*Industrial and Corporate Change*, 2020, Vol. 29, No. 1, 125–142
doi: 10.1093/icc/dtz069
Advance Access Publication Date: 12 December 2019
Original article

OXFORD

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

# Relining the garbage can of organizational decision-making: modeling the arrival of problems and solutions as queues

**Peter W. Glynn,[1] Henrich R. Greve,[2,*] and Hayagreeva Rao[3]**

[1]School of Engineering, Stanford University, 475 Via Ortega, Stanford, CA 94305, USA. e-mail: glynn@stanford.edu, [2]INSEAD, 1 Ayer Rajah Avenue, Singapore 138676, Singapore. e-mail: henrich.greve@insead.edu and [3]Graduate School of Business, Stanford University, 655 Knight Way, Stanford, CA 94305, USA. e-mail: hrao@stanford.edu

*Main author for correspondence.

## Abstract

The garbage can account of organizations where problems, solutions, and people chase each other is often invoked but rarely studied since its publication 44 years ago. It has been critiqued for being a metaphor rather than a model, and offering a deterministic rather than stochastic account. We reline the garbage can model of organizational decision-making by modeling the arrival of problems, people, and solutions as queues that get matched randomly. We show that queuing models allow us to understand the effect of using either experts, supervisor approval, teams, and deviation from supervision on problem resolution and oversight. Our approach shows that manager approval increased the standard deviation of problem resolution, whereas queues are processed faster and have lower variance when there is oversight or teamwork, or when manager approval is bypassed due to independent action by problem solvers. It also shows the costs of using an organizational hierarchy to address problems with different levels of difficulty, or specialization to address a mixture of fundamentally different problems. Thus, a stochastic garbage can model provides insights into why organizations make many decisions but often fail to resolve problems!

**JEL classification:** D2, D7, D9

Behavioral accounts of organizations build on the idea of bounded rationality (Simon, 1947) and describe how the cognitive limitations of individuals, especially, with respect to attention, underpin organizational responses to problems. A prominent exposition is the garbage can model of organizations that was geared to initially understand organizations having problematic preferences, unclear technology, and fluid participation and was outlined in an influential *Administrative Science Quarterly* (ASQ) paper that has received more than 10,000 citations since then (Cohen *et al.*, 1972). A central insight of the paper is that choice opportunities (any decision forum) collect problems, solutions, and decision makers. There is a continual stream of problems, people, solutions and choices, and every now and then they coincide in a decision forum, and a decision is produced (Cohen *et al.*, 1972). Hence, temporal

© The Author(s) 2019. Published by Oxford University Press on behalf of Associazione ICC. All rights reserved.

sorting of problems and solutions is crucial, and decision forums were termed garbage cans to emphasize this point. Indeed, as Cohen *et al.* (2012: 26–28) note 40 years later,

> "These extensions [beyond bounded rationality] imply a temporal understanding of events, in contrast with an intentional or consequential one. The framing of decisions may be to a considerable extent determined by temporally unfolding processes of participation and attention."

The garbage can model of organizations has been described as an "arresting and radical vision of organizations" (Padgett, 2013: 473). Since March and Olsen (1979) presented a number of case studies showcasing the core elements of the garbage can model, it has influenced political science through models of how Congressional agenda setting is triggered by policy windows (Kingdon, 1984), or models of how governmental priorities in budgets are shaped by the allocation of media and Congressional attention (Jones and Baumgartner, 2005). In organizational studies, it has also received significant attention (e.g., Levitt and Nass, 1989).

The model has not been uncontroversial. A belated critique in the political science literature depicts the garbage can model more as a metaphor more than a model, and a deterministic rather than stochastic account of how people, problems, and solutions are chasing each other (Bendor *et al.*, 2001). Padgett (2013) attributes the critique to "the extreme herding behavior that was built into details of the 1972 ASQ code. While not inconsistent with the verbal theory, this nonessential design feature—that both people and problems respond by the same rule of flowing to choices with the lowest energy deficits—drove a number of the original simulations results" (p. 478).

Since then, a number of studies have demonstrated support for the basic findings of the garbage can model (see Lomi and Harrison, 2012). Two in particular deserve mention. For example, Lomi *et al.* (2012) study the match between bugs and programmers in open source software. They find that there is a core group of programmers surrounded by a larger group of nomads, hard problems attract fewer programmers, and the larger the number of programmers drawn to a bug, the more people it attracted. Knudsen *et al.* (2012) conducted a laboratory experiment to discern if the assumptions of the ASQ paper were realistic. They found that 57% of the time, people do what the ASQ model suggested: they are attracted to choice settings with the lowest energy deficits. However, in that 34% of the time, they do the opposite and go to places with the highest energy deficit, so there is more heterogeneity than the ASQ paper implied. They found that the deviations occurred when people had unrestricted access to people but limited hierarchical access to problems that inhibited the herding critiqued by Bendor *et al.* (2001). A striking finding was that randomization solved problems better than the ASQ model and the experimental subjects; thereby, suggesting that the world is not a moving landscape in which people and problems and solutions move.

In general, research on the garbage can has been powered by computer simulations rather than by actual empirical studies, and these simulations focus on isolating specific elements of the garbage (e.g., fluid participation) in order to shed light on how temporal sorting occurs through mechanisms, such as fluid participation and access structures. In a review of these studies, Padgett (2013: 473) lamented that

> we seem to prefer to isolate and study specific mechanisms rather than to revel in their collective interaction and cacophony… Cohen, March, and Olsen then wanted simultaneously to be scientific engineers and to be pioneering botanists. The first side of them survives in their progeny today, whereas the second side has faded.

These considerations provide the motivation for our paper to develop a stochastic model of temporal sorting. The modeling apparatus used in the original paper was a FORTRAN program fed with specific parameters, a modeling approach that is both weaker in conclusions than the formal modeling and analytical approximation that is possible now, and also sufficiently cumbersome that it does not allow the careful investigation that can now be applied to models. We reline, as it were, the garbage can, by modeling the arrival of people, problems, and solutions as queues that are stochastically matched. Our work is not intended as a replication with novel techniques, but as an extension, so our formulation and application of the model will subtract some elements found in Cohen *et al.* (1972) and add others. Our work does share their goal of developing a "behavioral theory of organized anarchy" (Cohen *et al.*, 2012: 2), so our model results should be seen as predictions of organizational behavior given situations that resemble those of the model. Because each model specification omits some features of organizational life, such as costly and slow communication in teams, it should be possible to pinpoint the sources of empirical deviations from our predictions. Some model findings can also be used to construct a normative theory, because they do give suggestions on what organizational designs will increase speed and reliability. Again, each model specification omits some features of organizational life, so these suggestions should be applied with caution.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

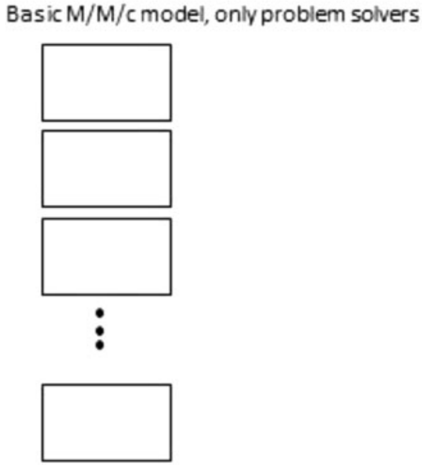


**Figure 1**. Independent problem solvers.

Our technical approach is queuing theory, which we choose because it is a powerful analytical approach that is a close match to the problems addressed by the garbage can model. We show that our extension of the garbage sheds light on how organizations make decisions (such as flight or oversight) often, but often fail to solve problems in a timely manner. We begin with a basic formulation of the queuing model, and add forums or procedures for decision-making, such as independent problem solvers, approval by manager, team work, and independent action. Thus, we examine distinct decision-making processes just as the original paper, but make different selections on which processes to investigate.

