

The concept of capacity

Class handout

[As class starts, Radu passes out a handout with the following case and asks the students to read it out loud as a group, each taking turns to read. Radu gives students some time to ponder and discuss among themselves in smaller groups.]

Case: Behind the Counter at Beanstalk Coffee

Beanstalk Coffee is a single-location specialty coffee shop located near a large urban university campus. The shop is known for its carefully sourced beans, its precise preparation methods, and its regulars, many of whom have been coming in at the same time every weekday for years.

The general manager, Theo, has been with Beanstalk for four years. He is proud of the shop's reputation and genuinely fond of his staff, most of whom have been with him for over a year—a rarity in the industry. Recently, however, he has grown concerned. The morning rush, which runs from roughly 7:30 to 9:30 a.m., has become noticeably harder to get through. A few regulars have mentioned, politely, that their orders are taking longer than they used to. One longstanding customer switched to the café down the street.

Theo pulled together what data he could. On a typical weekday during the morning rush, the shop serves approximately 150 customers in the two-hour window from 7:30 am to 9:30 am. Theo was also careful to benchmark his staff's performance, which was 90 seconds per drink, based on industry standards for a skilled barista at a specialty coffee shop. Theo also knew that standard time calculations for this type of work, which account for a normal pace and a personal fatigue and delay (PFD) allowance of roughly 20%, suggest a barista should be able to complete a drink order in about 108 seconds under sustainable working conditions.

Theo's two baristas, Ebrahim and Dae-jung, work the morning rush together, splitting the drink orders between them. They are both experienced, well-liked by customers, and, as far as Theo can tell, competent. But in the past three months, Ebrahim has called in sick four times—more than in the prior year combined. Dae-jung has started arriving late for many shifts. Two part-time staff hired earlier in the year lasted less than a month each.

At a recent staff meeting, Theo raised the issue of service speed. He suggested the team try to move a faster during the rush. Ebrahim looked at the floor. Dae-jung looked at Ebrahim. Neither said anything. Later that day, a junior staff member Lena, who handles pastries and has been at Beanstalk for eight months, found Theo in the back office and knocked on the doorframe.

"Can I ask you something?" she said.

"Of course."

"What makes you think they're not already working as hard as they can?"

Theo did not have an answer.

Question for consideration:

How would you go about determining whether Ebrahim and Dae-jung are working as hard as they can?

Class discussion

RADU: Do you think Ebrahim and Dae-Jung are working as hard as they can?

[Silence for a few seconds.]

CAMILA: I am not sure the case gives us enough to answer it. But I also want to say that I stopped at a coffee shop on my way here this morning and I do not think the workers there were doing their best.

RADU: What made you think that?

CAMILA: They were just standing around.

RADU: Standing around when no one was in line?

CAMILA: No, standing around while I was in line waiting.

RADU: And you thought they should be working harder?

CAMILA: They should work as hard as they can. That is their job.

RADU: Let's examine that for a minute. Haoyu, when you ran your restaurant, should the staff be working at their absolute maximum as long as there are customers waiting? Or even when there are no customers, should they be cleaning, restocking, staying busy the whole time?

HAOYU: If they worked for me, yes. I got my staff to work until they bled.

RADU: Hopefully not into the food.

[Everyone laughs.]

BINESI: But this is wrong. Everyone needs a breather. If you push people with no relief, they burn out. That is not sustainable.

HAOYU: If they burn out, I find someone else.

RADU: Can you? That is actually an interesting empirical question, and I suspect the answer is harder than it sounds. But let's come back to that. Binesi raises something important. If you work people at maximum intensity with no slack, they eventually break down physically, or mentally, or both. We can all agree on that, yes?

MING: I can certainly relate. The full-time MBA program is basically a controlled experiment in how people burn out. *[Uncomfortable laughter.]*

RADU: *[Smiling.]* I will take that as feedback. But you raise a good point. You are all experiencing something like this. When you say you are overloaded, what do you mean exactly?

