attentive reader might be puzzled, since the binomial exam scores "converge" to normal as $K \to \infty$, yet we never see the tail where x_q dominates x_p. That is because this tail is always beyond K. In fact it is not the exam scores, but normalized exam scores, which converge to normal, and the means of x_p and x_q are diverging at the rate K (not $\sqrt{K}$).

discrete or have piecewise continuous densities f and g with no atoms. Then fine grading is optimal in the class of all monotonic, anonymous grading schemes.

Proof: Consider the case of two students, a worker and a shirker. A monotonic grading scheme γ differs from fine precisely on the set

$$W = \{(x_L, x_H) \in \mathbb{R}^2 : x_L \neq x_H, \text{ yet } \gamma \text{ gives } x_L \text{ and } x_H \text{ the same grade}\}.$$

Suppose first that the densities f and g are discrete. By anonymity, $(\alpha, \beta) \in W$ if and only if $(\beta, \alpha) \in W$. But from the uniform domination $x_H \gtrsim_U x_L$, we know that if $\beta > \alpha$, then $f(\beta)/g(\beta) \geq f(\alpha)/g(\alpha)$. Hence replacing the masking grading γ on W with fine grading *lowers* the expected exam payoff to the shirker by

$$\sum_{\substack{(\alpha,\beta)\in W \\ \alpha<\beta}} [f(\beta)g(\alpha) - g(\beta)f(\alpha)] \geq 0.$$

Next suppose that f and g are piecewise continuous. Since W is measurable it can be approximated arbitrarily closely (in Lebesgue measure) by a union of small squares $Q = \{(x_L, x_H) : \alpha - \varepsilon \leq x_L < \alpha + \varepsilon \text{ and } \beta - \varepsilon \leq x_H < \beta + \varepsilon\}$ whose interiors do not contain any points of discontinuity of f and g. By anonymity, we may assume that the mirror square $Q^* = \{(x_L, x_H) : \beta - \varepsilon \leq x_L < \beta + \varepsilon \text{ and } \alpha - \varepsilon \leq x_H < \alpha + \varepsilon\}$ is also part of the approximation. These squares have area approximately equal to $\varepsilon^2 f(\alpha)g(\beta)$ or $\varepsilon^2 g(\alpha)f(\beta)$. If $\beta > \alpha$, then by uniform stochastic dominance, $f(\beta)g(\alpha) \geq g(\beta)f(\alpha)$. The masking on W thus hides the fact that x_H would have come ahead of x_L more often than behind x_L when both variables are in W. Hence masking W does not improve the incentive to work. A similar argument could be given with more than two students. ■

Now we focus on absolute grading, giving conditions under which a cell $[a, b)$ should *not* be cut by any optimal partition.

Theorem 4 (Coarseness in the Optimal Grading): *Let there be N iid students who can work or shirk. Suppose that on some interval $[a, b)$, x_L uniformly dominates x_H. Then for any partition $\mathcal{P}$ that cuts $[a, b)$, there is another partition $\mathcal{P}^*$ that gives at least as much incentive to work without cutting $[a, b)$. If x_L strictly uniformly dominates x_H on $[a, b)$, then every optimal grading partition is coarse on $[a, b)$.*

Proof: Consider the following picture:

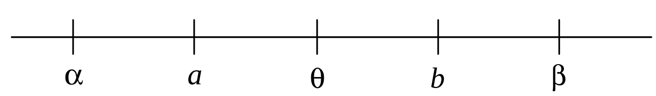

Figure 4: Cutting (a, b) at θ

Let $\mathcal{P}$ be a partition that cuts $[a, b)$ just once, that is, suppose that $\theta \in (a, b)$, and that $[\alpha, \theta) \in \mathcal{P}$ and $[\theta, \beta) \in \mathcal{P}$, where $\alpha \leq a < \theta < b \leq \beta$, where $-\alpha$ or β might be infinite. (The cases where the closed end comes on the right instead of the left are handled the same way.) For the rest of the proof all probabilities will be taken conditional on x_L and x_H being in $[\alpha, \beta]$. For ease of notation, we suppress this conditionality. Thus when we write $P(x_L \in [a, b))$, we really mean $P(x_L \in [a, b))/P(x_L \in [\alpha, \beta))$, etc.

¿From the Corollary to Lemma 1, we know that the expected payoff to L in $\mathcal{P}$ (conditional on both x_L and x_H in $[\alpha, \beta)$) is

$$P(x_L \in [\theta, \beta)) - P(x_H \in [\theta, \beta)).$$

Suppose first that $P(x_L \in [\theta, b)) \geq P(x_H \in [\theta, b))$. Then

$$\begin{aligned} & P(x_L \in [b, \beta)) - P(x_H \in [b, \beta)) \\ \leq \quad & P(x_L \in [\theta, \beta)) - P(x_H \in [\theta, \beta)). \end{aligned}$$

It follows from the Corollary to Lemma 1 that the partition obtained from $\mathcal{P}$ by moving the cut from θ to b (i.e., by replacing $[\alpha, \theta)$ and $[\theta, \beta)$ with $[\alpha, b)$ and $[b, \beta)$) weakly lowers the expected exam payoff to L, without cutting $[a, b)$, as was to be proved.

Suppose on the other hand that $P(x_L \in [\theta, b)) < P(x_H \in [\theta, b))$. From the dominance of x_L over x_H on $[a, b)$, we conclude that $P(x_L \in [a, \theta)) < P(x_H \in [a, \theta))$. It follows immediately that

$$\begin{aligned} & P(x_L \in [a, \beta)) - P(x_H \in [a, \beta)) \\ < \quad & P(x_L \in [\theta, \beta)) - P(x_H \in [\theta, \beta)) \end{aligned}$$

showing that the partition obtained from $\mathcal{P}$ by moving the cut from θ to a (i.e., by replacing $[\alpha, \theta)$ and $[\theta, \beta)$ with $[\alpha, a)$ and $[a, \beta)$) lowers the expected payoff to L, without cutting $[a, b]$, as was to be shown.

It only remains to consider the case where $\mathcal{P}$ cuts (a, b) multiple times. If there are a finite number of cuts, that case can be reduced to the case where (a, b) is cut once, by arbitrarily choosing any cut $\theta \in (a, b)$, and then choosing the highest cut $c < \theta$ and the lowest cut $d > \theta$, and replacing (a, b) with $(a, b) \cap (c, d)$. By the *uniform* dominance of x_L over x_H on $[a, b)$, x_L dominates x_H on $[a, b) \cap [c, d)$. The same proof can then be repeated. This reduces the number of cuts by 1. The reduction can then be iterated.

Finally, suppose there is a subinterval $[c, d) \subset [a, b)$ on which $\mathcal{P}$ is perfectly fine. Change $\mathcal{P}$ to $\mathcal{P}^*$ by completely masking $[c, d)$. Since x_L dominates x_H on $[c, d)$, this masking can only (weakly) lower the expected exam payoff to L. In this way we reduce the problem to finitely many cuts. This proves the first claim of the theorem. The second claim is proved the same way. ■

Theorem 4 shows that if work leads to a normal distribution $N(\mu, \sigma)$ of scores, and shirk leads to $N(\tilde{\mu}, \tilde{\sigma})$, where $\sigma \neq \tilde{\sigma}$, then one tail of scores will be completely masked in any optimal partition.

Theorem 4 also leads to a sufficient condition for the optimality of a partition $\mathcal{P}$. We say that x_L uniformly dominates x_H *inside* a partition $\mathcal{P}$ if $x_L \succsim_U x_H$ inside every cell $[a,b) \in \mathcal{P}$. We say that x_H uniformly dominates x_L outside a partition $\mathcal{P}$ if $x_H \succsim_U x_L$ on $\mathcal{P}$, i.e., across the cells of $\mathcal{P}$.

Theorem 5 (Uniform Inside and Outside Domination Implies Optimality): *Let there be N iid students who can work or shirk. Let $\mathcal{P}$ be a partition such that for every cell $[a,b) \in \mathcal{P}$, $x_L \succsim_U x_H$ on $[a,b)$, and such that $x_H \succsim_U x_L$ on the partition $\mathcal{P}$. Then $\mathcal{P}$ is optimal.*

Proof: Suppose $\mathcal{P}'$ does better than $\mathcal{P}$. From Theorem 4, we know that there is $\mathcal{P}''$ that does at least as well as $\mathcal{P}'$, and which does not cut any cell in $\mathcal{P}$. But every cell in $\mathcal{P}''$ is refined by cells in $\mathcal{P}$. Since $x_H \succsim_U x_L$ on $\mathcal{P}$, it follows that the payoff to x_L is weakly lower in $\mathcal{P}$ than in $\mathcal{P}''$, a contradiction. ■

Suppose that all students are homogeneous, with independent, and normally distributed exam scores. If work raises a student's expected exam score, without changing its variance, then Theorem 5 implies that an optimal grading scheme is to post the exact scores.