## 1. Garbage cans and queues: a basic model and extensions

A queueing model has three pieces: a process that determines the arrival of problems, a set of one or more problem solvers, and a probability distribution that determines the duration of solution time for problems (see Adan and Resing, 2015). A very simple queue would have a Poisson process that determines problem arrival epochs, one problem solver, and an exponential distribution for the total time required to solve each problem. It would be a way to approximate a single employee or vendor serving customers who have requests of varying complexity. In this simple queuing model, every problem is solved, though the time to solve it is determined stochastically and hence has some variation. We will start with a simple elaboration of this model, which introduces $c$ different problem solvers rather than one, and assigns the problem to the first available problem solver. The model is (Figure 1):

1. Problem arrival, Poisson with arrival rate $\lambda$ (or, equivalently, mean inter-arrival time $1/\lambda$);
2. Problem solvers, $c$ in total of which the first available addresses the first-arriving queued problem;
3. Problem solution, exponential distribution with mean $1/\mu$.

This is called an M/M/c model. This specification is obviously a simplification to make the model tractable, but has some useful features:

1. Problem arrivals are unpredictable, as one would expect when an organization solves problems rather than conducts routine tasks.
2. Busy problem solvers do not take on additional problems, also a realistic model of how organizations seek to address unexpected problems.
3. The distribution of solution time has much probability weight on quick resolutions, but also some very lengthy one, which matches a world in which many unexpected problems turn out to be simple, but some require considerable work to be solve.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

We use this model as a baseline, and note its key features. The proportion of time that at least one problem solver is busy is $1 - \pi_0$, where

$$\pi_0 = \left( \sum_{k=0}^{c-1} \frac{\rho^k}{k!} + \frac{\rho^c}{c!} \frac{1}{1 - \rho/c} \right)^{-1}$$

where $\rho = \lambda/\mu$ and $\lambda < c\mu$ is assumed. (If $\lambda \geq c\mu$, the amount of work in the system increases without bound over time. This means that if $\lambda \geq c\mu$, the load on the organization is too high, and additional problem solvers are needed to handle the arriving workload.) The likelihood that a problem arrives and finds all problem solvers busy is (Appendix 1 has a guide to solutions)

$$\pi_b = \pi_0 \frac{\rho^c}{c!} \frac{\rho/c}{1 - \rho/c}$$

But we are more interested in the following quantities:

1. The average number of problems in the queue is

$$\frac{\rho/c}{1 - \rho/c} \pi_b$$

When $\rho$ is close to $c$, this can be approximated by $\frac{1}{1-\rho/c}$.

2. The average total solution time from arrival to completed solution is

$$E(S) = \frac{1}{\mu} + \frac{1}{\lambda} \pi_0 \frac{\rho^2}{c!} \frac{\rho/c}{1 - \rho/c}$$

When $\rho$ is close to $c$, this can be approximated by $\frac{1}{\mu c} \frac{1}{1-\rho/c}$.

3. The standard deviation of the total solution time is $\sqrt{E(S^2) - (E(S))^2}$, where

$$E(S^2) = \frac{2\pi_b}{1 - c + \rho} \frac{1}{c^2 \mu^2} \frac{1}{(1 - \rho/c)^2} + 2\left(1 - \frac{\pi_b}{1 - c + \rho}\right) \frac{1}{\mu^2}$$

if $\rho \neq c - 1$. If $\rho$ is close to $c$, the standard deviation is approximately

$$\frac{1}{\mu c} \frac{1}{1 - \rho/c}$$

If $\rho = c - 1$,

$$E(S^2) = 4\pi_b \frac{1}{\mu^2} + \frac{2}{\mu^2}$$

A set of graphs can show these effects. In Figure 2, we compare single problem solver (so $c = 1$) with 3 ($c = 3$) and 10 ($c = 10$), and show the average number of problems in the queue, the expected time to solve a problem, and the standard deviation of a problem-solving time. The graphs show how multiple problem solvers reduce the expected solution time and its standard deviation, with the graphs being very similarly shaped for these two statistics. Naturally, both increase significantly as the load ($\rho$ increases, which the problem arrival rate divided by the expected solution time).

The graphs also have substantive implications. A simple one is that they suggest some of the appeal of large organizations: because they can have a greater number of identical problem solvers, they are on average faster, and have lower variation in the solution time. These advantages are not costly as long as the scale of the organization matches the problem inflow, because the advantages come through the greater ability to absorb variability in the problem inflow. A slightly more complex implication is that the graphs suggest a major problem of scaling organizations. The processing time and its standard deviation both increase rapidly as the inflow increases. That's problematic when the load $\rho$ varies over time, as in many services that are time or date dependent. Most organizations cannot flexibly contract and expand to match $\rho$, leading to periods with extreme queuing because of high demand on their services.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