MING: I mean there is too much work and not enough time. It is not that the work is hard, it is just too much of it. Too many cases, too many readings, too many group projects.

RADU: Good. So overload is not about any single task being impossible. It is about the accumulation of demands exceeding what can be absorbed over time. Now, Farah, can you give me

a cleaner definition? What does it mean to say someone is working as hard as they possibly can?

FARAH: I think it means they are at the maximum pace they can sustain without burning out. Not the fastest they could go for five minutes. The fastest they can go hour after hour and still function the next day.

RADU: I love that definition. Let me give it a name. We will call this a person's *capacity*. A resource is operating at capacity when it is working at the maximum sustainable pace. Not the peak they can reach in a sprint, but the pace they can maintain. [*Writes on the board.*]

Capacity = maximum sustainable throughput rate of a resource

Recall that “pace” for us means the rate at of processing jobs.

MING: But surely you can push past that, at least for a while? Like right before an exam.

RADU: Yes, absolutely. You can work someone beyond their capacity for a while. That is how you get sprints, crunch times, crises. But what happens if you do it continuously?

MING: You get something like the first semester of this program. [*More uncomfortable laughter.*]

RADU: Right. And what are the symptoms?

BINESI: People making mistakes, people not coming to class.

RADU: [*Pausing.*] Does any of that sound familiar from the case?

[*All the students swallow as a moment of recognition passes through the room.*]

CAMILA: Yes, Ebrahim calling in sick, Dae-jung arriving late, the two part-timers who didn't last a month.

RADU: So are these people lazy? Quitters?

CAMILA: The evidence says they are. They are not showing up. That seems like the definition of being a quitter to me.

BINESI: [*Animated.*] But Camila, did you show up to every class last semester?

CAMILA: I choose to exercise my Fifth Amendment rights. [*Everyone laughs.*]

BINESI: We all took a class or two off. Is it because we were lazy? Are we quitters?

[*The students again swallow with recognition.*]

RADU: I think it is valuable to put yourself into the shoes of these workers. Their behavior does not necessarily reveal character flaws. They may be signals that the operation is running its people past their capacity on a regular basis, which brings us back to Lena's question. Let's try to answer it properly. Can we figure out from the case data whether Ebrahim and Dae-jung are working at, or beyond, their capacity?

FARAH: Yes. We have enough data for that. The case says they serve 150 customers in the two-hour rush, split between two baristas. So that is 75 drinks each in two hours.

RADU: Good. And how long does each drink take?

FARAH: The standard time—the sustainable pace with the PFD allowance built in—is 108 seconds.

RADU: Remind everyone what PFD means, I assume you have heard of this before?

FARAH: Oh yes, it means personal, fatigue, and delay allowance. It is the extra time built into the standard to account for the fact that workers are human and they need a moment to recover, to use the restroom, to handle small interruptions. The 20% allowance brings 90 seconds performing time up to 108 seconds.

RADU: Now, how long would it take one barista to complete 75 drinks at this pace?

FARAH: $75 \times 108 = 8,100$ seconds, which ... [*Looks to the ceiling for a moment.*] ... is 135 minutes.

RADU: And how long is the morning rush?

FARAH: 120 minutes.

RADU: Wait ...135 minutes of work in 120 minutes?

MING: Yes, they are working fast than standard pace. They are completing a job every 96 seconds. So they are going faster than the standard.

RADU: And what does that imply?

BINESI: This means the PFD allowance is not being taken. They are not getting the recovery time built into the standard. Their rest.

RADU: The 96-second pace looks efficient from the outside—they are finishing drinks quickly, the line is moving. But it is only achievable because they are not recovering. They are not working within their capacity.

CAMILA: [*Quietly.*] And they have been doing this every morning.

BINESI: This is exactly what happens in social work. Caseloads kept creeping up, and we kept finding ways to absorb them—eating lunch at our desks, skipping documentation, answering calls after hours. From the outside it looked fine. From the inside everyone was disappearing.