Similarly, if the K exam questions are identical, independent trials, and if hard work allows a student to raise his probability of getting each answer right, then again an optimal grading scheme is to reveal the exact scores.

But consider the two leading examples of this section. There we found that giving just two grades, A and B, improved incentives beyond what could be achieved by fully revealing the scores. Theorem 5 and Lemma 3 guarantee that these are indeed optimal partitions. In the first example, x_L uniformly dominates x_H on $\{0,1\}$, while x_H uniformly dominates x_L across the partition cells $\{0,1\}$, $\{2\}$, since $.36/.64 > 0/1$. In the second example, inside the cell $[0,50)$, x_H has probability zero, so x_L trivially uniformly dominates it. Inside the other cell $[50,100)$, $f(t)/g(t) = 2/2t = 1/t$ is strictly falling, so x_L uniformly dominates x_H. Across cells we can check that x_H uniformly dominates x_L. On $[0,50)$, we can define the effective density of a worker as $f_{\mathcal{P}}(t) = 0$, and that of a shirker as $g_{\mathcal{P}}(t) = .5$. On $[50,100)$ the effective densities become $f_{\mathcal{P}}(t) = 2$ and $g_{\mathcal{P}}(t) = 1.5$. Clearly $f_{\mathcal{P}}(t)/g_{\mathcal{P}}(t)$ is increasing.

In our next theorem we describe *necessary* conditions for a partition to be optimal, when agents are homogeneous. The "outside" condition, that $x_H \succsim_U x_L$ on $\mathcal{P}$, is the same as the sufficient "outside" condition appearing in Theorem 5. But the "inside" condition $x_L \succsim x_H$ on each cell $[a,b)$ in $\mathcal{P}$ is weaker than the sufficient "inside" condition $x_L \succsim_U x_H$ appearing in Theorem 5. For the theorem we need to impose slightly stronger conditions.

Definition: The exam performances x_L and x_H, with densities f and g, are called *generic* iff f and g are both discrete or both continuous, and there is a countable set $\{\cdots < a_i < a_{i+1} < \cdots\}$ such that $f(t)/g(t)$ is strictly increasing on $[a_i, a_{i+1})$ for all even i and strictly decreasing for all odd i.

Theorem 6 (Optimality Implies Domination): *Let there be N iid students who can work or shirk. Let $\mathcal{P}$ be an optimal absolute grading partition. Then for any cell $[a,b) \in \mathcal{P}$, $x_L \gtrsim x_H$ on $[a,b)$. Furthermore, if exam performances are all discrete,* or *all have piecewise differentiable densities,* or *are generic, then $x_H \gtrsim_U x_L$ on the totally ordered set $\mathcal{P}$.*

Proof: Consider any cell $[a,b)$ in $\mathcal{P}$ such that $P(x_L \in [a,b))P(x_H \in [a,b)) > 0$. Suppose there is some $\theta \in [a,b)$ with

$$\frac{P(x_L \in [\theta,b))}{P(x_L \in [a,b))} - \frac{P(x_H \in [\theta,b))}{P(x_H \in [a,b))} < 0.$$

Change $\mathcal{P}$ to $\mathcal{P}^*$ by replacing $[a,b)$ with $[a,\theta)$ and $[\theta,b)$. By the Corollary to Lemma 1, this must lower the expected exam payoff to the shirker against each worker. But this means that $\mathcal{P}^*$ is a better partition than $\mathcal{P}$, a contradiction proving that $x_L \gtrsim x_H$ on $[a,b)$.

Consider two consecutive cells $[a,b)$ and $[b,c)$ in $\mathcal{P}$ whose union $(a,b] \cup (b,c]$ has positive probability of being reached by both x_L and x_H. Then it is clear from the Corollary to Lemma 1 that

$$\frac{P(x_H \in [b,c))}{P(x_H \in [a,c))} \geq \frac{P(x_L \in [b,c))}{P(x_L \in [a,c))},$$

otherwise the partition $\mathcal{P}^*$ obtained from $\mathcal{P}$ by replacing the two cells $[a,b)$ and $[b,c)$ with the single cell $[a,c)$ would lower the expected exam score to L, contradicting the optimality of $\mathcal{P}$. Hence the likelihood ratio property holds for $f_{\mathcal{P}}$ and $g_{\mathcal{P}}$ across consecutive masked intervals. This same logic applies when exam scores are discrete.

The partition $\mathcal{P}$ must consist of intervals, each of which is fine or masked. If $f_{\mathcal{P}}(x)/g_{\mathcal{P}}(x)$ is increasing, then by Lemma 3 we have that $x \gtrsim_U y$ on $\mathcal{P}$. Suppose to the contrary that there is $\alpha < \beta$ with $f_{\mathcal{P}}(\alpha)/g_{\mathcal{P}}(\alpha) > f_{\mathcal{P}}(\beta)/g_{\mathcal{P}}(\beta)$. Then we can assume that either (1) α and β are in the same fine interval, or (2) α is in a fine interval and β is in the next (coarse) interval, or the reverse (3).

Suppose α and β are in the same fine interval. If the exam scores are generic, then on some open interval $(a,b) \subset (\alpha,\beta)$, $f(t)/g(t)$ is strictly decreasing, hence (by Theorem 4) (a,b) should be masked, contradicting the optimality of $\mathcal{P}$. If f and g are piecewise differentiable, and there is a point $\theta \in (\alpha,\beta)$ of differentiability with $(d/dt)(f(\theta)/g(\theta)) < 0$, then on some small interval (a,b), with $\theta \in (a,b) \subset (\alpha,\beta)$, $f(t)/g(t)$ is strictly decreasing, again contradicting Theorem 4. Otherwise there must be a jump down of f/g at a nondifferentiable point $\theta \in (\alpha,\beta)$.

Nearby, $f(t)/g(t) > M > m > f(s)/g(s)$ for all $a < t < \theta < s < b$. Modify $\mathcal{P}$ by masking the interval $[\theta-\varepsilon, \theta+\varepsilon) \subset [a,b)$. We claim that this will lower the payoff to x_L, contradicting the optimality of $\mathcal{P}$. Conditional on both x_H and x_L falling into $[\theta-\varepsilon,\theta+\varepsilon)$, $P\{x_L \in (\theta,\theta+\varepsilon)$ and $x_H \in [\theta-\varepsilon,\theta)\} - P\{x_H \in (\theta,\theta+\varepsilon)$ and $x_L \in [\theta-\varepsilon,\theta)\} > K > 0$, independent of ε. But conditional on both x_L and x_H lying in $(\theta,\theta+\varepsilon)$, or both lying in $[\theta-\varepsilon,\theta)$, the probability of x_L coming ahead minus the probability of x_H coming ahead converges to 0 as $\varepsilon \to 0$, because of the continuity of f and g on $[\theta-\varepsilon,\theta)$. Thus perfectly fine grading on $[\theta-\varepsilon,\theta+\varepsilon)$ gives L an edge that should have been eliminated by masking.

It only remains to consider the case where the drop in $f_{\mathcal{P}}(x)/g_{\mathcal{P}}(x)$ occurs at θ because θ is the cut between a perfectly fine cell $[c,\theta)$ of $\mathcal{P}$ and a masked cell $[\theta,d)$ of $\mathcal{P}$ (or vice versa). We argue that $\mathcal{P}$ could not be optimal, because moving the cut from θ to $\theta-\varepsilon$ would lower the payoff to x_L. We rely on the continuity of f to the left of θ, which holds for the generic case or the piecewise differentiable case.

