THE ECONOMIC JOURNAL

JANUARY 1992

The Economic Journal, **102** (*January* 1992), 1–8
Printed in Great Britain

HOW TO COUNT TO ONE THOUSAND*

Joel Sobel

People make mistakes. If the mistakes occur in a well understood way, then people can organise their activities to reduce the cost of making the mistakes. In this paper I look at a particular repetitive task that must be performed by error-prone agents. The process by which errors are made is known. Consequently the activity may be performed in a way that takes into account the possibility of mistakes. The task I examine can be viewed as counting to one thousand. The introduction describes the task and motivates the results.

Imagine that you must make a cash transaction and have just received change in one dollar bills. You expect a total of $1,000. How do you make sure that it is all there? You start counting. But what happens if you get confused or distracted along the way and forget the count? You must start over in order to be certain that you were given correct change. This paper investigates the properties of counting schemes that allow you to subdivide the task. First you count twenty five bills, say, and put them aside. Then you count another twenty five bills, and so on. The procedure has an advantage: If you make a mistake you do not need to start the task over from the beginning, you need only recount the current stack of no more than twenty five bills. It also has a disadvantage: After you have divided the bills into stacks of twenty five you must go back and confirm that there really are forty stacks. The counting scheme that minimizes the expected time it takes to confirm the size of the stack involves dividing the counting into substacks and then counting these stacks. If the number of dollars to be counted is large enough, then the counting process has several layers where stacks are grouped into stacks of stacks, which are grouped into stacks of stacks of stacks, and so on until the task is completed. A hierarchical structure arises simply because there is a positive probability of making mistakes.

I assume that people know when they have made a mistake and that everything that has not been saved (or put aside as a separate stack) is lost. Under these assumptions I deduce properties of the counting scheme that minimises the expected time needed to complete a task. If there is a constant probability of making a mistake at each step of the counting process, then

* I am grateful to the National Science Foundation and the Sloan Foundation for financial support.

(ignoring integer problems, which are negligible for large tasks) tasks are subdivided into units that depend only on the probability of making mistakes and not on the size of the entire job. So if it is efficient to count to one thousand in units of twenty five, then it is also efficient to count to one million by twenty fives. As a consequence the efficient number of levels needed to perform a task increases with the logarithm of the size of the task. The average time to complete a step in a task increases linearly with the size of the task.

The paper gives a formal treatment of an observation of Herbert Simon (1962). Simon suggests that all types of organisational structures evolve into forms that nest integrated units within integrated units. This hierarchical organisation is able to isolate the impact of disturbances and lessen the effect of shocks and mistakes. Simon (1962) tells a parable of two watchmakers, one first assembles subunits and then puts these subunits together to complete the watch, and the other who does not subdivide the construction. Many pieces must be assembled to make a watch and there is a chance that a watchmaker will be disturbed during the process. Simon argues that if a disturbance causes a watchmaker to reassemble any partially completed unit, then the watchmaker who first builds subunits will be able to complete more watches than his competitor. This paper provides a mathematical description of the parable.

In a series of papers Sah and Stiglitz (1985) and (1986) study the properties of different ways of making decisions in an environment in which people make mistakes. They compare organisations where a series of individuals must approve a project to ones where any single individual may approve a project. They find that the first form of organisation leads to a lower approval rate than the second. They identify other properties characteristic of different ways to organise decision making.

The simple framework of this paper lets me say things about the optimal size of a task. By deriving the task size that has the minimum average cost per step, I identify an efficient scale of operation. Other studies on hierarchical structures in firms try to deduce the optimal size of the firm. Calvo and Wellisz (1978) and (1979) and Williamson (1975) develop a hierarchical theory of the firm based on incentives. Only workers at the lowest level of a pyramid contribute directly to output. Workers only work if there is a large enough probability that their boss will catch them if they try to shirk. Higher level workers are needed only to supervise their immediate underlings. Assumptions about the supervision technology determine the efficient number of layers between the productive workers and the manager. Geanakoplos and Milgrom (1984) derive properties of an organisation assuming that individuals inside the organisation differ in their access to and ability to process information. They ignore incentive problems. I also ignore incentive problems. The advantages of a particular form of organisation stem only from the possibility of making mistakes. The optimal task size is not determined in the counting model. In Section II I show that expected time needed to count to n is (approximately) a linear function of n; there are essentially constant returns to scale in the size of the task. I describe a different model in Section III. Here the counting is done by an assembly line. Individual counters have the special job of performing a fixed step of the task (for example, counting the tenth bill). While these specialised workers are unlikely to get confused, each worker must be present in order to have a successful count. If there is a positive probability of absenteeism and workers can be trained to perform only a single step of the task, then this type of assembly line exhibits decreasing returns to scale, but if there are workers who can fill several positions, then average costs first decrease and then increase.