HAOYU: But come on people! That 108 seconds is the “standard”. The standard is for chumps. It is a tough market out there in food services. You need to be higher than the standard if you want to compete.

RADU: Are you suggesting that Ebrahim and Dae-jung have a higher than average capacity?

HAOYU: I would hope so. If not, I can find someone who does.

RADU: Suppose you try to hire someone else. Don’t we have the same problem of trying to learn what their capacity is? How can we be sure their capacity is higher than average?

HAOYU: Make them scared enough or feel lucky enough, I am sure we can extract a high capacity from them.

BINESI: Haoyu! I am surprised by you! Their capacity is their capacity; what do you mean by “extract” a high capacity? This just burns them out. Are you okay with that?

HAOYU: If that’s what it takes.

BINESI: Takes to do what? [*Binesi turns away abruptly with a scowl from Haoyu, who does not dare to say more.*]

MING: Can I change the subject a bit?

RADU: Please.

MING: Isn’t there some way to calculate the capacity of a person? Like with a machine, it has a spec sheet which tells you its operating pace, which I guess comes from some calculation. Can’t we do a similar calculation on a person?

RADU: This is one of the deepest questions in all of operations management. [*Pauses.*] A machine has a rated capacity and its manufacturer will tell you how many units per hour it can process under normal operating conditions. Do you know how the manufacturer comes up with that?

MING: I guess through some experiments?

RADU: Yes, they experiment with running it faster and faster until it breaks. Can we do that for humans?

MING: Experiments on humans, yeah, there are rules against that.

FARAH: Well, this is what the standard time is trying to capture. Someone does a time study, observes the task, applies a rating factor for pace, and adds the PFD allowance. That gives you the standard time, which is the basis for calculating capacity.

RADU: That is certainly the intention Farah. But there is a difference between the capacity of a machine and the capacity of a person that I want us to sit with. A machine is often pretty stable. It does not have a difficult morning. It does not care whether the work is meaningful or whether the manager is fair. However, a person's capacity is not a constant. It fluctuates with mood, health, relationships, rest, motivation, and a dozen other things we cannot measure or control. The standard time is an average over a population of average workers. Ebrahim and Dae-jung are not average workers—they are specific people with specific histories.

BINESI: Exactly, each person is different! And even if you could measure someone's capacity exactly, is it right to demand all of it? If I know you can sustainably do 45 tasks per shift, am I entitled to ask for all 45? What if you would prefer to do 35 and have ten minutes to breathe and think?

CAMILA: Why wouldn't we have a right to all of their time? We're paying them for their time? It's an exchange. I give them money and they give me their capacity.

BINESI: I think this is where companies can get it very wrong. Even if you never ask for more than someone's capacity, or even less than their capacity, you can still treat them in ways that deplete it over time. A toxic environment, a dismissive manager, no recognition—all of these erode what someone is able to give. It can become really hard to coax it out of them.

HAOYU: Right, a bunch of lazy cry babies. I am with Camila. If you take my money, you should grow up and do the work. Or I find someone who can.

RADU: How do you know for sure that you can find someone who are better than the two you already have with the same pay, Haoyu?

HAOYU: [*Pauses.*] Oh, I see your point, are you saying if I want someone better, I may need to pay them more? Hmm, that could increase my cost.

RADU: Yeah, that's one possibility. Or more likely, you'll find someone who handles thing well in the very beginning, but gradually burns out too, just like Ebrahim and Dae-Jung.

HAOYU: In that case, I just find yet another new worker.

[*The class laughs.*]

RADU: [*Radu also laughs.*] So it seems like you want to create an operational pattern we might call "let's consume newcomers' passion and fire them when they burn out". Is this the kind of boss you want to be, Haoyu?

[Haoyu thinks for a while and does not answer.]