Indeed, the change in expected payoff to L from moving the cut to $\theta-\varepsilon$ is

$$\begin{aligned}
&P(x_H\in[\theta,b))P(\theta-\varepsilon\le x_L<\theta)-P(x_L\in[\theta,b))P(\theta-\varepsilon\le x_H<\theta)\\
&\quad+\ P(\theta-\varepsilon\le x_H<\theta)P(\theta-\varepsilon\le x_L<\theta)[P(x_H>x_L|\theta-\varepsilon\le x_L,x_H<\theta)\\
&\quad-\ P(x_L>x_H|\theta-\varepsilon\le x_L,x_H<\theta)].
\end{aligned}$$

Observe that the third term goes to zero as ε^2 when $\varepsilon\to 0$, whereas the first two terms are of the order of ε. As $\varepsilon\to 0$, $P(\theta-\varepsilon\le x_H<\theta)$ converges to $\varepsilon f(\theta)$, and $P(\theta-\varepsilon\le x_L<\theta)$ converges to $\varepsilon g(\theta)$. Thus if $f(\theta_-)/g(\theta_-)>f_{\mathcal{P}}(\theta_+)/g_{\mathcal{P}}(\theta_+)$, then the first two terms add to less than zero. This shows that the extra masking obtained by lowering the cut θ to $\theta-\varepsilon$ reduces the expected exam payoff to L, contradicting the optimality of $\mathcal{P}$. ■

We can use Theorems 4 and 6 to completely characterize the optimal partitions for the normally distributed case, and more generally, for the generic case, even though Theorem 5 cannot be applied. Consider again the situation where $f\sim N(\mu,\sigma)$ and $g\sim N(\tilde{\mu},\tilde{\sigma})$ with $\mu>\tilde{\mu}$ and $\sigma<\tilde{\sigma}$. We have seen that the function $f(t)/g(t)$ is differentiable and single-peaked, strictly rising for $-\infty<t\le\bar{x}$ and strictly falling for $\bar{x}\le t<\infty$. See Figure 5. Thus the normal case is differentiable and generic. We know from Theorem 4 that any partition that cuts $(\bar{x},\infty)$ can be strictly improved by a partition that leaves $(\bar{x},\infty)$ uncut.

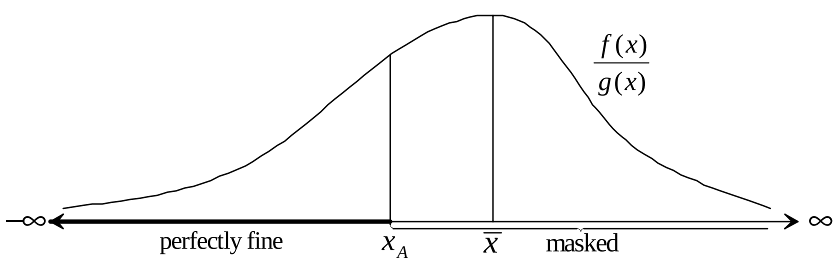

Figure 5: Normally Distributed: $\mu>\tilde{\mu}$, $\sigma>\tilde{\sigma}$

Moreover, since $f(t)/g(t)$ is strictly increasing on $(-\infty,\bar{x})$, we know from Theorem 6 that if there is any cut in $(-\infty,\bar{x})$, say at x_A, then $(-\infty,x_A)$ should be perfectly fine. It follows that the optimal partition must be of the form $\{(-\infty,x_A),[x_A,\infty)\}$ with $(-\infty,x_A)$ perfectly fine and $[x_A,\infty)$ completely masked, and $x_A<\bar{x}$. The point

x_A is uniquely defined by the greatest $x \leq \bar{x}$ such that

$$\frac{f(x_A)}{g(x_A)} = \frac{P(x_H \geq x_A)}{P(x_L \geq x_A)}.$$

At $x = \bar{x}$, $f(\bar{x})/g(\bar{x}) > P(x_H \geq \bar{x})/P(x_L \geq \bar{x})$. As x falls to the left of $\bar{x}$, $f(\bar{x})/g(\bar{x})$ also falls, but $P(x_H \geq x)/P(x_L \geq x)$ rises as long as $f(x)/g(x) > P(x_H \geq x)/P(x_L \geq x)$. Since $f(\bar{x})/g(\bar{x}) > 1$ and $\lim_{x\to-\infty} f(x)/g(x) = 0$, x_A exists.

If $x > x_A$, then the partition $\{(-\infty, x), [x, \infty)\}$ violates the outside condition of Theorem 6, while if $x < x_A$ it violates the inside condition on the cell $[x, \infty)$, as seen by cutting this cell at x_A.

The general picture is as follows

Theorem 7 (Optimal Partitions for Generic Densities): *Consider the case of generic densities f and g. In any optimal partition $\mathcal{P}$, all the cuts are in the rising segments (a_i, a_{i+1}) of f/g, where i is even. If there is a cut in (a_i, a_{i+1}), then the set of all cuts in (a_i, a_{i+1}) is a fine interval $[\alpha_i, \beta_i) \subset (a_i, a_{i+1})$ in P, or else a point $\alpha_i = \beta_i$. In either case,*

$$\frac{f(\beta_i)}{g(\beta_i)} = \frac{\int_{\beta_i}^{y} f(t)dt}{\int_{\beta_i}^{y} g(t)dt}$$

where y is the smallest cut to the right of β_i, and

$$\frac{f(\alpha_i)}{g(\alpha_i)} = \frac{\int_{x}^{\alpha_i} f(t)dt}{\int_{x}^{\alpha_i} g(t)dt}$$

where x is the biggest cut to the left of α_i. Thus the optimal partition has at most one more cell than the number of extremal points of f/g.

Proof: Immediate from Theorems 4 and 6, and Lemma 3, using the same logic as in Figure 5. ■

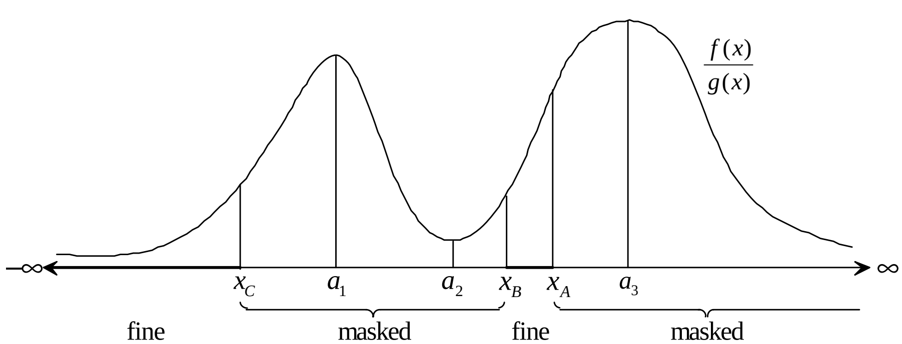

Figure 6: Generic Regular Densities

5 Grading on a Curve

We have assumed that students care only about their relative grade. It would seem therefore that relative grading, i.e., grading on a curve, would provide the best incentives. But in fact the contrary is true. We shall prove that when all the students are disparate or homogeneous, it is always better to grade according to an absolute scale, no matter how many students are in the class.

With a large number of students of each type, there is practically no difference between grading on a curve and absolute grading. Giving an A to top 10% of students can be replicated with very high probability by giving an A to all scores above x_A, for some appropriate threshold x_A.

5.1 Disparate Students

Let there be N_i students of type $i = 1, ..., \ell$, as in Section 3.3. Grading on a curve means specifying integers $K = (K_A, K_B, ..., K_Z)$ with $K_A + K_B + \cdots + K_Z = N = N_\ell + \cdots + N_1$, where the top K_A student exam scores get A, the next K_B get B and so on. (The probability of ties is zero.)

If there is only one disparate student of each type, then the student of type i will score below $\ell - i$ students and above $i-1$ students whether he works or shirks. With grading on a curve, his letter grade must therefore be independent of his effort, and so grading on a curve provides no work incentive whatsoever.

Consider a general population $N = (N_1, ..., N_\ell)$, and any grading on a curve $K = (K_A, K_B, ..., K_Z)$. We can find an absolute grading scheme that creates the same incentives to work for types $2, ..., \ell$, and at least as much for type 1.

Let μ_i measure the distribution of scores of a type i agent when he works (so $\mu_i(T) = \text{Prob}(x_i^H \in T)$ for every $T \subset \mathbb{R}$). Define $\mu \equiv \sum_{i=1}^{\ell} N_i \mu_i$. For any relative grade α, cut $\mathbb{R}$ at the *minimum* point x such that

$$K_A + \cdots + K_\alpha = \mu[x, \infty).$$

It is easy to check that the absolute partition defined by these cuts does the job.