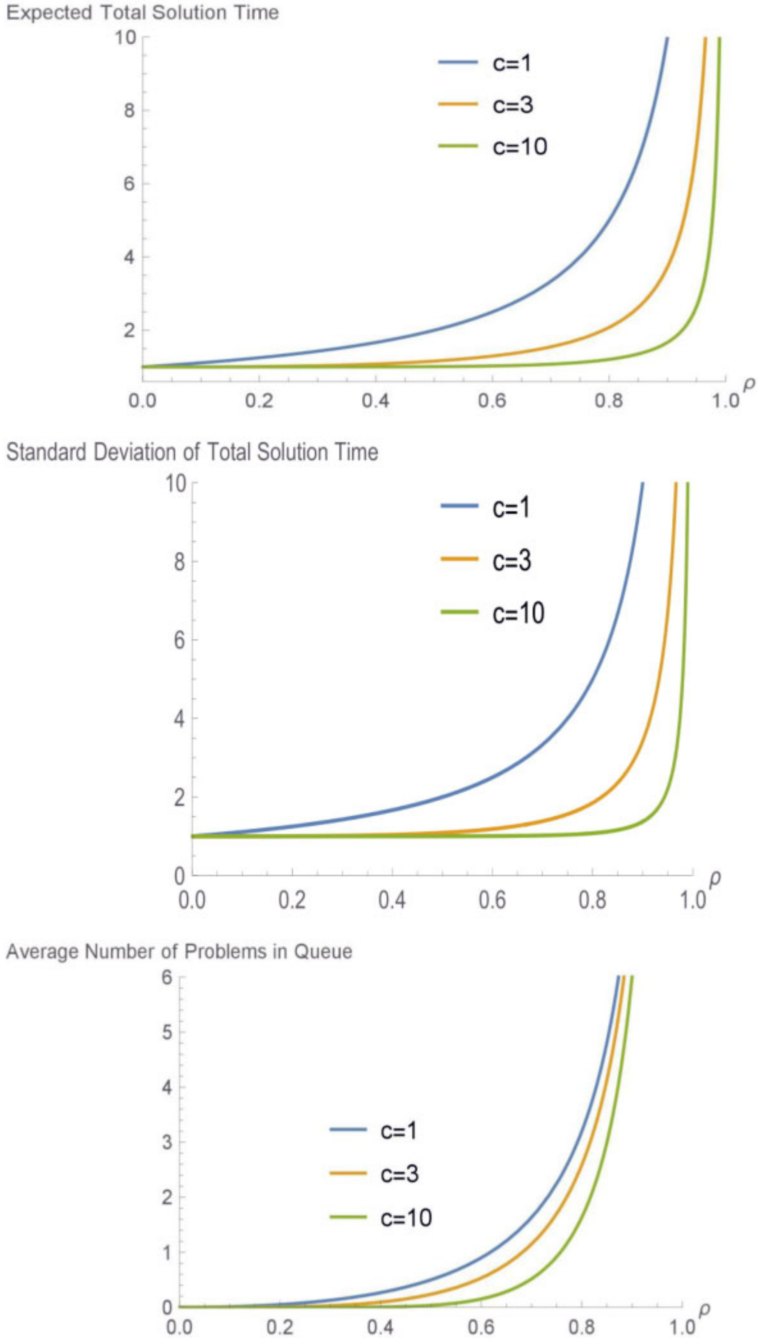


**Figure 2.** Problem solution times.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

### 1.1 Solution by oversight and flight

In the simplest of queues, all decisions are problem resolutions. However, an organization can also make decisions that solve no problem (oversight) and decisions that drop problems from consideration (flight) (Cohen *et al*., 1972: 8). Often choices involve both oversight and flight, as when one problem is dropped from consideration without any resolution, and the organization starts considering the next problem in the queue. This means that nothing happens to the problem, but the problem solver has decided to ignore it and move on, so there is now one less problem in the queue. Decision-making with oversight and flight was noted as an organizational phenomenon in the original formulation of the garbage can theory (Cohen *et al.*, 1972), and has also been observed in later research (e.g., Takahashi, 1997; Levitt and Nass, 1989). In this simple queue model, one way of solving by oversight (and flight) is if problems are solved by oversight when they have waited longer than an exponentially distributed amount of time having mean $1/\theta$. One motivation for this mechanism is that problem solvers dislike long-lived problems, perhaps because they are seen as overly difficult or because the organization has incentives connected to the number of problems solved. Once a problem is solved by oversight, its processing time is assumed to be zero.

1. The average number of problems in the queue is

$$\pi_0 \frac{\rho^c}{c!} \sum_{k=1}^{\infty} \frac{k\left(\rho/c\right)^k}{\prod_{j=1}^{k}\left(1+\frac{j\theta}{\mu c}\right)}$$

where

$$\pi_0 = \left( \sum_{k=0}^{c-1} \frac{\rho^k}{k!} + \frac{\rho^c}{c!} \sum_{k=0}^{\infty} \frac{\left(\rho/c\right)^k}{\prod_{j=0}^{k}\left(1+\frac{j\theta}{\mu c}\right)} \right)^{-1}$$

2. The average solution time for a problem, either by oversight or by resolution, is

$$E(S) = \frac{\pi_0}{\lambda} \frac{\rho^c}{c!} \sum_{k=1}^{\infty} \frac{k\left(\rho/c\right)^k}{\prod_{j=1}^{k}\left(1+\frac{j\theta}{\mu c}\right)} + p_s \frac{1}{\mu}$$

where $p_S$ is the probability solved by resolution and is given by

$$p_s = \pi_0 \sum_{k=0}^{c-1} \frac{\rho^k}{k!} + \frac{c}{\rho} \frac{\rho^c}{c!} \sum_{k=1}^{\infty} \frac{\left(\rho/c\right)^k}{\prod_{j=1}^{k}\left(1+\frac{j\theta}{\mu c}\right)} \pi_0$$

3. The standard deviation of the solution time for a problem is $\sqrt{E(S^2) - (E(S))^2}$, where

$$E(S^2) = \frac{2\pi_0}{c\lambda\mu} \frac{\rho^c}{c!} \sum_{k=1}^{\infty} \frac{\left(\rho/c\right)^k}{\prod_{j=1}^{k}\left(1+\frac{j\theta}{\mu c}\right)} \sum_{l=1}^{k} \frac{l}{1+\frac{l\theta}{c\mu}} + \frac{2}{\mu} \frac{\pi_0}{\lambda} \frac{\rho^c}{c!} \sum_{k=1}^{\infty} \frac{\left(\rho/c\right)^k}{\prod_{j=1}^{k}\left(1+\frac{j\theta}{\mu c}\right)} \sum_{l=1}^{\infty} \frac{1}{1+\frac{l\theta}{\mu c}} + p_s \frac{2}{\mu^2}$$

4. The proportion of decisions made by oversight and flight is $1 - p_S$.

Although these expressions are more complex than those of the basic model, they are easy to interpret by keeping in mind that $1/\theta$ is the mean time until oversight is applied, so a greater $\theta$ means that problems are dropped sooner. Inspection of the formulae for the mean and standard deviation of the solution time shows that increased $\theta$ reduces the solution time mean and standard deviations. Thus, organizations gain efficiency through solving problems through oversight. Naturally the cost of this efficiency is reduced effectiveness, if we define effectiveness as actual resolution of problems.

Again, the effects can be displayed through graphs. The graphs shown in Figures 3 and 4 continue to compare 1, 3, and 10 single problem solvers, but now they also show different speeds of dropping problems from consideration. Comparing the vertical axes of the graphs, we see that the top graph with the fastest dropping of problems has more rapid resolution of problems (by solution or oversight) than the medium and the bottom one, with the difference