RADU: Let's take a step back for a moment. Do you think Theo did the calculations we did here, the 135 vs. 120 comparison? Was he aware Ebrahim and Dae-jung were working faster than standard?

[A long pause.]

MING: I am genuinely conflicted about this one. Theo seems to like these workers and believes they are doing a good job. So I feel he finds them to be above average workers. But the fact he asked them to work harder, it really does suggest that he felt they were not working as hard as they could.

BINESI: But I also think he must have been aware things were stressful for them. After all, they were increasing their absenteeism. That should have been a hint to them that they were burning out.

HAOYU: Or that they were lazy.

CAMILA: Haoyu brings up something I have been thinking about. What if the workers do have the capacity, say greater than average capacity, but are choosing not to use all of it? Not burning out, just...holding back?

RADU: Deliberately working below their capacity?

CAMILA: Yes, it's certainly possible.

FARAH: I have seen this. In the engineering firm I worked at, experienced engineers would slow down the new ones to work slower than they were capable. Not obviously. Just a look, a word. The new ones learned quickly to work at the slower pace.

RADU: And why would they do that?

FARAH: Self-preservation. If you show what an energetic engineer can really do, that becomes the new normal. Everyone ends up doing more work.

CAMILA: So there are two failure modes. Either workers are being pushed past their capacity or they are working below their capacity deliberately. From the outside, both might look the same.

RADU: Precisely. And this is why Lena's question is so important. Before you ask people to work harder, it might be important to answer first: do I know what their capacity actually is? And am I asking for more than that, or less?

MING: But even us students here can't agree on whether they are working at or below their capacity. Even the calculation did not seem to help. How can we resolve this?

RADU: Indeed, how to resolve this? I will let you all think about it more yourselves. Let's end class here.

Email

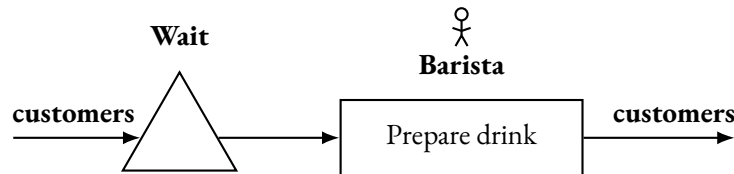
Dear students,

Today's class ended without a resolution, and I want to be honest: I planned it that way. The question of whether Ebrahim and Dae-jung are working above or below their capacity cannot be settled from the

case data alone, and I did not want to pretend otherwise. What we can do in this email is to sharpen the concepts and give you the tools to think about it more carefully.

Defining capacity more precisely. We arrived at a working definition in class: capacity is the maximum sustainable throughput rate of a resource. Let me flesh this out more carefully, because the word “sustainable” is doing a lot of work.

Consider a single worker handling one type of task, say, preparing drinks at a coffee bar. We can draw this as the simplest possible process flow:



The barista is a resource. She processes one customer at a time. Her capacity C is the maximum number of customers she can serve per hour, or equivalently, $1/C$ is the minimum time she needs per customer, without the pace becoming unsustainable.

The key phrase is “without becoming unsustainable.” This is where industrial engineering has developed a useful concept: the *standard time*. The standard time for a task is:

$$T_s = T_{\text{normal}} \times (1 + \text{PFD})$$

where T_{normal} is the time a skilled worker takes at a normal pace, and PFD is the personal, fatigue, and delay allowance, which is the fraction of time that must be reserved for the worker to recover. A typical PFD allowance is 10–20%, depending on the physical and cognitive demands of the work. For the barista in our case, a 90-second performing time with a 20% PFD allowance gives a standard time of $90 \times 1.2 = 108$ seconds per drink.