5.2 Homogeneous Students

Theorem 8 (Absolute Grading Beats Grading on a Curve): *Let there be N iid students who can work or shirk. Let all their score densities be piecewise differentiable. Let $\mathcal{P}$ be an optimal absolute grading partition. Then $\mathcal{P}$ gives at least as much incentive for work as any grading on a curve.*

The proof relies on the inside and outside domination criteria for any optimal partition given in Theorem 6. Starting from an optimal partition, we prove the stronger result that conditional on the number of students who get each absolute grade, no grading on a curve will do better.

For the proof we first establish a simple lemma.

Lemma 4: *Denote scores in $[\theta, \infty)$ as A. Suppose $N-1$ students work hard, and each has probability p of getting an A, while one student shirks and has probability $q < p$ of getting an A. Suppose all scores are independent. If exactly K students wind up with A, the conditional probability that the shirker got A is less than K/N, while the probability any hard worker got A is more than K/N.*

Proof: The conditional probability the shirker got A is

$$\frac{q\begin{pmatrix} N-1 \\ K-1 \end{pmatrix} p^{K-1}(1-p)^{N-K}}{q\begin{pmatrix} N-1 \\ K-1 \end{pmatrix} p^{K-1}(1-p)^{N-K} + (1-q)\begin{pmatrix} N-1 \\ K \end{pmatrix} p^{K}(1-p)^{N-K-1}}$$

which is strictly monotonically increasing in q (as can easily be seen by dividing numerator and denominator by the numerator). But when $q = p$, symmetry implies that the expression must be exactly K/N. Hence the probability the shirker got A is less than K/N. Since exactly the proportion K/N students did get A, the probability of the good students getting A must then be more than K/N. ■

Proof of Theorem 8: Let $\mathcal{Q}$ be any partition of class rank $\{1, 2, ..., |N|\}$, representing an arbitrary grading on a curve.

For any possible exam scores $x = (x_n)_{n \in N}$, and any absolute interval G in $\mathcal{P}$ (whether coarse or fine), let $\mu_G(x)$ be the number of exam scores lying in G. For any curved grade γ, let $\mu_G^\gamma(x)$ be the number of scores in G that also get curved grade γ. Note that since there are no ties, $(\mu_G^\gamma)_{G \in \mathcal{P}}$ can be deduced from $\mu \equiv (\mu_G)_{G \in \mathcal{P}}$. A picture helps to clarify the situation.

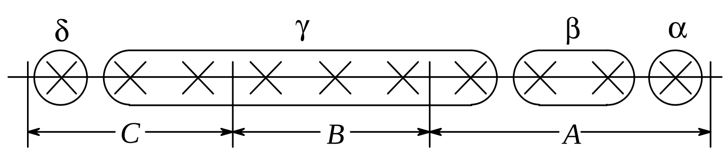

Figure 7. Absolute vs. Relative Grading

In the picture there are 4 A's, 3 B's, and 3 C's on the absolute scale. The relative scale gives grade α to the top score, β to scores 2 and 3, γ to $\{4, 5, 6, 7, 8, 9\}$, and δ to the 10th highest score. We can deduce that $\mu_A^\gamma = 1$, $\mu_B^\gamma = 3$, and $\mu_C^\gamma = 2$.

Define the join $\mathcal{P} \vee \mathcal{Q}$ of $\mathcal{P}$ and $\mathcal{Q}$ as follows: given scores $(x_n)_{n \in N}$, the exam grade for x_n is strictly higher according to $\mathcal{P} \vee \mathcal{Q}$ than the exam grade for x_m iff *either* the absolute letter grade for x_n is strictly higher than for x_m, *or* the relative grade for x_n is strictly higher than for x_m. Conditional on μ, that is achieved by cutting the curved grade cell γ into three curved grades γ_A, γ_B, and γ_C with cardinalities 1, 3, and 2, respectively.

We will now argue that if we grade according to the join $\mathcal{P} \vee \mathcal{Q}$ then the expected exam payoff of the shirker is no more than it was in $\mathcal{Q}$. If a relative grade, such as γ

in $\mathcal{Q}$, is refined in $\mathcal{P} \vee \mathcal{Q}$, we show that the expected score of the shirker, conditional on μ and on his being in γ to begin with, will not go up. Since splitting γ does not affect scores against students outside γ, it suffices to show that the expected exam score of the shirker against the other students in γ must be at most zero.

Let μ be an arbitrary interval distribution of scores such that $\mu_G \leq 1$ for every fine interval G in $\mathcal{P}$. By subdividing fine intervals into smaller and smaller fine intervals, the probability that two scores fall in a single fine interval goes to zero, so we can restrict attention to such μ.

Let γ be any curved grade with $P(x_L \in \gamma|\mu) > 0$. Let $\mathcal{P}^*$ be the collection of intervals G in $\mathcal{P}$ such that $\mu_G^\gamma \geq 1$. Clearly $\mathcal{P}^*$ has a finite number of elements.

Let $\tilde{q}$ be the probabilities of a shirker getting each absolute grade, conditional on the grade distribution μ and the shirker being in γ.

We shall show that if $C < B$ are intervals in $\mathcal{P}^*$ with $\tilde{q}_B > 0$, then $\tilde{q}_C > 0$ and

$$\frac{\tilde{q}_B}{\tilde{q}_C} \leq \frac{\mu_B^\gamma}{\mu_C^\gamma}.$$

Suppose γ includes the *top* $\mu_C^\gamma \leq \mu_C$ scores in C. If C is fine, then by hypothesis $\mu_C^\gamma = \mu_C = 1$, and $P(x_L \in C \cap \gamma|\mu \;\&\; x_L \in C) = 1$. If the interval C is masked, then by Theorem 6 the shirker dominates inside the cell, hence in either case

$$P(x_L \in C \cap \gamma|\mu \;\&\; x_L \in C) \geq \frac{\mu_C^\gamma}{\mu_C}.$$

By Bayes Law,

$$\begin{aligned} \tilde{q}_C &= P(x_L \in C|\mu \;\&\; x_L \in \gamma) \\ &= P(x_L \in C \cap \gamma|\mu \;\&\; x_L \in C)P(x_L \in C|\mu)/P(x_L \in \gamma|\mu). \end{aligned}$$

Thus

$$\tilde{q}_C \geq \frac{\mu_C^\gamma}{\mu_C} \frac{P(x_L \in C|\mu)}{P(x_L \in \gamma|\mu)}.$$

Similarly, if only the *bottom* part of scores in B are included in γ, then

$$P(x_L \in B \cap \gamma|\mu \;\&\; x_L \in B) \leq \frac{\mu_B^\gamma}{\mu_B}.$$

Thus arguing as with C,

$$\tilde{q}_B \leq \frac{\mu_B^\gamma}{\mu_B} \frac{P(x_L \in B|\mu)}{P(x_L \in \gamma|\mu)}.$$

If $\tilde{q}_B > 0$, then $P(x_L \in B|\mu) > 0$, and by Lemma 4,

$$\frac{P(x_L \in B|\mu)}{P(x_L \in C|\mu)} \leq \frac{\mu_B}{\mu_C}.$$

Hence $P(x_L \in C|\mu) > 0$. Therefore

$$\frac{\tilde{q}_B}{\tilde{q}_C} \leq \frac{\mu_B^\gamma P(x_L \in B|\mu)}{\mu_B P(x_L \in C|\mu)} \frac{\mu_C}{\mu_C^\gamma} \leq \frac{\mu_B^\gamma}{\mu_C^\gamma},$$

as claimed.

It follows that the shirker's expected exam payoff according to $\mathcal{P} \vee \mathcal{Q}$ against the other $\sum_{G \in \mathcal{P}^*} \mu_G^\gamma - 1$ scores in μ must be non-positive. Thus, conditional on μ alone, the expected exam payoff of a shirker is lower when grading by $\mathcal{P} \vee \mathcal{Q}$ than when grading by $\mathcal{Q}$.

Thus we have shown that the expected exam payoff to a shirker is lower under $\mathcal{P} \vee \mathcal{Q}$ than under $\mathcal{Q}$.

To conclude the proof, we need only show that the expected exam payoff of the shirker in $\mathcal{P}$ is even lower (weakly) than his expected exam payoff in $\mathcal{P} \vee \mathcal{Q}$. This follows at once from the fact (see Theorem 2) that conditional on being in a masked cell of $\mathcal{P}$, the score of the shirker dominates the score of a worker. But then, by Lemma 2, any monotonic grading scheme within cells of $\mathcal{P}$ (as is induced by $\mathcal{P} \vee \mathcal{Q}$) will weakly increase the payoff of the shirker. ■

6 Games of Money and Status: Setting Optimal Wage Schedules

Suppose now that the agents are workers in a firm who obtain direct utility from the purchasing power of wages and status utility from higher wages.