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

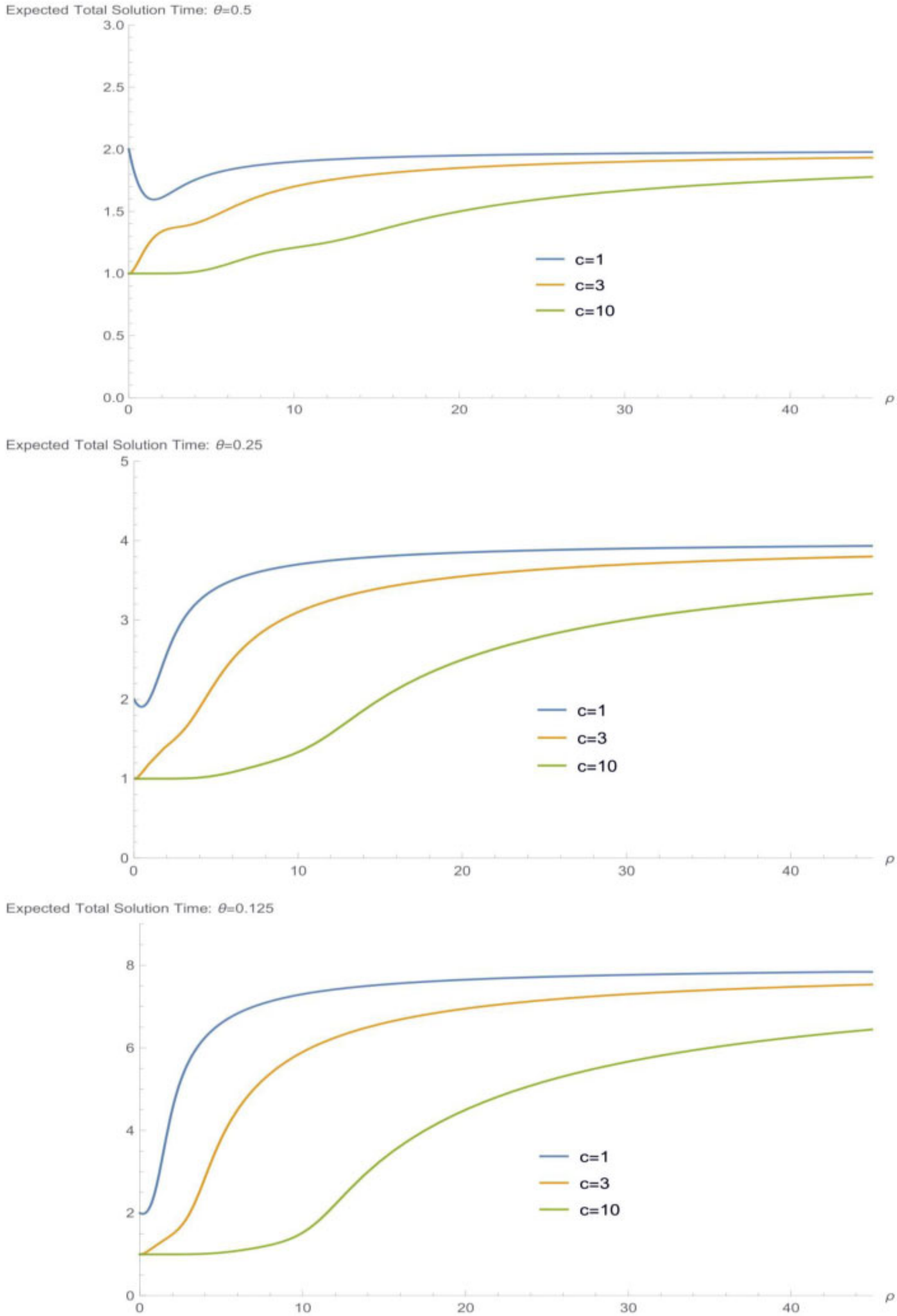


**Figure 3.** Solution and oversight: solution times.

being particularly clear to the right, when the problem load is high. Figure 4 suggests a role of increased standard deviation in generating this outcome, as it is clearly larger in the lower figures, especially when the load is high.

Again there are multiple substantive implications of these findings. Going back to the problem of organizations facing extreme queuing at specific times, having problem solvers with any kind of incentive connected to the average speed of resolution will create obvious risks for the organization. For example, the telephone operator who

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

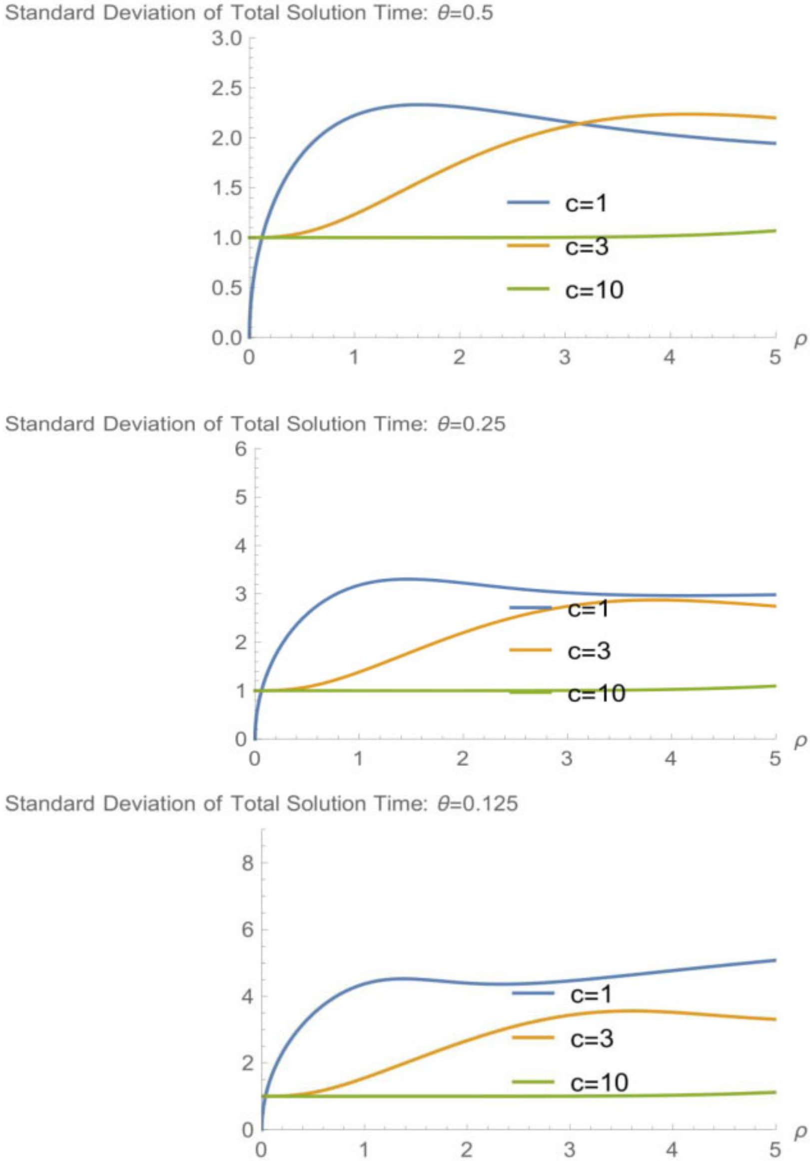


**Figure 4.** Solution and oversight: standard deviation.

understands that a request is time consuming might seemingly by accident disconnect the call in order to process more of the simpler requests. Such accidents are more likely to happen if phone operators have any kind of incentive system linked to the number of calls processed. More formally, organizations may have mechanisms for pre-sorting of requests and allocation of resources that in practice means that the more time-consuming requests will systematically end in oversight during periods of queuing.

However, a model of problems aging may not be the most realistic, as organizations have significant memory when problems are continuously reinforced by either external or internal actors. For example, social movements, investors, or the media can continue to pressure organizations to address a given problem. An alternative model assumes limited capacity for having problems in the queue, such that problems are dropped if the number waiting exceeds $m$.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

1. The average number of problems in the queue is

$$\pi_0 \frac{\rho^c}{c!} \sum_{k=1}^{m} k \left({}^{\rho}/{}_{c}\right)^k$$

where

$$\pi_0 = \left( \sum_{k=0}^{c-1} \frac{\rho^k}{k!} + \frac{\rho^c}{c!} \sum_{k=0}^{m} \left({}^{\rho}/{}_{c}\right)^k \right)^{-1}$$

2. The average solution time which includes dropped problems (having solution time zero) is

$$E(S) = \frac{1}{\mu}\left(1 - \pi_{\text{drop}}\right) + \pi_0 \frac{\rho^c}{c!} \sum_{j=0}^{m-1} \frac{j+1}{c\mu} \left(\frac{\rho}{c}\right)^j$$

where $\pi_{\text{drop}} = \pi_0 \frac{\rho^{c+m}}{c!c^m}$.

3. The standard deviation of the problem solution time is $\sqrt{E(S^2) - (E(S))^2}$, where

$$E(S^2) = \frac{2}{\mu^2} \pi_0 \sum_{j=0}^{c-1} \frac{\rho^j}{j!} + \pi_0 \frac{\rho^c}{c!} \sum_{j=0}^{m-1} \left(\frac{\rho}{c}\right)^j \left( \frac{j+1}{c^2\mu^2} + \frac{1}{\mu^2} + \left(\frac{j+1}{c\mu} + \frac{1}{\mu}\right)^2 \right)$$

4. The proportion of decisions made by oversight and flight is $\pi_{\text{drop}}$.

Again we can easily see the effect of oversight by concentrating on its parameter $m$, the maximal number of problems in the queue. We see that the summation over $m$ means that the mean and standard deviation time both increase as $m$ increases. Again, making decisions by oversight increases organizational efficiency and reduces organizational effectiveness.

Note that a key feature of the simulations done here is that the organization will be busy solving problems; it will be in a high-load condition. We set the parameters accordingly. This is done partly for realism, because organizations seek to eliminate spare capacity (slack), or conversely, ignore fewer problems if they have spare capacity, and partly because the comparisons between different conditions are more meaningful when the load is high.

## 2. Organizational structures

After comparing the decision-making structures, we examine some organizational structures through models that incorporate simple versions of how organizations solve problems. Our approach is to simplify the procedures sufficiently that we can illustrate clearly the main feature of the structure and how it affects problem solving, without introducing additional details that make the source of the effects unclear. We focus on two basic structure. First, organizations have a hierarchy with managers approving solutions that problem solvers in the level(s) below propose. Second, organizations have teams of problem solvers who bring diverse expertise together to solve a single problem.