The capacity of a resource can then be calculated as:

$$C = \frac{\text{available time per shift}}{T_s}$$

For one barista over a two-hour rush with a standard time of 108 seconds per drink:

$$C = \frac{\frac{120 \text{ minutes}}{\text{two-hour rush}}}{\frac{108 \text{ seconds}}{\text{drink}} \cdot \frac{1 \text{ minute}}{60 \text{ seconds}}} \approx 66.7 \text{ drinks per two-hour rush.}$$

This is the maximum the barista can produce while still taking her PFD allowance. If you ask for more, you are asking her to give up her recovery time—to skip lunch, to not use the restroom, to push through fatigue. She may do it. But you are drawing down a reserve that needs to be replenished, and eventually it runs out.

Utilization. Once we have a measure of capacity, we can define a second concept: *utilization*. Utilization is the fraction of a resource’s capacity that is actually being used:

$$u = \frac{R}{C}$$

where R is the actual throughput rate and C is the capacity. A utilization of 1 (or 100%) means the resource is being used to its full sustainable capacity. A utilization above 1 means the resource is operating beyond its sustainable capacity.

In the Beanstalk case: two baristas, each handling 75 drinks in 120 minutes, gives a throughput rate of $R = 75/120 = 0.625$ drinks per minute, or 37.5 drinks per hour per barista. The capacity at standard time is $C = 66.7$ drinks per two-hour rush, or about 33.3 per hour. So:

$$u = \frac{37.5 \text{ drinks/hr}}{33.3 \text{ drinks/hr}} \approx 1.13.$$

A utilization of 1.13 means Ebrahim and Dae-jung are operating at 113% of the standard sustainable capacity.

Taylor's two kinds of soldiering. The class ended with Camila and Binesi on opposite sides of a question: are Ebrahim and Dae-jung working beyond their capacity, or below it? This is not a new dispute. It sits at the center of one of the oldest arguments in the history of management.

Frederick Winslow Taylor, writing in *The Principles of Scientific Management* in 1911, was obsessed with the question of whether workers were giving their full effort. He believed, based on his years observing workers in steel mills and machine shops, that deliberate output restriction—what he called *soldiering*—was “the greatest evil with which the working people of both England and America are now afflicted” [6]. He distinguished two forms.

The first was *natural soldiering*: the simple human tendency to take it easy, to work at a comfortable pace rather than a demanding one. Taylor did not attribute this to laziness exactly, but to a kind of rational self-protection in the absence of any reason to work harder.

The second was *systematic soldiering*: a coordinated, deliberate, collective effort by workers to restrict output. This, Taylor believed, was driven by three rational fears [6]:

1. If workers produced more, management would raise the expected quota, and they would end up doing more work for the same pay.
2. If pay was piece-rate, a higher output would prompt management to cut the per-unit rate, so the worker would earn no more despite working harder.
3. The most widespread fear: if workers became productive enough, some of them would become unnecessary, and jobs would be eliminated.

These are not irrational fears. Each one represents a real pattern that workers had experienced, often repeatedly. Taylor's response was to argue that the fears were misguided that the “scientific management”, by creating genuine gains for both sides, would make fear obsolete. But he was arguing against a history that gave workers every reason to distrust management's assurances. There is a strong case to be made that the origin of the academic field we call “operations management” has its origin with Taylor and his “scientific management” movement.

The Hawthorne experiments. Taylor's observations were anecdotal—a manager's account of what he saw on his factory floor. The question of whether systematic output restriction was real and widespread

got a different kind of answer a decade or two later, in one of the most famous series of studies in the history of organizational behavior.

Between 1924 and 1932, a team of researchers from Harvard, led by Elton Mayo, conducted a series of experiments at the Hawthorne Works of the Western Electric Company in Cicero, Illinois. The best-known phase involved women assembling telephone relays, whose productivity kept rising regardless of what the researchers did—raising or lowering the lights, changing the length of breaks, adjusting the temperature. This puzzling result eventually led to the concept of the “Hawthorne effect”: people work differently when they know they are being observed.