I. EXPECTED COMPLETION TIMES OF ATOMIC TASKS

In this section I discuss the expected time it takes to successfully complete a task with n steps without making subdivisions. A task is atomic if it is not subdivided. It is efficient to subdivide large tasks into smaller units. The next section studies the optimal way to subdivide large tasks.

I assume that the probability of succeeding in the $(k+1)$th step of the process is $p_{n-k} \in (0, 1)$ and the time it takes to perform the $(k+1)$th step of the process is a_{n-k}. When p_k and a_k are independent of k, the expected time of completion is just the expected time to achieve n consecutive successes from independent Bernoulli trials. The time needed to complete n tasks is

$$\left(\sum_{i=1}^{n} q_i a_i\right) \Big/ q_0, \tag{1}$$

which reduces to $(1-p^n)/[(1-p)p^n]$ in the case where p_i is independent of i and $a_i = 1$ for all i. (I omit the derivation of this and other formulas from the paper; details are contained in an appendix, which I will supply to the interested reader upon request.)

Imagine that there are n workers described by their speed a_i and their accuracy p_i. It is possible to make strong statements about what types of workers will be asked to do which jobs. If the workers differ only in their speed, then it is optimal to assign the jobs to workers in order of their speed, with the fastest workers doing the first steps in the task. The intuition is clear: The first step must be repeated whenever there is an mistake. If one step can be done more quickly with no sacrifice of accuracy, then it is efficient for it to be the first task. Similarly, if the workers differ only in their accuracy, then it is efficient to assign the workers in increasing order of accuracy, with the more accurate workers doing the later steps. Since the later in the task a mistake is made, the more time is lost, expected completion time is reduced by placing the most accurate workers at the end of the counting process.

When different workers perform different steps in the counting process, the form of the expected time of completion has implications about how workers will be trained. Assume that workers begin with equal speed and accuracy, but that through training their skills can be improved. Further assume that there is a continuously differentiable function that determines the effectiveness of money spent on training. Under these conditions it follows from (1) that it is never efficient to improve the skills of workers uniformly. Instead, if training improves accuracy, then workers who perform the later steps should be trained more than the other workers. If training improves speed, then the workers who perform the first steps should be given the most training. A small probability of making mistakes leads to asymmetric development and use of workers' skills.

Now suppose that the workers have identical skill levels and that I wish to hire and allocate workers so that in a steady state all of the workers are occupied. If N_k denotes the number of different projects that are k steps from completion, then in a steady state equilibrium, $N_{k-1} = pN_k$ for $k = 1, 2, \ldots, n$. Hence the number of workers assigned to one step should be proportionally greater than the number of workers at the next step of the task. The assignment of labour inside a given n-step task leads to a hierarchical structure with more (or faster) workers assigned to the lower tasks. This characteristic arises only because the lower tasks must be performed more frequently than the higher ones. The ratio of the number of workers at one step to the number of workers in the next higher step depends only on the accuracy of workers at the step closer to completion. In this model the relative accuracy of a worker determines her span of control.

Finally, since the expected time of completion is equal to $(1-p^n)/[(1-p)\,p^n]$ when each step succeeds with probability p, the average time per step to complete an n-step task, $(1-p^n)/[n(1-p)\,p^n]$, is increasing in n. The production process exhibits decreasing returns to scale. In the next section I show that if subdivision is possible then the production process exhibits constant returns to scale.

II. EFFICIENT SUBDIVISION OF TASKS

I have computed the expected length of time needed to complete a task of fixed length. In order to finish a large task in the least time it will generally be optimal to subdivide. This section uses the result of the Section I to characterise the optimal number of subdivisions under the assumption that all steps succeed with probability p ($p_i = p$ for all i) and one unit of time is needed to perform a step ($a_i = 1$ for all i).

Suppose that the task has n steps. The counter first handles a subtask of size r; then finishes the rest of the task. Ignoring integer restrictions on the size and number of bundles, which can be shown to be negligible when n is large, a dynamic programming argument demonstrates that the counting problem reduces to finding a pile size r to solve:

$$\min\,(1-p^r)/[p^r(r-1)], \qquad (P)$$

when $n > 1$, the corresponding optimal choice of k satisfies $k = (n-1)/(r-1)$, and the minimum expected time to complete the project is

$$[(n-1)\,(1-p^r)]/[p^r(r-1)\,(1-p)]. \qquad (2)$$

The objective function in (P) is independent of n, therefore the optimal pile size is also independent of n. Hence, ignoring the integer problem, it is efficient to create bundles of the same size r regardless of the size of the entire task. An intuition for the result comes from the logic of dynamic programming and the observation that the expected completion time of an atomic task is convex in the number of items in the task. When the counter creates the first stack, she reduces the size of the task. The efficient stack size in the new, smaller problem is equal to the size of the first stack. Otherwise, by convexity, the counter could reduce the expected completion time of the original problem by averaging the stack sizes.