### 2.1 Managers

Problem solving is often delegated to subordinates or experts, but the final solution is approved by supervisors. These can be anything from a line supervisor inspecting work, to a manager approving a budget, or an executive signing off on a proposed strategic action. Indeed, the approval of certain classes of solutions by a manager defines the hierarchy, which is a basic feature of organizations. The so-called "classic" organization theory was preoccupied with the question of how best to subdivide organization and manage each unit (Gulick and Urwick, 1937; Simon *et al.*, 1950).

An important part of this theory was the idea that organizations have goal hierarchies, with the top-level goals of the organization as a whole gradually broken into subgoal further down the hierarchy (March and Simon, 1958).

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

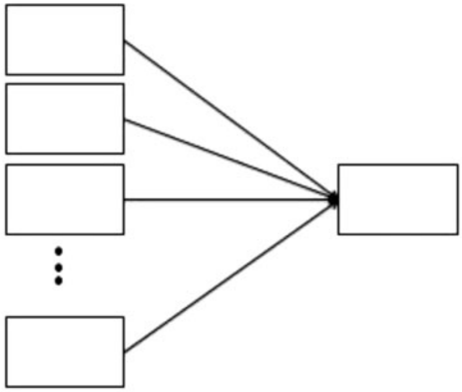


**Figure 5.** Problem solvers and manager.

Because each subgoal is incomplete, problem solvers in the lower part of the organization will not be fully informed of whether each solution is the best for the organization as a whole, so approval from a manager with a more complete understanding of the organizational goal is needed. Classic organizational theory saw subdivision as a means towards greater efficiency through better training and learning of each type of problem solver, but was also aware of the costs of adding hierarchical levels. Adding a manager to our model captures this concern and allows assessment of its consequences.

Again the total arrival rate to the system is $\lambda$, so each of the $c$ subordinates receives problems at rate $\lambda/c$. We assume that each subordinate takes an exponentially distributed time with mean $1/\mu$ and that the supervisor requires an exponentially distributed time with mean $1/\mu_1$. This parameter ensures that the manager has a capacity to approve that matches the subordinates' capacity to solve. An implicit assumption is that a manager will spend less time on each solution, the more subordinates there are (span of control) or the greater the expected problem-solving speed of each subordinate. We assume that $\lambda < c\mu$ and that $\lambda < \mu_1$, for otherwise the amount of work in the system grows without bound. Figure 5 illustrates this decision-making structure.

As before, we are interested in

1. The average number of problems in the queue (and not receiving service/approval) is

$$\frac{\rho^2/c}{1-{}^{\rho}/_{c}}+\frac{\rho_1^2}{1-\rho_1}$$

where $\rho_1 = \lambda/\mu_1$ (and we assume $\lambda < \mu_1$).

2. The average time required for problem solution is

$$\frac{1}{\lambda}\left(\frac{\rho}{1-{}^{\rho}/_{c}}+\frac{\rho_1}{1-\rho_1}\right)$$

3. The standard deviation of the time required for problem solution is

$$\sqrt{\frac{1}{\mu^2}\frac{1}{\left(1-{}^{\rho}/_{c}\right)^2}+\frac{1}{\mu_1^2}\frac{1}{(1-\rho_1)^2}}$$

The effect of the manager is easy to interpret by comparing with the basic M/M/c model of a queue without a manager. We see that the mean completion time increases (the extra term in the expression). The standard deviation of the time required for problem solution also has two terms, and we note that the division by the number of problem solvers in the basic M/M/c standard deviation ($1/c$) no longer occurs. The reason is the manager connects the queues because one

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

manager now approves solutions from multiple problem solvers, causing congestion and leading to increased standard deviation. Thus, managerial approval reduces the stability of organizational problem solving times.

There are straightforward substantive implications. First, if the organization is under high load $\lambda$, adding an approval at a higher level of hierarchy is potentially costly. Because the reason for the approval stage is to ensure that relatively uninformed workers make decisions that match the overall goals of the organization, the model suggests the value of the alternative path of making the workers better informed and avoiding the approval step. Naturally, this could involve an additional and costly problem of ensuring that the incentives for fulfilling organizational goals are aligned well, but this problem can also occur at the approval stage. A second implication has to do with the advantage of large organizational size showed in the initial model. This advantage relies on the problem solving queues being independent from each other, but the consequence of an approval step is to connect the queues, and hence cause the larger organization to lose its advantage over a small organization.

## 2.2 Teams

Organizations often require a team of multiple members to take part in problem solving, as in task forces and committees. The properties of team composition and team work processes has been the subject of significant work over time, with findings that now show differences according to the type of team and the task solved by the team (Jones and George, 1998; Bantel and Jackson, 1989; Edmondson, 1999; Horwitz and Horwitz, 2007), and has become a specialization in the field. We step back from the questions regarding internal team structures and processes, and instead we view the team as a device for seeking to use the horizontal dimension of the organization better. Organizational members are different from each other, unlike the identical problem solvers in the M/M/c model, and these differences include skill sets that vary across individuals, even if they hold identical jobs. If the organization is presented with problems that require unequal mixtures of skills to solve, then the regular assignment of problems to individuals done in the M/M/c model may slow problem solving because each problem solver slows down when facing problems that require a skill set that differs from the one he or she holds. To show the potential effects of teams, we simplify the problem by not assuming any such effect of slowing down when facing problems that are difficult for the individual, but we do assume that a team divides any problem into subtasks that are distributed to each team member.

A simple formulation of a team is one in which member $i$ of the committee individually works on one of $c$ problem subtasks (each having an exponential completion time with mean $1/(c\mu)$). Upon completing the subtask, the team member hands the problem on to member $i+1$, who then commences the next subtask. The $c$ subtasks are completed serially, after which the manager works on the problem for an exponentially distributed amount of time having mean $1/\mu_1$, thereby completing the problem solution. The logic behind this formulation is the same as the logic behind teams. Each team member has distinct expertise that needs to be brought to the problem in order to ensure a solution, and the combination of this matching of expertise and problem characteristics can be modeled as a serial solution. In reality, the solution may involve passing the problem back and forth in multiple rounds, but this structure can be simplified into a single series without losing its basic characteristics. Actual teams in organizations have meetings in which all are present, but the conversations in these meetings usually consist of serial, not simultaneous, statements. It would be possible to envision teams in which there is some sort of parallel processing of the problem, but that would be equivalent to the basic M/M/c model with each problem solver consisting of multiple individuals, so it is a simpler structure than the sequential team processing we model here. Thus, we model the team structure as shown in Figure 6.