But the final phase of the Hawthorne studies, the Bank Wiring Observation Room, told a different and less optimistic story [5]. Researchers observed fourteen male workers who are wirers, solderers, and inspectors, for six months, with minimal interference. What they found was that the group had developed its own informal standard for what constituted a fair day’s work, and they enforced it with considerable social pressure. Workers who exceeded the group norm, called “rate busters”, were pressured, ridiculed, and sometimes physically tapped on the arm (“binging”) to bring them back into line. Workers who fell significantly short, called “chiselers”, were also sanctioned. The group, not management and not the formal incentive system, was the real regulator of effort.

This was Taylor’s systematic soldiering, documented independently by researchers who were not looking for it. The workers were not simply lazy; they were enforcing a collectively negotiated view of what was fair, in full awareness that working faster would likely result in higher quotas or lower rates.

The right interpretation of Beanstalk is probably that neither the full-capacity burnout story nor the full-capacity soldiering story is exactly right. The case shows symptoms consistent with overload—absenteeism, turnover, declining morale. But it is worth keeping both possibilities in mind, because they call for different responses. Overload calls for relief: more staff, reduced demand, or redesigned processes. Soldiering calls for something harder: rebuilding the kind of trust between workers and management that makes it rational for workers to show what they can actually do.

Haoyu’s model, taken to its extreme form. In class, Haoyu propose a model of constant burnout: if a worker breaks down, you find another. I joked about it, but I want to take the position seriously, because this is far more than just a thought. A 2021 New York Times investigation into Amazon found that the company had, by design, embraced enormous worker turnover. Jeff Bezos reportedly feared that a long-tenured workforce would drift into what he called a “march to mediocrity” [3]. The churn was staggering: attrition running at roughly 150% a year, meaning that Amazon replaced the equivalent of its entire hourly workforce about every eight months.

For a long time this looked. From the outside, like Haoyu’s strategy vindicated at continental scale: the reserve of willing newcomers seemed bottomless, and each one could be run near capacity and replaced before the cost of burnout came due. But run the model far enough and it collides with a wall Haoyu did not anticipate. A later internal memo, leaked in 2022, warned that at these rates Amazon would deplete the available pool of hireable labour in parts of the United States within a couple of years, and in some regions, within months. The strategy of “just find someone else” is only free so long as there is always someone else. Burn through people fast enough, in a big enough operation, and you discover that the newcomer’s passion Haoyu wanted to consume is not an infinite resource but a common one.

Does management have the right to all of a worker’s capacity? I want to end with the question Binesi raised in class, because I do not think it is merely philosophical.

Suppose you could measure a worker's capacity precisely, i.e., you knew, to the minute, how much they could sustainably produce in a shift. Would you be entitled to ask for all of it?

This is where the history of management offers a cautionary example. In 1880, the industrialist George Pullman built a company town on the south side of Chicago to house the workers of his Pullman Palace Car Company. The town was beautiful: clean streets, parks, a library, an opera house, a church. Pullman genuinely believed he was doing his workers a favor: by providing good housing, moral surroundings, and no saloons, he was cultivating the kind of workers who would show up reliably, work hard, and stay. Contemporary observers praised it as a visionary experiment in industrial welfare.

But the workers experienced it differently. Pullman owned everything: the houses, the shops, the church, the newspaper, and rent was deducted directly from wages. Visiting in 1885, nearly a decade before any strike, the reform economist Richard T. Ely already saw the trouble: the town, he wrote, was "benevolent, well wishing feudalism, which desires the happiness of the people, but in such way as shall please the authorities", and, he concluded, fundamentally "un-American" [1]. The grievance was less about wages than about the totality of control; the workers had no sphere of life that was their own.

When the Panic of 1893 cut into company profits, Pullman slashed wages by around a quarter. Some real paychecks fell further once fixed rents were deducted, while holding rents and dividends steady and refusing to negotiate. In May 1894, the workers struck. The strike spread into a national railroad boycott and was broken, after deadly violence, by federal troops. The national commission that Cleveland's administration appointed to investigate found Pullman's paternalism partly to blame and described the company town, pointedly, as "un-American" [7].