In keeping with our focus on ordinal status, we assume that if the N workers are paid wages $w = (w_1, ..., w_N)$, they obtain utility

$$u^n(w) = w_n + \#\{j : w_n > w_j\} - \#\{j : w_j > w_n\}.$$

The employer seeks to minimize his total wage bill, subject to providing incentives to work of at least d_n, for each worker n. The wage schedule he sets must be an anonymous and monotonic function of outputs.

As before, wages can be awarded on the basis of absolute performance or relative performance (analogous to grading on a curve). If the status payoff were eliminated, leaving behind only utility for money, then paying relative wages would resemble a tournament with multiple prizes: one can think of different players who are awarded the same prize as receiving the same grade.[24]

With status, we shall focus instead on wage schedules that are based on absolute performance, paying each worker according to his own output.[25] We conjecture that, as with games of pure status, absolute wages are better for the employer than relative wages anyway. (In the case of disparate workers, this is obvious from Section 5.1.)

Our model is thus the same as the classical wage problem, except that we have attached status to wages (and assumed participation constraints are not binding).

[24]In much of the tournament literature, the focus is on a single prize (see, e.g., Lazear–Rosen [1981] and Green–Stokey [1983]). Even when multiple prizes are allowed for, it is optimal to hand out only one (see Moldovanu–Sela [2001]). We reach a different conclusion because, in contrast to tournaments, the status of coming ahead is not merely instrumental in getting a bigger prize but is valued in and of itself.

[25]The focus of Moldovanu–Sela–Shi (2005) is on relative wages.

6.1 Disparate Workers

First consider the disparate case with $N_i \geq 1$ workers of type $i = 1, ..., \ell$, exactly as in Section 3, with exam scores reinterpreted as output produced. (Thus if a worker of type i shirks, he still produces more output than any worker of type $i - 1$.) We saw there that the principal could provide the greatest status incentives to work by clumping outputs into $\ell+1$ grades. Now we ask the question: given that the principal can use both status and wages as motivators, should he use them in equal proportion for all workers? Or should he, for example, reserve status mostly for higher worker types?

Since workers are disparate, strictly monotonic wages (such as piece-rate contracts) are suboptimal, because they squander the motivating power of status. Exactly as in Section 3, it will be optimal to set $\ell + 1$ different grade wages, with the upper (p_i)-fraction of i's outputs assigned wage w_i and the lower $(1 - p_i)$-fraction assigned wage w_{i-1} for $1 \leq i \leq \ell$, where $w_0 = 0$. (In order to signal that grade i confers higher status than grade $i-1$, the principal will need to assign $w_i > w_{i-1}$. To guarantee compactness, we allow him the freedom to set $w_i = w_{i-1}$ and still create a difference in status, e.g., via a title.[26] It will turn out that under the conditions of Theorem 9, he will never exercise this freedom.)

The employer's optimization problem is given below. One critical aspect of the problem is that we have capped the maximum wage at an arbitrary, but high, level M. The employer seeks to minimize his wage bill, subject to incentivizing everyone to work:

$$
\begin{aligned}
\min_{p,w} \quad & \sum_{i=1}^{\ell}[(1-p_i)w_{i-1} + p_i w_i]N_i = \min_{p,w}\left\{\sum_{i=1}^{\ell} w_{i-1}N_i + \sum_{i=1}^{\ell} p_i(w_i - w_{i-1})N_i\right\} \\
\text{s.t.} \quad & \tilde{I}_1 \equiv p_1[(N_1 - 1) + (1-p_2)N_2] + p_1(w_1 - w_0) \geq d_1 \\
& \quad \vdots \\
& \tilde{I}_i \equiv p_i[(N_i - 1) + (1-p_{i+1})N_{i+1} + p_{i-1}N_{i-1}] + p_i(w_i - w_{i-1}) \geq d_i, \text{ for } 1 < i < \ell \\
& \quad \vdots \\
& \tilde{I}_\ell \equiv p_\ell[(N_\ell - 1) + p_{\ell-1}N_{\ell-1}] + p_\ell(w_\ell - w_{\ell-1}) \geq d_\ell \\
& \qquad 0 \leq p_i \leq 1 \\
& \qquad 0 = w_0 \leq w_1 \leq \cdots \leq w_\ell \leq M
\end{aligned}
$$

Note that each agent i's incentive $\tilde{I}_i$ consists of the old (from Section 3.3) status incentive I_i plus a wage incentive $p_i(w_i - w_{i-1})$. We assume that the cap $M > \sum_{i=1}^{\ell} d_i$.

If there were no status attached to higher wages, the employer would simply pay each worker type its disutility of effort, setting $p_i = 1$ and $w_i - w_{i-1} = d_i$ for all

[26] We could have postulated instead that w_i confers higher status than w_j if, and only if, $w_i \geq w_j + \delta$ for some threshold $\delta > 0$. Then our last constraint in the principal's optimization problem would read: $0 = w_0$, $w_i + \delta \leq w_{i+1}$ for $1 \leq i \leq \ell - 1$, $w_\ell \leq M$. It is worth noting that as $d \to \infty$, $p_i(w_i - w_{i-1}) \to \infty$ since the status incentive terms $I_\ell^i(p)$ are bounded by $N_1 + \cdots + N_\ell$. Thus the constraints $w_i + \delta \leq w_{i+1}$ are automatically satisfied for large enough d (given any δ).

$i = 1, ..., \ell$. Wages would never rise faster than disutilities as the ability i increased. But Theorem 9 shows that when wages confer status, it is uniquely optimal (no matter what the population distribution $N_1, ..., N_\ell$) to pay an astronomical wage to a tiny number of top performers. The shape of the wage schedule for the other workers depends on $N_1, ..., N_\ell$, as well as $d_1, ..., d_\ell$, and will be described in Section 6.1.1.

Theorem 9 (Elite Performers Get Exorbitant Wages): *Let there be ℓ disparate types of workers, with $N_i \geq 1$ of each type $i = 1, ..., \ell$. Suppose their disutilities $(d_1, ..., d_\ell)$ of work are such that at any feasible wage schedule (p, w) they each get a wage incentive (in addition to their status incentive),*[27] *i.e., $p_i(w_i - w_{i-1}) > 0$ for each $i = 1, ..., \ell$.*

Then there is a unique optimal status partition and wage schedule (p, w). In fact $p_i = 1$ for all $i = 1, ..., \ell - 1$, so that any two workers of the same type below ℓ get the same status and wage. But $p_\ell \leq \sum_{i=1}^{\ell} d_i/M$, and $w_\ell = M$. In other words, for large M, a tiny elite $p_\ell N_\ell$ out of the highest type ℓ is paid the exorbitant salary M, while the rest of their type obtain the same status and pay as type $\ell - 1$.

Proof: It will be useful to keep in mind throughout that raising p_i has the effect of raising $\tilde{I}_i$ and $\tilde{I}_{i+1}$ and lowering $\tilde{I}_{i-1}$ without disturbing other incentives. Similarly, raising w_i raises $\tilde{I}_i$ and lowers $\tilde{I}_{i+1}$, with no other effect.

Raising p_1 improves the status incentive for agents of type 1 and 2, without disturbing the incentives of the others, allowing the employer to reduce w_1. Thus $p_1 = 1$.

Inductively assume that $p_1 = \cdots = p_{i-1} = 1$ for $i < \ell$. Suppose $p_i < 1$.

Set $\tilde{w}_{i-1} = w_{i-1} + \varepsilon N_i$ and $\tilde{p}_i = p_i + \varepsilon$. Then the status incentive of $i - 1$ goes down by $\varepsilon N_i p_{i-1} = \varepsilon N_i$, but his wage incentive goes up by the same amount: since $p_{i-1} = 1$, $p_{i-1}(\tilde{w}_{i-1} - w_{i-2}) = p_{i-1}(w_{i-1} - w_{i-2}) + \varepsilon N_i$.

Also, the status incentive of i goes up by

$$\Delta I_i(\varepsilon) \equiv \varepsilon[(N_i - 1) + (1 - p_{i+1})N_{i+1} + N_{i-1}].$$

This allows us to reduce his wage incentive by the same amount. So, set $\tilde{w}_i$ to satisfy

$$\tilde{p}_i(\tilde{w}_i - \tilde{w}_{i-1}) \equiv p_i(w_i - w_{i-1}) - \Delta I_i(\varepsilon).$$

For small ε, $\Delta I_i(\varepsilon)$ is small, so $p_i(w_i - w_{i-1}) > 0$ implies that $\tilde{p}_i(\tilde{w}_i - \tilde{w}_{i-1}) > 0$, which in turn implies $\tilde{w}_i > \tilde{w}_{i-1}$, retaining the monotonicity of the revised wages. We shall be assuming ε small enough to guarantee monotonicity in all future wage revisions, without explicitly saying so.