As before, we are interested in

1. The average number of problems are queued is

$$\frac{\rho^2/_c}{1-\rho/_c}+\frac{\rho_1}{1-\rho_1}$$

2. The average time required for problem solution is

$$\frac{1}{\lambda}\left(\frac{\rho}{1-\rho/_c}+\frac{\rho_1}{1-\rho_1}\right)$$

3. The standard deviation of the time required for problem solution is

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

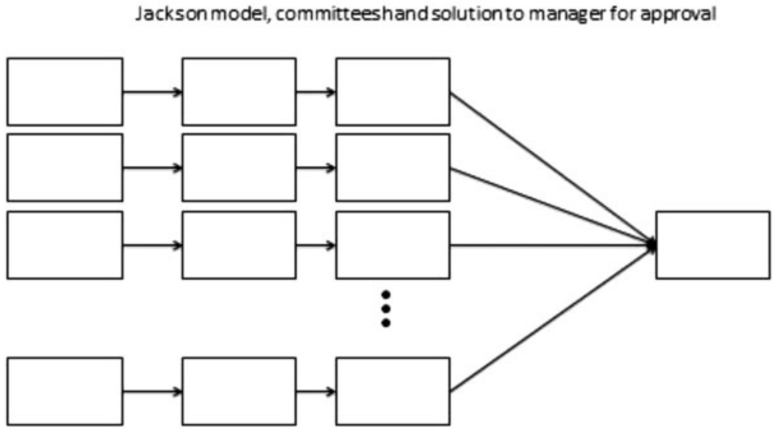


**Figure 6.** Team and manager.

$$\sqrt{\frac{1}{c\mu^2}\frac{1}{\left(1-{}^{\rho}/_{c}\right)^2}+\frac{1}{\mu_1^2}\frac{1}{(1-\rho_1)^2}}$$

The team is also easy to compare to the basic M/M/c model and the model with managerial approval. Note that we have standardized the committee member expected time so that the team member expected times add up to the time of a single problem solver. (This is counter to the theoretical idea that team members work faster because they only use their most expert knowledge. It is also counter to the heuristic that teams are intrinsically slow. However, it gives a cleaner comparison to the nonteam organization than any of these ideas.) Such a serial queue has a lower standard deviation of the time of completion than a single problem solver because each team member independently draws the stochastic completion time, thus reducing the effect of outlier values of the time. Hence, in team decision-making followed by managerial approval, it is the manager, not the team, who introduces variability into the problem solution time. The team on its own has lower standard deviation than a single problem solver with the same mean time. Thus, teams stabilize problem solving times.

A major substantive implication is the contrast of these findings with popular belief. Why is it still common to believe that teams are slow? There are multiple possible reasons for this belief. The first is that people do not consider the counterfactual. Problems assigned to teams are usually more difficult than problems that are assigned to individuals. Naturally they are solved more slowly, because difficult problems take longer time than easy problems. The second and more valid explanation is that the handoffs between team members involve communication, which is time consuming and not always accurate. Thus, a more realistic model of team problem solving might add a communication delay between each step of the sequence. Such a model is less interesting than the simple model we show here because it is trivially true that communication delays will slow down a queue, and can make it slower than a one-step queue with the same expected problem-solving time. Given that this communications problem is solved well, these results show that the organizational process of using teams to handle problems that require some specialization is not only a good use of individual skill sets, it also has an added bonus of absorbing variability in the solving process and hence give more stable solution time.

## 3. Deviations from organizational structures

It is also useful to make comments on how organization in practice vary their procedures, or break procedures, for the sake of expediency. It is well-known that organizational members deviate from the prescribed procedures, often because they find the procedures cumbersome and unnecessary, or simply because the problem they are solving is seen as urgent. The most classic deviation is to solve a problem that should have gained a manager's approval, but without waiting for the approval. This is often easy to do because the problem solvers see all problems that arrive to them, but the manager only sees the problems forwarded for approval. Also, many problem solvers in organizations

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

work on a mix of problems with solutions that need approval and problems with solutions that do not need approval.

We consider here the model in which one has $c$ problem solvers working in parallel, as in the first model of this section. We modify the model as follows: Each problem solver can choose to bypass the manager with probability $1-p$. This reduces the load on the manager by a factor of $1-p$. To adjust the managerial capacity so that it is identical to the earlier model, we now assume that the manager's processing time is exponentially distributed with mean $1/(\mu_1 p)$.

1. The average number of problems that are queued is

$$\frac{\rho^2/_c}{1-\rho/_c}+\frac{(\rho_1/_p)^2}{1-\rho_1/_p}$$

where we assume that $\lambda < p\mu_1$.

2. The average time required for problem solution is

$$\frac{1}{\lambda}\left(\frac{\rho}{1-\rho/_c}+\frac{\rho_1/_p}{1-\rho_1/_p}\right)$$

3. The standard deviation of the time required for problem solution is $\sqrt{E(S^2)-(E(S))^2}$, where

$$E(S^2)=\frac{2(1-p)}{\mu^2(1-\rho/_c)^2}+2p\left(\frac{1}{\mu^2(1-\rho/_c)^2}+\frac{1}{p^2\mu_1^2(1-\rho_1/_p)}\right)$$

The consequences of bypassing the manager are easy to see from these formulae. The probability $p$ that a problem solver decides to forward the solution to the manager for approval occurs only in places in which higher values of $p$ increases the value, so the mean and standard deviation of the problem solution time are increased by forwarding more problems to the manager for approval, and decreased by bypassing the manager for more problems. Breaking the rule of managerial approval and instead bypassing the manager increases organizational efficiency and stability of problem solution times.

The obvious substantive implication is that this model results suggests a positive role of the organizational member who makes tradeoffs between following rules and having timely solutions to problems. This is well recognized by organizations that have procedures allowing lower level employees to skip approval steps that would normally be required when timeliness is needed, as such rules are just ways of granting permission to a more efficient decision-making when the situation calls for it. A natural concern with breaking rules or having rule-breaking rules is that the unapproved solutions may not fit the organizational goal well, and the organizational member could strategically use this procedure as a result of having personal interests contrary to the organizational goals. As with many organizational procedures, efficiency comes at a cost.

## 4. Complex organizations

The organizational structures and procedures we have considered so far have started with one problem solver as the basic unit and extended the structure one step either vertically or horizontally. This cover the most basic principles of organizing, but is far simpler than the complex organizational structures that we see in operation. Taking the next step towards studying organizational structures involves a dilemma, however, because the complexity has to be simplified in order to gain rigor. We take Cohen *et al.* (1972) as an indication of one potentially fruitful approach, and like them we examine hierarchical decisions and specialized decisions. Here, hierarchy does not mean approval, as in the earlier section on the effect of a manager, but instead that problems that exceed a certain level of difficulty or importance are allocated to problem solvers higher up in the hierarchy. Specialized decisions likewise mean that problems differ in characteristics so that they are unsolvable for some problem solvers but not others, but here we envision that all problem solvers are at the same rank, but hold different knowledge.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

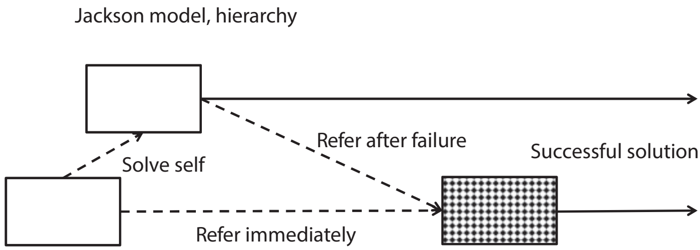


**Figure 7**. Jackson model, hierarchy.

## 4.1 Hierarchy

Figure 7 shows a simplified version of the hierarchical structure. Problems arrive to the lower level problem solver, who diagnoses each problem as either a low-level problem that the problem solver can solve or as a high-level problem that needs to be referred to the next level. The high-level problem solver is shown by the filled marking, the low-level problem solver is shown both in the diagnosis role and the solving role, and the arrows are either dashed to indicate referral or solid to indicate successful solution. As the figure shows, there is some probability that the lower-level problem solver tries to solve a high-level problem and fails, and needs to refer to the high-level problem solver after failing. This event was not studied by Cohen *et al.* (1972), but is a useful addition to the model.

More generally, instead of one problem solver at each level we have a pool of low-level problem solvers acting and referring to a pool of high-level problem solvers, giving waiting times and solving times at both levels of the solution hierarchy. We let the problems vary in difficulty as a draw from the uniform distribution with value $x$, and let $f(x)$ be the likelihood that it can be solved successfully at the lower level. The low-level problem solvers refer all problems with difficulty greater than $\bar{x}$ , and the high level problem solvers solve all problems regardless of whether they are referred directly or referred after a failed solution attempt. This setup, with some added details, has been examined by Hasija *et al.* (2005), who found that an analytical solution was not possible. However, for organizations under high inflows of problems relative to their capacity, precise approximations are possible and yield interesting findings.