The Pullman story is extreme. Few employers today literally own their workers' housing. But its underlying logic is not extinct. The idea that an employer is entitled to organize the conditions of workers' lives in order to maximize their productive capacity, including their sleep, their diet, their commute, their social environment, is alive in subtler forms. The question of where the employer's claim on capacity ends and the worker's own life begins remains genuinely contested.

It would be comforting to file Pullman away as a nineteenth-century curiosity, but the arrangement he pioneered—the employer supplying not just a wage but the housing, the meals, the surroundings in which a worker lives, is very much alive, and often in its kindest-looking forms. Many schools in Hong Kong, and in a number of other places where housing is scarce and expensive, provide teachers with subsidized or free accommodation, sometimes on or beside the campus itself. Universities house faculty; hospitals house residents; mining and oil operations house crews in camps; tech firms run shuttle buses, cafeterias, and sleep pods. Some of this is straightforwardly generous, a real benefit in a market where a teacher or nurse could never otherwise afford to live near work. But notice what every one of these arrangements also does: it shortens the commute, removes a daily friction, keeps the worker close, and makes leaving the job costlier, because leaving now means losing your home, not just your paycheck. The same act can be a gift and a claim on capacity at once, and usually it is both. This is why the arrangement is so durable and so hard to judge cleanly.

Pullman's mistake was not the idea of housing his workers; it was owning everything, leaving them no sphere that was their own, so that when he cut wages there was nowhere to stand. The modern versions survive precisely because they stop short of that totality—they take a large share of the worker's life, but not all of it. Where exactly the acceptable share ends is the question Binesi was really asking, and it has no settled answer.

Taken one step further we may ask the bold question: What is the price on employee's *last* unit of capacity? The first unit of someone's effort is cheap; they came to work willing to give it. The last unit—the part that eats into their recovery, their evening, their reserve, is the most expensive thing you can buy from them, because it is the part they least want to sell. This is not a new observation. It is the core of what economists call the labour–leisure trade-off, an idea traceable at least to William Stanley Jevons, who argued in 1871 that the pain of labour grows the longer one works, so that beyond some point the disutility of one more hour outweighs the wages it earns [2]. The modern statement is the upward-sloping labour supply curve: to buy an additional hour, you must pay a higher rate, because the leisure being given up is worth more and more to the worker [4].

A worker will hand over the pace that fits comfortably inside their capacity for the wage you posted. To buy the pace above it, the sprint that skips the PFD allowance, you are no longer paying for labor; you are paying someone to spend down a reserve they will have to replenish on their own time. Rationally, that should command a steep premium—overtime rates, hazard pay, the wage it takes to make burnout worth it to the person burning. Haoyu's instinct, "I'll just find someone who does it for the same pay," assumes that last, most expensive unit of capacity can be had at the price of the first. It often cannot. Either you pay far more for it than you expect, or more commonly, you don't pay for it at all, and the worker covers the cost via absences and errors, and eventually in leaving. Ebrahim's four sick days are, in a sense, an invoice.

Warmest, Radu.

References

- [1] Richard T. Ely. Pullman: A social study. *Harper's New Monthly Magazine*, 70:452–466, February 1885.
- [2] William Stanley Jevons. *The Theory of Political Economy*. Macmillan, London, 1871.
- [3] Jodi Kantor, Karen Weise, and Grace Ashford. The amazon that customers don't see. *The New York Times*, June 2021. Published June 15, 2021.
- [4] Lionel Robbins. On the elasticity of demand for income in terms of effort. *Economica*, (29):123–129, 1930.
- [5] Fritz J. Roethlisberger and William J. Dickson. *Management and the Worker*. Harvard University Press, Cambridge, MA, 1939.
- [6] Frederick Winslow Taylor. *The Principles of Scientific Management*. Harper & Brothers, New York, 1911.
- [7] United States Strike Commission. Report on the chicago strike of june–july 1894. Technical report, U.S. Government Printing Office, Washington, DC, 1895. Commission headed by Carroll D. Wright.