[27] This is guaranteed if, for example, $N_{i-1} + N_i + N_{i+1} < d_i$ for all i (with $N_0 = N_{\ell+1} = 0$). Would anybody work for free, just for the status of coming ahead of all his peers?

Note that

$$
\begin{aligned}
\tilde{w}_i &= \frac{p_i}{\tilde{p}_i} w_i + \tilde{w}_{i-1} - \frac{p_i}{\tilde{p}_i} w_{i-1} - \frac{1}{\tilde{p}_i} \Delta I_i(\varepsilon) \\
&= \frac{p_i}{p_i + \varepsilon} w_i + \left(1 - \frac{p_i}{p_i + \varepsilon}\right) w_{i-1} + \varepsilon N_i - \frac{1}{p_i + \varepsilon} \Delta I_i(\varepsilon) \\
&< w_i + \varepsilon N_i - \varepsilon[(N_i - 1) + (1 - p_{i+1}) N_{i+1} + N_{i-1}] \\
&\leq w_i - \varepsilon(N_{i-1} - 1) \leq w_i
\end{aligned}
$$

since $w_i > w_{i-1}$ and $p_i + \varepsilon < 1$ (if $p_i < 1$ and ε is small) and $N_{i-1} \geq 1$.

Finally, the status incentive of $i+1$ goes up by

$$\Delta I_{i+1}(\varepsilon) \equiv \varepsilon p_{i+1} N_i.$$

Therefore the wage incentive of $i+1$ can be reduced by the same amount. So set $\tilde{w}_{i+1}$ to satisfy

$$p_{i+1}(\tilde{w}_{i+1} - \tilde{w}_i) = p_{i+1}(w_{i+1} - w_i) - \Delta I_{i+1}(\varepsilon).$$

Since $\tilde{w}_i < w_i$, clearly $\tilde{w}_{i+1} < w_{i+1}$. Hence recursively setting

$$\tilde{w}_j - \tilde{w}_{j-1} = w_j - w_{j-1} \text{ for } j > i+1$$

further lowers wages without changing incentives.

It only remains to show that the wage bill defined in the employer minimization problem has gone down. The only terms that increase are

$$w_{i-1} N_i \text{ and } p_{i-1}(w_{i-1} - w_{i-2}) N_{i-1}$$

while many terms are reduced, including

$$p_i(w_i - w_{i-1}) N_i \text{ and } p_{i+1}(w_{i+1} - w_i) N_{i+1}.$$

The increases add up to

$$\varepsilon N_i N_i + \varepsilon N_i N_{i-1},$$

while just these two reductions add to

$$
\begin{aligned}
& \Delta I_i(\varepsilon) N_i + \Delta I_{i+1}(\varepsilon) N_{i+1} \\
=\ & \varepsilon N_i[(N_i - 1) + (1 - p_{i+1}) N_{i+1} + N_{i-1}] + \varepsilon p_{i+1} N_i N_{i+1} \\
=\ & \varepsilon N_i^2 + \varepsilon N_i N_{i-1} + \varepsilon N_i (N_{i+1} - 1).
\end{aligned}
$$

Since $N_{i+1} \geq 1$, the reduction is at least as big as the increase. But we have ignored many other strictly positive reductions (for example in $w_{i+1} N_{i+1}$). This contradiction proves that $p_i = 1$, for $i = 2, ..., \ell - 1$.

Now suppose $w_\ell < M$. Since $p_\ell(w_\ell - w_{\ell-1}) > 0$, clearly $p_\ell > 0$. Lower p_ℓ by ε. This raises the status incentive of type $\ell - 1$ workers by εN_ℓ, enabling us to lower the wage incentive for type $\ell - 1$ by the same amount.

Recalling that $p_{\ell-1}=1$, set $\tilde{w}_{\ell-1}$ to satisfy

$$(\tilde{w}_{\ell-1}-w_{\ell-2})=(w_{\ell-1}-w_{\ell-2})-\varepsilon N_\ell.$$

This drop in p_ℓ unfortunately lowers the status incentive of type ℓ by $\varepsilon(N_\ell-1+N_{\ell-1})$. Therefore we must raise the wage incentive of ℓ, choosing $\tilde{w}_\ell$ to solve

$$(p_\ell-\varepsilon)(\tilde{w}_\ell-\tilde{w}_{\ell-1})=p_\ell(w_\ell-w_{\ell-1})+\varepsilon(N_{\ell-1}+N_\ell-1).$$

Fortunately, there is no group $\ell+1$ to be affected by the change in p_ℓ, which is why it will turn out to be optimal to *lower* p_ℓ as long as $w_\ell<M$, whereas it was shown to be optimal to *raise* p_i all the way to 1 for any $i<\ell$.

Indeed the terms in the wage bill that change are

$$w_{\ell-1}N_\ell+p_{\ell-1}(w_{\ell-1}-w_{\ell-2})N_{\ell-1}+p_\ell(w_\ell-w_{\ell-1})N_\ell.$$

The net change in those terms is

$$\begin{aligned} & -\varepsilon N_\ell^2-\varepsilon N_\ell N_{\ell-1}+\varepsilon(N_{\ell-1}+N_\ell-1)N_\ell \\ = & -\varepsilon N_\ell<0 \end{aligned}$$

showing that the wage bill can be reduced, a contradiction. This proves that $w_\ell=M$.

Clearly, for any $1\le i\le \ell-2$, the wage incentive

$$w_i-w_{i-1}=p_i(w_i-w_{i-1})=d_i-I_i=d_i-(N_i+N_{i-1}-1).$$

This recursively proves that w_i is uniquely determined (starting from $w_0=0$), for $i=1,...,\ell-2$. Also

$$w_{\ell-1}-w_{\ell-2}=d_{\ell-1}-[N_{\ell-2}+N_{\ell-1}+(1-p_\ell)N_\ell-1]$$

and

$$M-w_{\ell-1}=\frac{d_\ell}{p_\ell}-[N_{\ell-1}+N_\ell-1]$$

(recalling that $w_\ell\equiv M$ in the last equation). It can be checked[28] that there is a unique solution $w_{\ell-1}$, p_ℓ of these two simultaneous equations, so that the optimal wage schedule is determined uniquely.

Summing the inequalities $w_i-w_{i-1}\le d_i$ over $i=1,...,\ell-1$ gives

$$w_{\ell-1}\le d_1+\cdots+d_{\ell-1}.$$

Since the wage incentive of ℓ is at most d_ℓ, we have

$$p_\ell(M-w_{\ell-1})\le d_\ell,$$

hence

$$p_\ell\le\frac{d_\ell+p_\ell w_{\ell-1}}{M}\le\frac{d_\ell+w_{\ell-1}}{M}\le\frac{\sum_{i=1}^{\ell}d_i}{M}. \qquad \blacksquare$$

[28] Adding the two equations yields a quadratic in the single unknown p_ℓ. This quadratic is convex and has a negative value at $p_\ell=0$ and therefore a unique positive solution.

Theorem 9 shows that when workers are disparate, the interplay of status incentives and wage incentives produces an optimal hierarchy that is substantially different from what was generated in the game of pure status discussed in Section 3. In the pure status case, it was essential that workers of the same type could end up with different status, so we found $p_i < 1$ for all $2 \leq i \leq \ell$. The opportunity to motivate workers via money ensures that only the top group ℓ should be split into different status groups, so now $p_i = 1$ for $1 \leq i \leq \ell - 1$.

Theorem 9 also gives an explanation via status for the exorbitant pay often seen at the very top of some real world hierarchies. It is cheaper to incentivize the managing directors of type $\ell - 1$ as much as possible via status rather than wages. To achieve this they must be able to get the same status as most of the senior managing directors of type ℓ, if they work hard. This fixes the wage of the latter group at the managing director's level. In order to incentivize the senior managing directors, they are given to understand that the CEO will be chosen from among their rank, and even though the chance of getting selected is small, the salary is huge.

This stratagem of paying a huge salary to the tiny fraction of top performers in a group is counterproductive at any level below ℓ, because monotonicity would force the employer to pay *all* workers of higher type at least as much.