Expression (2) shows that the counting process exhibits (approximately) constant returns to scale. The average time to complete a step in an n-step process is $(n-1)/n$ times the constant $(1-p^r)/[p^r(r-1)(1-p)]$. When I examined atomic processes in Section I, the counting process exhibited decreasing returns to scale because larger tasks carried the risk of more costly mistakes. If it is feasible to subdivide the task, then the costs of mistakes are bounded. Indeed, if the size of the task is doubled, then one feasible way to perform the task is to divide it into two parts, perform each half separately, and then combine the two completed halves. Provided that it is not costly to combine the two halves, and my assumptions guarantee that it is not, doubling the size of the task can (essentially) no more than double the expected completion time.

The solution to (P) is characterised by a first-order condition since $(1-p^r)/[p^r(r-1)]$ is convex. Hence the optimal size of a pile is determined as the solution to

$$p^r = 1+(r-1)\log p. \tag{3}$$

The optimal choice of r characterised by (3) is increasing in p.

I can also compute the number of layers there are in the process. If each stack consists of r items, then n/r stacks are needed to count the n items, n/r^2 stacks are needed to count the first set of stacks, and so on. Therefore if there are l levels, then total number of stacks k is equal to $\Sigma_{i=1}^{l} n/r^i = k$. Since $k = (n-1)/(r-1)$, it follows that $l = (\log n)/(\log r)$. That is, the number of levels increases with the logarithm of the size of the task. As expected, increases in the probability of making a mistake lead to smaller subdivisions and as a result more layers in the counting pyramid.

III. ASSEMBLY LINES

It is natural to assume that people who must perform the task of counting make mistakes. Specialisation might solve the problem. If a sequence of individuals could be hired with each one assigned a particular step in the process, then there is little reason to think that anyone would make a mistake. I have in mind a situation where one thousand people are hired (when $n = 1{,}000$). For each $i = 1, 2, \ldots, 1{,}000$, person i is assigned the job of counting the ith element in the sequence. Imagine that the employees are ordered in an assembly line, with worker number 1 counting the first bill, and passing the task to worker number 2, who counts and passes to the third worker, and so on. Provided that workers can remember their index i and the identities of the workers with adjacent indices, there should be no danger of getting confused. Assigning specific tasks to specific workers could eliminate failures due to human fallibility. The job of remembering is divided into manageable pieces and divided among different workers. The assembly line technology is extremely sensitive to absenteeism. if one of the workers fails to do his job, then the job may go undone. In this section I investigate the assembly line technology.

n steps must be performed in order to accomplish the task. A pool of workers are available to perform individual steps in the task. Workers are identical and each worker is able to perform any step. However, there is a probability that a worker will not show up. Assume that the probability a particular worker shows up is $\pi \in (0, 1)$, and that these probabilities are independent across workers. If one worker is hired for each step, then the probability that the task is done is π^n, and the expected completion time is π^{-n}. Assume that the wage per worker per unit of time is equal to one and that all workers hired must be paid (whether they show up or not). The second assumption makes sense if it is not possible to write a labour contract where payment is contingent upon showing up, which is a plausible restriction if 'not showing up' is interpreted as a failure to supply an unobservable minimum effort level. It follows that the expected cost of completing the task if one worker is assigned to each step is equal to $n\pi^{-n}$. Assembly lines do not work well if workers can perform only one step and all workers are essential to the production process. I want to investigate the implications of relaxing each of these assumptions.

First assume that it is possible to hire many workers for each step in the process. Assume that a worker must be assigned to a particular step, but only one worker capable of performing each step need be present in order to complete the task. Consequently, if k workers are hired to perform step i, then the probability that the step will actually be done is equal to $1-(1-\pi)^k$. The manager's objective is to hire k_i workers for step i $(i = 1, 2, \dots, n)$ to minimise the expected total wage bill. It is cost minimising to hire the same number of workers for each step. If k workers are hired for each step, then the probability that the task will succeed is $[1-(1-\pi)^k]^n$, the expected time to completion is $[1-(1-\pi)^k]^{-n}$, and the expected cost is $nk[1-(1-\pi)^k]^{-n}$. The average cost per step in an n-step task, $k[1-(1-\pi)^k]^{-n}$, is increasing in n for all k: The technology exhibits decreasing returns to scale.