Extensions

[Posted as homework on the course's website for this class session.]

1. The standard time formula includes a PFD allowance, which is set by an industrial engineer based on the physical and cognitive demands of the task. Why might this be challenging? Who might influence the decision on this amount?
2. A machine has a rated capacity determined by its manufacturer. A person's capacity is estimated by a time study, where an industrial engineer observes workings undertaking various tasks and uses a stop watch to measure various durations associated with the tasks. What are the key differences between these two kinds of capacity measurement? Under what circumstances would you trust a time study, and under what circumstances would you be skeptical of it?
3. Theo could respond to his capacity problem in at least three ways: hire more staff, reduce the number of drinks offered during the rush to speed up preparation, or stagger arrival times by offering a discount for orders placed before 7:30 a.m. Which would you recommend, and why?
4. The Hawthorne Bank Wiring Room study found that workers enforced their own output norms, independently of what management set. What conditions at Beanstalk might encourage or discourage this kind of informal norm-setting among the baristas? If Ebrahim and Dae-jung have developed informal norms, what are they likely to be, and why?
5. Utilization above 100% is mathematically possible but physically unsustainable. Sketch what you would expect the relationship between utilization and absenteeism to look like as a graph. At what utilization level would you expect absenteeism to start rising noticeably? What other signals, besides absenteeism, might indicate that a resource is being pushed past its capacity?
6. Taylor argued that systematic soldiering was rational given the incentive structures workers faced. Design a compensation structure for the Beanstalk baristas that would reduce the rational incentive to soldier without simply demanding more output for the same pay. What tradeoffs does your design involve?
7. A hospital emergency department, a commercial kitchen, and a software development team all have workers whose capacity is hard to measure. Pick one of these and describe the specific challenges involved in estimating capacity. What proxies or indirect measures might you use? What would you be missing?
8. Google's offices are famous for amenities—free meals, gyms, recreational spaces—that keep employees on campus for longer hours. Using the concepts from today's class, construct an argument that these amenities are a capacity management strategy rather than a benefit. Then construct the strongest counter-argument.
9. Theo's instinct was to ask his baristas to work harder. Before doing that, what information should he have gathered? Write a one-page memo from Theo to himself, drafted the evening after Lena's question, that outlines what he now needs to find out and how.

Reflections

[Posted as a reflection exercise on the course's website for students to write about in a journal due at the end of semester.]

1. Have you ever been asked to work harder by someone who had no idea how hard you were already working? What did that feel like? What did you do?
2. Do you know what your own capacity is? Not in the abstract, but concretely: how much can you sustainably produce, in the kind of work you do, before the quality starts to slip or you start to break down? How did you learn this, and how confident are you in it?
3. Haoyu said he could just find someone else if his workers burned out. Is this true in the work you know? What is actually lost when an experienced worker leaves, and how long does it take to recover?
4. Think of a job you have held where you were working below your capacity. Was it natural soldiering, systematic soldiering, or something else? What was the reason, and do you think the employer was aware of it?
5. Binesi asked whether management is entitled to all of a worker's capacity. Where do you come down on this, and why? Does your answer change depending on whether the worker is paid hourly or salaried?
6. Ming said the MBA program felt like working beyond her capacity. Is that a design flaw or a feature? Would anything be lost if the program were redesigned to keep students comfortably within their capacity at all times?
7. Think of a manager or supervisor you have respected. How did they know when you had more to give, and how did they ask for it? What made it feel different from Theo's approach?
8. The email notes that Google's amenities keep workers on campus for longer hours. Have you ever worked somewhere whose benefits or culture had this effect on you, deliberately or not? How did you feel about it at the time, and how do you feel about it now?