6.1.1 Wage Differentials for Disparate Workers

The conclusions about exorbitant pay for the CEO and $p_i = 1$ for all $i = 1, ..., \ell - 1$ are quite robust; they hold regardless of the distribution of abilities $N_1, ..., N_\ell$, or the disutilities of work $d_1, ..., d_\ell$.

But the wage differentials $w_i - w_{i-1}$ for $i < \ell$ do depend on the N_j's and d_j's. Our analysis is based on the following corollary:

Corollary to Theorem 9: *Under the conditions of Theorem 9,* $I_1 = N_1 - 1 < I_2 = N_1 + N_2 - 1$*;* $I_i = N_{i-1} + N_i - 1$ *for* $i = 3, ..., \ell - 2$. *Also,* $I_{\ell-1} = N_{\ell-2} + N_{\ell-1} + (1 - p_\ell)N_\ell - 1 \approx N_{\ell-2} + N_{\ell-1} + N_\ell - 1$. *Finally,* $I_\ell = p_\ell(N_{\ell-1} + N_\ell - 1) \approx 0$.

Thus for $2 \leq i \leq \ell - 2$,

$$(w_i - w_{i-1}) - (w_{i-1} - w_{i-2}) = (d_i - d_{i-1}) + N_{i-2} - N_i.$$

Proof: The incentive formulae are trivially generated by plugging $p_i = 1$ for $1 \leq i \leq \ell - 1$ into the status incentives for each agent, and by observing that $p_\ell \leq \sum_{i=1}^{\ell} d_i/M \approx 0$ if M is large.

The wage differentials were explicitly computed in the proof of Theorem 9. ■

A natural case to consider is the one where the population N_i declines in size as the ability type increases. If disutilities do not fall as fast (i.e., if $N_{i-2} - N_i > d_{i-1} - d_i$, which occurs for example, if disutilities are constant), then we conclude from the corollary that wage differentials escalate as we go up the ability ladder from $i = 2$ to $i = \ell - 2$. See Figure 8.

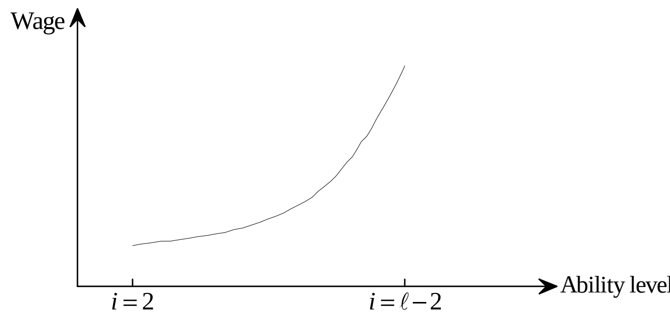

Figure 8: Pyramidical Ability Distribution

Another natural case arises in a population that is bell-shaped around the mean ability. When $N_i - N_{i-2} > d_i - d_{i-1}$ for small i and $N_{i-2} - N_i > d_{i-1} - d_i$, for large i, we get a wage schedule which is first concave and then convex. See Figure 9.

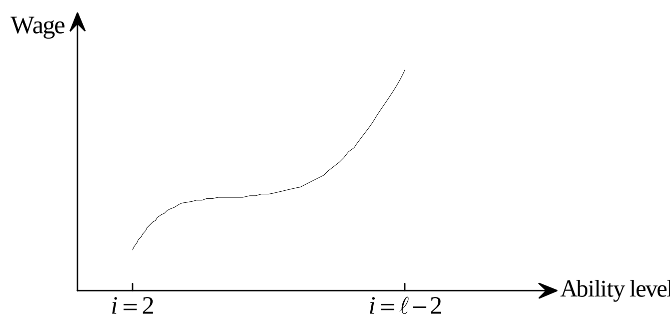

Figure 9: Bell-Shaped Ability Distribution

The simplest case is when $N_i = N \; \forall i$ and $d_i = d \; \forall i$. Below we graph the situation exactly, for all ℓ, when $\ell = 6$, $N_i = 2$, and $d_i = 3$.

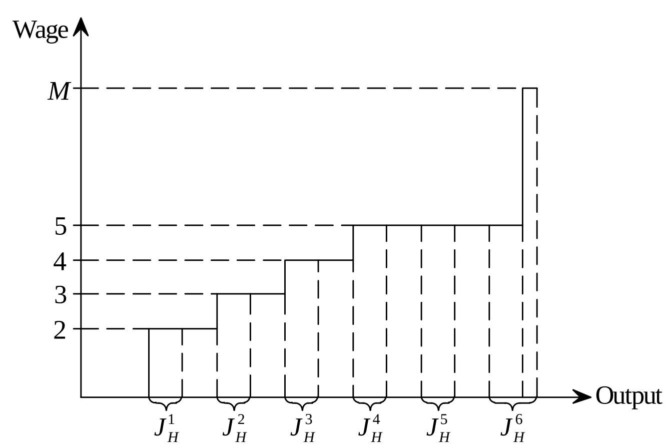

Figure 10

6.2 Homogeneous Workers

Consider again the model of Section 4, but with scores interpreted as output produced by homogeneous workers. We begin by showing that if no status attached to higher wages, then the employer should use trigger wages, paying a constant wage to any output above a threshold, and nothing for output below it. When status enters the picture, the optimal wage schedule is exactly the same, except that the constant wage is reduced by the status incentive. The reason is that the top cell in the optimal pure status partition (derived in Section 4) begins precisely at the threshold wage of the pure wage game. Thus for the homogeneous case, wage incentives and status incentives run in parallel, and reinforce each other.[29]

Recall that f and g are the output densities of the worker and shirker. For simplicity we shall suppose[30] that the possible outputs lie in a finite set $Q \subset \mathbb{R}$, and that for all $x \in Q$, either $f(x) > 0$ or $g(x) > 0$. A (monotonic, anonymous, absolute) *wage schedule* is simply a nondecreasing function w mapping outputs to wages, $w : Q \to \mathbb{R}_+$. The set of all such w is denoted $\mathcal{W}$.

The following lemma shows that if an employer does *not* have any recourse to status incentives, but must motivate his employees exclusively via wage incentives, then he should pick a *trigger wage*, i.e., a wage schedule $w \in \mathcal{W}$ which is zero below a trigger output θ and a positive constant M for all outputs at least θ.

Lemma 6: *If g does* not *stochastically dominate f, the problem*

$$\begin{aligned} \min_{w \in \mathcal{W}} \quad & \sum_{x \in Q} f(x) w(x) \\ \text{s.t.} \quad & \sum_{x \in Q} [f(x) - g(x)] w(x) \geq d \end{aligned}$$

is solved by every trigger wage of the form

$$w^*(x) = \begin{cases} 0 & \text{if } x < \theta \\ M & \text{if } x \geq \theta \end{cases}$$

where

$$\theta \in Q^* \equiv \arg\max_{\theta \in Q} \frac{\sum_{x \geq \theta} f(x)}{\sum_{x \geq \theta} g(x)}$$

and

$$M = \frac{d}{\sum_{x \geq \theta} [f(x) - g(x)]}.$$

[29] For cells below the top, the employer will pay the same zero wage, but allocate titles to the different cells to generate the same status as in the pure status game described in Section 4.

[30] If Q is a compact interval, we can approximate it by a fine finite grid and then use a limiting argument to derive the analogous result for a continuum of outputs.

Proof: Consider first the class $\mathcal{W}_T \subset \mathcal{W}$ of all trigger wages. Since g does not dominate f, there is a θ with $\sum_{x\geq\theta}[f(x) - g(x)] > 0$. Hence there are feasible $w \in \mathcal{W}_T$ satisfying the constraint. In fact there are only a finite number of such θ, and once θ is specified, $M(\theta)$ is determined by the constraint. The optimal trigger wage is then given by the pair $(\theta, M(\theta))$ that minimizes the expected wage. But since

$$\arg\min_{\theta\in Q} \frac{\sum_{x\geq\theta} f(x)M}{\sum_{x\geq\theta}[f(x) - g(x)]M} = \arg\max_{\theta\in Q} \frac{\sum_{x\geq\theta} f(x)}{\sum_{x\geq\theta} g(x)},$$

it is obvious that w^* is a best wage schedule in $\mathcal{W}_T$.