It is wasteful to hire several workers for each step to minimise the costs associated with absentees. If two workers arrive for the same task, then one of them contributes nothing to output. Suppose that some workers can be hired to perform any one of a large number of tasks should the need arise. Hiring 'jack-of-all-trades' as substitutes is likely to be a way to lower the average cost of completing the task. A complete analysis of this possibility requires a careful consideration of the relative wage that must be paid to a versatile substitute and how the wage varies with the range of steps that the worker is able to perform (and her accuracy). Rather than carry out that extended exercise, all I will do is point out that if versatile substitutes are available, then the average cost of production is likely to first decrease and then increase in the size of the task. Therefore efficient size of a 'firm' operating with an assembly line technology with substitutes is positive and finite. To illustrate this assume that one specialised worker is hired for each step, and that as above each of these workers shows up with probability π. Further assume that an additional worker is hired who is able to perform any of the steps in the task. This worker is effective with probability ρ and must be paid λ. The probability that the task will be completed is equal to $\pi^n + n\pi^n(1-\pi)\rho$, and the expected wage bill is $(n+\lambda)/[\pi^n + n\pi^{n-1}(1-\pi)\rho]$. In this case the average cost per step is increasing in n for large n, but decreasing in n for small tasks.

V. EXTENSIONS

My purpose has been to show the effect of the possibility of errors on organisational form. I carried out the computations in a setting too simple to describe any real phenomena. In this section I discuss several extensions of main ideas.

There are several ways to generalise the error structure in the model. I have assumed that the objective was to perform the counting task accurately in the minimum expected time. In many circumstances complete accuracy may be too much to hope for. Assume that the counter begins with a probability distribution on the number of items and is allowed to decide how long to count and what number to guess when the counting is over. It is natural to assume that utility decreases with the time spent counting and increases with the accuracy of the guess. This problem is more difficult than the one I have discussed. Subdivision is a good idea, but there is no reason to think that the divisions will have a regular pattern. One special case is easy. If there is a constant probability that the next item counted is the last one, then the counter learns nothing about the size of the task from any step. It is best either to spend no time counting (and make the best guess regarding the size of the task), or to follow the counting procedure characterised in Section II.

I assume that people know when they make a mistake. Alternatively I could assume that there is a fixed probability of making a mistake of any magnitude and that the counter does not know when a mistake is made. There need not be a gain from subdividing a task in this situation. However it is not difficult to think of situations where subdividing tasks is valuable even with this type of error structure. Rubinstein (1988) describes a game where two prisoners must coordinate the date of an escape months in advance. They cannot communicate but are able to watch the sun rise and set. If it is difficult to keep track of the days, then neither prisoner will be sure whether his count agrees with the other prisoner's. Coordination may be difficult to achieve. When the prisoners wish to coordinate on an escape date several months in the future, their chances would improve a great deal if they could see the moon as well as the sun. While the prisoners still lack common knowledge of the date, both could be more confident that they have correctly counted three full moons and five days, for example, than eighty nine days.

Rubinstein's coordination problem is a strategic one. Whether or not the escape succeeds depends on both the prisoners making their break at the same time. Bendor and Mookherjee (1987) have investigated another gametheoretic setting where a hierarchical structure can be used to improve outcomes. They study an n-player variant of the repeated prisoners' dilemma with imperfect monitoring and show that there are parameter values (group size, the level of idiosyncratic uncertainty, the amount of monitoring, and rate of discounting) for which it is possible to sustain cooperation by dividing the players into subgroups, randomly monitoring the effort of subgroups, and punishing subgroups that are found to be playing noncooperatively, even though it is not possible to sustain cooperation without subdivisions.

University of California

Date of receipt of final typescript: May 1991

References

Bendor, J. and Mookherjee, D. (1987). 'Institutional structure and the logic of ongoing collective action.' *American Political Science Review*, vol. 81, pp. 129–54.

Calvo, G. and Wellisz, S. (1978). 'Supervision, loss of control, and the optimum size of the firm.' *Journal of Political Economy*, vol. 78, pp. 943–52.

—— and —— (1979). 'Hierarchy, ability, and income distribution.' *Journal of Political Economy*, vol. 79, pp. 991–1010.

Geanakoplos, J. and Milgrom, P. (1984). 'A theory of hierarchies based on limited managerial attention.' Mimeo, School of Management, Yale University.

Rubinstein, A. (1988). 'Comments on the interpretation of game theory.' ICERD Discussion Paper, London School of Economics.

Sah, R. and Stiglitz, J. (1985). 'Human fallibility and economic organization.' *American Economic Review*, vol. 75, pp. 292–7.

—— and —— (1986). 'The architecture of economic systems: hierarchies and polyarchies.' *American Economic Review*, vol. 76, pp. 716–27.

Simon, H. (1962). 'The architecture of complexity.' *Proceedings of the American Philosophical Society*, vol. 106, pp. 467–82. Reprinted as chapter 4 in *The Sciences of the Artificial*. Cambridge and London: Massachusetts Institute of Technology, 1969.

Williamson, O. (1975). 'Hierarchical control and optimum firm size.' *Journal of Political Economy*, vol. 75, pp. 123–38.