Waiting times and processing times are both affected by the choice of the ratio of low level problem solvers to high level solvers, and this in turn depends on how the skills of each level match with the inflow of problems. One extreme case is that low level problem solvers that rarely solve problems, but instead refer most to the next level and solve only those with a very low difficulty $\bar{x}$. Another extreme case is low-level problem solver who try to solve a significant proportion of the problems presented to them, referring up either when $x$ is clearly very high or they have failed to solve the problem.

Both types of hierarchies are known in organizations. The hierarchy with a dominant high level is typically used when keeping waiting times low is valuable, such as when servicing problems in large assets such as ships, trains, or land transportation facilities like roads and bridges. The hierarchy with active and numerous low level problem solvers is typically used when the waiting time of problems is less costly for the organization (though it may be costly for the client), as in service establishments such as medical treatment or banking. The economic value of using a large low tier lies in the lower pay of each problem solver at this level, but to use this structure efficiently the lower level problem solvers need to be trained to be just good enough to attempt solving a high proportion of the problems presented, and preferably also to correctly refer those that are untreatable at this level. Interestingly, this hierarchical structure can usually be staffed efficiently simply by using the old heuristic that each layer of problem solvers should have a number of problem solver equal to the square root of those at the layer below.

## 4.2 Specialization

Figure 8 shows an example of the specialized structure. In this structure, there are three specializations with the divisions between them shown by lines, and the two problem solvers in each specialization can only solve their designated type of problem. We assume that the problems self-sort into the correct specialization, or alternatively that there is a gatekeeper sorting the problems with no waiting time. Between the second and third specialization we have also placed a problem solver who can solve either of these problem types, and marked this problem solver with

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

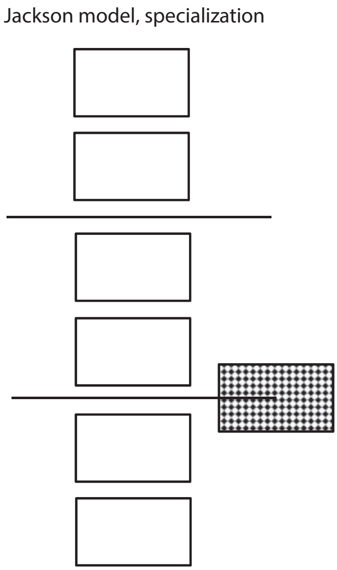


**Figure 8.** Jackson model, specialization.

shading to indicate that it is a more advanced and expensive problem solver than those who can only solve one type of problem.

Without doing any modeling, we can make some observations about the effect of the specialized structure. First, note that each specialization is an M/M/c structure, so if we had no problem solvers who can address two types the overall organization would simply be the sum of $N$ (three in this case) M/M/c structures, which makes the expectation and standard deviation of the waiting times easy to calculate. The most important comparison is between the single queue with $c$ problem solvers and $n$ queues with $c/n$ problem solvers. The formulae for the M/M/c queue indicate increase in average and standard deviation of the solution time as $c$ is reduced, even if $\lambda$ is adjusted downward accordingly, so any division by specialization will make the solution times longer and less predictable. Given this disadvantage of specialization, the question is whether flexible problem solvers such as the one placed between the second and third specialization in Figure 8 can help.

This question has been studied by Bassamboo *et al.* (2012), who explored what minimal degree of flexibility was needed to make the specialized organizational structure efficient. This problem is also too difficult for an analytical solution, but they found precise approximations provided the organization has high inflows of problems relative to their capacity for solving them. The first question is whether the generalist (nonspecialized) problem solvers need to be able to help more than two specializations, or whether it was sufficient that each specialist can help only two. The answer depends on the queue structures, but it is nearly always enough that the generalists can solve exactly two kinds of specialized problems. Next, assuming that each specialty has the same inflow of problems ($\lambda$) and capacity for solving it, the next question is how many generalists are needed. Assuming that waiting time costs and hiring costs are linear, but not assuming any prioritization of waiting time relative to hiring cost, a clear and simple answer appeared: it is most efficient to have a number of two specialization generalists equal to the square root of the number of specialists (Bassamboo *et al.*, 2012). As in the previous section, an old heuristic on the proportions of different types of staff turns out to work quite well.

For organizations, the conclusions from this modeling exercise may come as a surprise. Everyday experience as consumers and clients suggests that many organizations use specialization without sufficient awareness of how it increases waiting times. Hospitals assign medical doctors to specialties even they are trained to treat multiples—and indeed, the same doctors who are normally specialists can act as generalists when assigned to emergency room duty. In some cases, the efficiency gains from generalists are also easily observed. Immigration lines are usually specialized by the type of passport (foreign, domestic, or even groups of nations like EU and ASEAN). Whenever a line is allowed to change specialization in order to serve a neighboring specialty that is queued up, the waiting time is

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

reduced. Moving from the flexible problem solving assumed by queuing models to more regular production, we note that most automobile makers design factories and logistics to produce one model or a few closely related models on the same production line. However, Toyota is famous for its flexible production lines that can mix multiple automobile models either at once or by quickly changing from one to the other. Contrary to common belief, and consistent with the findings of Bassamboo *et al.* (2012), most of Toyota's manufacturing is not done in flexible production lines, but in conventional production lines. These are paired up with smaller flexible production lines that can change models at low cost or reduce speed at low cost, whereas conventional production lines work best when they have sufficient incoming orders to operate at their designed maximal speed.

## 5. Discussion and conclusions

Our modest goal in this article was to revisit the garbage can model, and reline it with the idea that problems are placed in queues and stochastically matched with people who look for solutions. Our approach demonstrates that managerial approval increased the standard deviation of problem resolution and indicates that queues are processed faster and have lower variance when there is oversight or teamwork, or when manager approval is bypassed due to independent action by problem solvers. It shows clearly the tradeoffs involved in making organizational structures and procedures. Oversight and flight from problems is efficient. Consistency of goals and maintenance of quality is inefficient. Teams are more reliable than individuals in their problem solving activities, so any problems seen in teamwork must be related to communications. Avoiding oversight from superiors is efficient for the organization too, not just the individual problem solver. Hierarchies of problem difficulty can be solved at modest cost, and so can mixtures of problems that different types of specialists are qualified to address. Some of these solutions appear to differ from common organizational practice, though they are also known from specific organizations. Some of these tradeoffs are already accepted in the literature, but we have shown them formally and quantified their cost.

Our article is best seen as an invitation to re-engage with the garbage can model and address a number of unresolved issues. We expect further work to proceed on decision-making structures such as the unsegmented, hierarchical and specialized structures. The findings we have presented so far require a series of specific assumptions about similarity of problem load across hierarchy levels and specializations, and the conclusions may look different if these are relaxed. Indeed, the rules for referral of problems in the hierarchical structure or generalization in the specialized structures are simple and can be made more elaborate. For example, future work can examine a variation of the specialized structure in which idle solvers may (with probability p) try to solve a problem outside their topic, but do so with much longer solution duration. This can be thought of as specialization with mistakes, or alternatively, specialization with helping across specialties. In modeling terms, mistakes and helping give equivalent results. While this approach is tractable in a regime with static assignment of tasks, it is harder to derive closed form solutions when assignment is dynamic.

We hope these directions inspire researchers to reflect on garbage can processes, and to illuminate how macro-outcomes in problem resolution and decision-making are the outcomes of stochastic matching across queues of problems, solutions, and people. We think the work in this article has provided a framework for future research that can build on the seminal work of Cohen *et al.* (1972) and extend it in multiple directions.

## Acknowledgments

We are grateful to Sameer Hasija, Thorbjørn Knudsen, and Phanish Puranam for helpful comments on the article.