But any $w \in \mathcal{W}$ is simply a convex combination of $\tilde{w} \in \mathcal{W}_T$. Since the minimand and the constraint are linear in the vector $(w(x))_{x\in Q}$, the convex combination can never be better than the best summand. Hence w^* is also optimal in $\mathcal{W}$. ■

Lemma 7: *Let $\theta \in Q^*$, and let $\mathcal{P}$ be an optimal, absolute status incentive partition (as defined in Section 4). Then the cell $[a, b)$ in $\mathcal{P}$ containing θ has $a = \theta$ or else $a \in Q^*$.*

Proof: Take any $\theta \in Q^*$. Suppose there is a cell $[a, b)$ in $\mathcal{P}$ with $a < \theta < b$. Since $\theta \in Q^*$,

$$\frac{\sum_{x\geq\theta} f(x)}{\sum_{x\geq\theta} g(x)} \geq \frac{\sum_{x\geq a} f(x)}{\sum_{x\geq a} g(x)} \tag{1}$$

and

$$\frac{\sum_{x\geq\theta} f(x)}{\sum_{x\geq\theta} g(x)} \geq \frac{\sum_{x\geq b} f(x)}{\sum_{x\geq b} g(x)} \text{ if } b \leq \max\{q \in Q\}. \tag{2}$$

Hence,

$$\frac{\sum_{a\leq x<\theta} f(x)}{\sum_{a\leq x<\theta} g(x)} \leq \frac{\sum_{x\geq\theta} f(x)}{\sum_{x\geq\theta} g(x)} \tag{3}$$

and

$$\frac{\sum_{\theta\leq x<b} f(x)}{\sum_{\theta\leq x<b} g(x)} \geq \frac{\sum_{x\geq\theta} f(x)}{\sum_{x\geq\theta} g(x)}. \tag{4}$$

Putting these last two inequalities together,

$$\frac{\sum_{a\leq x<\theta} f(x)}{\sum_{a\leq x<\theta} g(x)} \leq \frac{\sum_{\theta\leq x<b} f(x)}{\sum_{\theta\leq x<b} g(x)}. \tag{5}$$

If $a \notin Q^*$, then inequality (1) is strict, and hence inequalities (3) and (5) are strict. But strict (5) contradicts Theorem 6, according to which g stochastically dominates f inside $[a, b)$. Thus $a \in Q^*$. ■

Lemma 8: *Let $\mathcal{P}$ be an optimal, absolute status incentive partition. Suppose $[a, b)$ is a cell in $\mathcal{P}$ with $a \in Q^*$. Then for every cell $[c, d)$ in $\mathcal{P}$ with $c \geq b$, $c \in Q^*$. Thus if $[e, \infty)$ is the top cell in $\mathcal{P}$, then $e \in Q^*$.*

Proof: ¿From the outside condition of Theorem 6, f uniformly dominates g across cells of $\mathcal{P}$. Hence

$$\frac{\sum_{x \geq c} f(x)}{\sum_{x \geq c} g(x)} \geq \frac{\sum_{x \geq a} f(x)}{\sum_{x \geq a} g(x)},$$

hence $c \in Q^*$.

¿From Lemma 7 we know that at least one cell $[a, b)$ in $\mathcal{P}$ has $a \in Q^*$. ■

We now consider the general problem of selecting the optimal wage schedule, taking into account its status incentives and wage incentives together.

Let $\mathcal{P}$ be a partition of Q into consecutive cells. A wage schedule w is *consistent* with $\mathcal{P}$, if w is constant on each cell of $\mathcal{P}$. The set of all wage schedules in $\mathcal{W}$ that are consistent with $\mathcal{P}$ is denoted $\mathcal{W}(\mathcal{P})$.

Let Π be the (finite) set of partitions of Q into consecutive cells.

Consider the problem of finding an optimal wage schedule w:

$$\min_{w, \mathcal{P}} \quad \sum_{x \in Q} f(x) w(x)$$

$$\text{s.t.} \quad \begin{cases} I(\mathcal{P}) + \sum_{x \in Q} [f(x) - g(x)] w(x) \geq d \\ \mathcal{P} \in \Pi \\ w \in \mathcal{W}(\mathcal{P}) \end{cases}$$

where $I(\mathcal{P})$ denotes the status incentive generated by $\mathcal{P}$, as defined in Section 4.

We are now ready to state the main result of our section.

Theorem 10: *Let $\mathcal{P}$ be any optimal, absolute status partition, and let $[a, \infty)$ denote its top cell. Then there is an optimal wage schedule $(\mathcal{P}, w)$ which has $\mathcal{P}$ as its partition, and a trigger wage at $\theta = a$.*

Proof: By Lemma 8, $a \in Q^*$. By Lemma 6, there is a trigger wage, with trigger at a, that minimizes the wage bill given the constraint that the wage incentive exceed $d - I(\mathcal{P})$. Clearly this trigger wage is consistent with $\mathcal{P}$.

No other consistent pair $(\mathcal{P}', w')$ can do better. $\mathcal{P}'$ provides an incentive $I(\mathcal{P}') \leq I(\mathcal{P})$, since $\mathcal{P}$ is status optimal, and therefore the wage w' must generate a bigger wage incentive $d - I(\mathcal{P}')$. It must therefore generate at least as high a wage bill as w. ■

As in the disparate case, we find that the top echelon of performers are rewarded well beyond all the others. The difference is that in the disparate case it was only a tiny elite, while here it is every score in the top cell, which might be quite big.

7 References

Cole, Harold L., George J. Mailath, and Andrew Postlewaite (1992). "Social Norms, Savings Behavior, and Growth," *Journal of Political Economy*, 100(6): 1092–1125.

——— (1995). "Incorporating Concern for Relative Wealth into Economic Models," *Federal Reserve Bank of Minneapolis Quarterly Review*, 19(3): 12–21.

——— (1998). "Class Systems and the Enforcement of Social Norms," *Journal of Public Economics*, 70: 5–35.

Corneo, Giacomo and Oliver Jeanne (1997). "On Relative Wealth Effects and the Optimality of Growth," *Economics Letters*, 54: 87–92.

Direr, Alexis (2001). "Interdependent Preferences and Aggregate Saving," *Annales-d'Economie-et-de-Statistique*, July–December, 63–64: 297–308.

Dubey, P. and Chien-wei Wu (2001). "When Less Scrutiny Induces More Effort," *Journal of Mathematical Economics*, 36(4): 311–336.

Dubey, P. and J. Geanakoplos (2004). "Grading Exams: 100, 99, ..., 1 or A, B, C? Incentives in Games of Status," Cowles Foundation Discussion Paper No. 1467, Yale University.

Dubey, P. and O. Haimanko (2003). "Optimal Scrutiny in Multiperiod Promotion Tournaments," *Games and Economic Behavior*, 42(1): 1–24.

Duesenberry, James S. (1949). *Income, Saving and the Theory of Consumer Behavior*. Cambridge: Harvard University Press.

Easterlin, Richard (1974). "Does Economic Growth Improve the Human Lot?" in Paul A. David and Melvin W. Reder, eds., *Nations and Households in Economic growth: Essays in Honor of Moses Abramowitz*. New York: Academic Press, pp. 87–125.

Fehr, E. and K.M. Schmidt (1999). "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114: 817–868.

Frank, Robert H. (1985). *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press.

Green, J.R. and N.L. Stokey (1983). "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91: 349–364.

Hopkins, Ed and Tatiana Kornienko (2003). "Ratio Orderings and Comparative Statics," Working Paper, University of Edinburgh.

Lazear, E. and S. Rosen (1981). "Rank-Order Tournaments as Optimal Labor Contracts," *Journal of Political Economy*, 89: 841–864.

Moldovanu, B. and A. Sela (2001). "The Optimal Allocation of Prizes in Contests," *American Economic Review*, 91(3): 542–558.

Moldovanu, B., A. Sela and X. Shi (2005). "Status Contests," Working Paper, Department of Economics, Bonn University.

Pollak, Robert (1976). "Interdependent Preferences," *American Economic Review*, 66(3): 309–320.

Robson, Arthur (1992). "Status, the Distribution of Wealth, Private and Social Attitudes to Risk," *Econometrica*, 60(4): 837–857.

Shaked, M. and J.G. Shanthikumar (1994). *Stochastic Orders and their Applications*. San Diego: Academic Press.

Spence, M. (1974). *Market Signaling: Informational Transfer in Hiring and Related Processes*. Cambridge: Harvard University Press.

Veblen, Thorstein (1899). *The Theory of the Leisure Class*. New York: Macmillan.