## References

Adan, I. and J. Resing (2015), *Queueing Systems*. Eindhoven University of Technology, Department of Mathematics and Computing Science: Eindhoven, The Netherlands.

Bantel, K. A. and S. E. Jackson (1989), 'Top management and innovations in banking: does the composition of the top team make a difference?' *Strategic Management Journal*, **10**(S1), 107–124.

Bassamboo, A., R. S. Randhawa and J. A. Van Mieghem (2012), 'A little flexibility is all you need: on the asymptotic value of flexible capacity in parallel queuing systems,' *Operations Research*, **60**(6), 1423–1435.

Bendor, J., T. Moe, and K. W. Shotts (2001), 'Recycling the garbage can: an assessment of the research program,' *American Political Science Review*, **95**(1), 169–190.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

Cohen, M. D., J. G. March and J. P. Olsen (2012), 'A garbage can model at forty: a solution that still attracts problems,' in A. Lomi and J. R. Harrison (eds), *The Garbage Can Model of Organizational Choice: Looking Forward at Forty*. Emerald Group Publishing Limited: Bingham, UK, pp. 19–30.

Cohen, M. D., J. G. March, J. P. Olsen et al. (1972), 'A garbage can model of organizational choice,' *Administrative Science Quarterly*, **17**(1), 1–25.

Edmondson, A. (1999), 'Psychological safety and learning behavior in work teams,' *Administrative Science Quarterly*, **44**(2), 350–383.

Gulick, L. and L. Urwick (1937), *Papers on the Science of Administration.* Institute of Public Administration, Columbia University: New York.

Hasija, S., E. J. Pinker and R. A. Shumsky (2005), 'Staffing and routing in a two-tier call centre,' *International Journal of Operational Research*, **1**(1/2), 8–29.

Horwitz, S. K. and I. B. Horwitz (2007), 'The effects of team diversity on team outcomes: a meta-analytic review of team demography,' *Journal of Management*, **33**(6), 987–1015.

Jones, B. D. and F. R. Baumgartner (2005), *The Politics of Attention: How Government Prioritizes Problems*. University of Chicago Press: Chicago, IL.

Jones, G. R. and J. M. George (1998), 'The experience and evolution of trust: implications for cooperation and teamwork,' *Academy of Management Review*, **23**(3), 531–546.

Kingdon, J. W. (1984), *Agendas, Alternatives, and Public Policies*. Little, Brown: Boston, MA.

Knudsen, T., M. Warglien and S. Yi (2012), *'Garbage Can in the Lab,' the Garbage Can Model of Organizational Choice: Looking Forward at Forty*. Emerald Group Publishing Limited: Bingham, UK, pp. 189–227.

Levitt, B. and C. Nass (1989), 'The lid on the garbage can: institutional constraints on decision making in the technical core of college-text publishers,' *Administrative Science Quarterly*, **34**(2), 190–207.

Lomi, A., G. Conaldi and M. Tonellato (2012), 'Organized anarchies and the network dynamics of decision opportunities in an open source software project,' in A. Lomi and J. R. Harrison (eds), *The Garbage Can Model of Organizational Choice: Looking Forward at Forty*. Emerald Group Publishing Limited: Bingham, UK, pp. 363–397.

Lomi, A. and J. R. Harrison (2012), *The Garbage Can Model of Organizational Choice: Looking Forward at Forty.* Emerald Group Publishing Limited: Bingham, UK.

March, J. G. and J. P. Olsen (1979), *Ambiguity and Choice in Organizations*. Universitetsforlaget: Bergen, Norway.

March, J. G. and H. A. Simon (1958), *Organizations.* John Wiley: New York, NY.

Padgett, J. F. (2013), 'The garbage can model of organizational choice: looking forward at forty,' in A. Lomi and J. R. Harrison (eds), *Research in the Sociology of Organizations*, Vol. 36. Sage Publications: Los Angeles, CA, pp. 472–482.

Simon, H. A. (1947), *Administrative Behavior.* Macmillan: New York, NY.

Simon, H. A., V. A. Thompson and D. W. Smithburg (1950), *Public Administration.* Transaction Publishers: Piscataway, NJ.

Takagi, H. (2014), 'Waiting time in the m/m/$m/(m+c)$ queue with impatient customers,' *International Journal of Pure and Applied Mathematics*, **90**(4), 519.

Takahashi, N. (1997), 'A single garbage can model and the degree of anarchy in Japanese firms,' *Human Relations*, **50**(1), 91–108.

Walrand, J. and P. Varaiya (1980), 'Sojourn times and the overtaking condition in Jacksonian networks,' *Advances in Applied Probability*, **12**(4), 1000–1018.

## Appendix: Key to solution sources

### M/M/c queue

- Formula for $\pi_0$: Adan and Resing (2015: 44)
- Formula for average number-in-queue: Adan and Resing (2015: 44)
- Average solution time: Adan and Resing (2015: 44)
- Standard deviation of solution time: Adan and Resing (2015: 46), obtained by integrating $P(S > t)$

*Resolution by oversight* (i.e., abandonment after waiting an exponentially distributed amount of time having mean $1/\theta$)

- Formula for $\pi_\theta$: Takagi (2014: 523)
- Formula for average number-in-queue: Takagi (2014: 525)
- Formula for average solution time:
  a. Formula for $p_S$ (=probability that a problem is served): Takagi (2014: 529)
  b. Formula for $E$(wait, abandon): Takagi (2014: 551)

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020

c. Formula for $E$(wait, service): Takagi (2014: 545)

$$E(S) = E(\text{wait, abandon}) + E(\text{wait, service}) + p_s \frac{1}{\mu}$$

or just use

$$E(S) = E(\text{wait}) + p_s \frac{1}{\mu}$$

- Formula for $E$(wait): Takagi (2014: 540)
- Formula for standard deviation of solution time:

$$E(S^2) = E(\text{waiting time})^2 + \frac{2}{\mu} E(\text{waiting time, service}) + p_s \frac{2}{\mu^2}$$

- Formula for $E(\text{waiting time})^2$: Takagi (2014: 541)

## *M/M/c/c + m* queue

1 and 4: standard theory for birth–death processes

2: A problem that arrives to find $c+m$ problems in the system is dropped, so solution time $= 0$. A problem that arrives to find $i$ problems in the system ($0 \leq i < c$) has mean solution time $= \frac{1}{\mu}$. A problem that arrives to find $i$ problems in the system ($c \leq i < c+m$) has mean solution time

$$\frac{i-c+1}{c\mu} + \frac{1}{\mu}$$

So,

$$E(S) = \frac{1}{\mu}(1 - \pi_{\text{drop}}) + \pi_0 \frac{\rho^c}{c!} \sum_{j=0}^{m-1} \frac{j+1}{c\mu} \left(\frac{\rho}{c}\right)^j$$

where

$$\pi_{\text{drop}} = \pi_0 \frac{\rho^{c+m}}{c! c^m}$$

3: To compute $E(S^2)$, use PASTA and the fact that the wait time in state $c+j$ is $(c+1)$ iid exponential rv's having mean $1/(c\mu)$.

## *For the Jackson model with a problem solver in parallel, plus a manager:*

1. We use standard Jackson network theory and note that the number-in-queue in an $M/M/1$ queue is $\rho^2/(1-\rho)$.
2. We compute the total number-in-system and use Little's law.
3. We use the paper by Walrand and Varaiya (1980). Since there is no overtaking in this network, sojourn times at successive stations visited along a path are independent.

## Committee with serial subtask handling

Solved as for previous problem.

Downloaded from https://academic.oup.com/icc/article-abstract/29/1/125/5673614 by Stanford University Medical Center user on 17 April 2020