

The Goal

A Process of Ongoing Improvement

Eliyahu M. Goldratt and Jeff Cox. North River Press, 1984. Fourth Revised, 30th Anniversary Edition, 2014.

<https://www.amazon.com/dp/O884271951>

Book Overview

The Goal is a business novel that teaches the Theory of Constraints through the story of Alex Rogo, a thirty-eight-year-old plant manager at UniCo's Bearington plant. Alex has three months to turn the unprofitable and unreliable plant around before the corporate division vice president, Bill Peach, closes it. His guide is Jonah, a physicist and former professor whom Alex encounters at an airport and consults by phone and in occasional visits over the following months. The novel unfolds across forty numbered chapters. It runs a parallel domestic plot line as Alex's marriage strains under the plant crisis, and uses the marriage as the second setting for the same Socratic method Jonah applies at the plant.

Goldratt's operating claim is that every system has a small number of constraints, usually one, and that the throughput of the whole system is limited by that one constraint. The classic management response, driving local efficiencies at every station and station-level cost measures across the plant, produces work-in-process inventory and no additional throughput. The alternative is to identify the constraint, exploit it, subordinate everything else to it, elevate it, and then repeat the cycle. The scheduling application is called drum-buffer-rope. The plant's constraint is the drum, buffer inventory protects the constraint from starvation, and a rope of material release ties the plant's input rate to the constraint's output rate. The book teaches these ideas through Alex's team's discovery of two bottlenecks at the Bearington plant, the NCX-10 machining center and the heat-treat oven, and their eight-month campaign to lift throughput. The 30th Anniversary Edition adds a codified statement of the Five Focusing Steps and an essay titled Standing on the Shoulders of Giants.

Every system is limited by a constraint. Identify it, exploit it, subordinate everything else to it, elevate it, and then find the next one. Local efficiencies at non-constraints produce inventory, not throughput. The whole system's output moves only when the constraint moves.

Chapter Summaries

Chapter 1: The Crisis

The first five chapters set the crisis. Alex arrives at the Bearington plant on a Monday morning to find Bill Peach's Mercedes in his parking spot and an unresolved customer emergency on his desk. Peach tells Alex the plant has three months to become profitable or it will be closed and Alex will be reassigned. Alex flies to a corporate meeting that afternoon and encounters Jonah in the terminal. Jonah asks Alex what the goal of his company is. Alex answers with efficiencies, cost reductions, quality, and shipments. Jonah says none of those are the goal. The goal is to make money.

- The precondition for improving a system is naming what the system is for. Goldratt argues that most operating dysfunctions come from teams optimizing measures that are not the goal, and treating the local measure as though it were.
- The goal of a manufacturing company, in Goldratt's frame, is to make money now and in the future. The three operational measures that track that goal are throughput, inventory, and operational expense.
- Throughput is the rate at which the system generates money through sales. Inventory is the money currently tied up in the system. Operational expense is the money spent to turn inventory into throughput. Every operating decision moves at least one of these measures.
- The chapter's implicit claim is that most cost-accounting measures obscure the goal by encouraging local optimization. Alex's plant hits its efficiency numbers and loses money.

Chapter 6: The Wrong Measures

Chapters six through ten are the Socratic sequence in which Jonah rejects the cost-accounting measures Alex has been trained on and replaces them with the three operational measures. The plant is running high machine efficiencies, low direct-labor cost per part, and full loads on expensive equipment, and it is losing money. Jonah argues that the plant is running the wrong measures. Machine efficiency and cost per part are local measures. Throughput, inventory, and operational expense are system measures.

- A plant that maximizes local efficiency will pile up inventory at every station that runs faster than the downstream station can consume it. The inventory is money tied up in work that has not yet become throughput.
- Efficient use of a non-constraint resource is not efficient use of the system. The resource works. The system waits. Goldratt calls the equation activated is not the same as utilized.
- The fastest way to raise return on investment is to raise throughput while holding inventory and operational expense constant. Cost-cutting alone reaches a floor. Inventory reductions release cash and shorten cycle time. Throughput has no ceiling.
- Alex's controller Lou, initially the most resistant character to the new measures, becomes the character who translates them into the plant's monthly reports. The translation is a necessary step. The measures are not usable if the finance function cannot report them.

Chapter 11: The Boy Scout Hike

Chapters eleven through fifteen are the middle turn. Alex takes his son Dave and Dave's Boy Scout troop on a fifteen-mile weekend hike, and while walking with the line of scouts he discovers a physical model of his plant. The slowest scout, Herbie, governs the pace of the whole line. The scouts ahead of Herbie stretch the line. The scouts behind Herbie bunch up. Alex realizes that the plant is a line of dependent events with statistical fluctuations, and that the whole plant is paced by its Herbie.

- The Herbie metaphor is Goldratt's most-used image. A system with dependent operations and variability in each is governed by its slowest step, not by the average step. Balanced plants where every station's capacity matches average demand cannot

deliver average throughput because the variability compounds.

- The scouts ahead of Herbie represent inventory built at non-constraints. The gap between scouts represents variation in each scout's walking speed. The bunched scouts behind Herbie represent the queue in front of the constraint.
- Alex's practical fix on the hike is to put Herbie at the front of the line and take some of Herbie's pack weight, redistributing it to the faster scouts behind. The operating translation is that the plant should reduce load on the constraint and release the constraint's capacity.
- Back at the plant Alex's team, Bob Donovan in production and Stacey in inventory and Ralph in data processing, begins hunting the plant's Herbie. They quickly identify two candidates, the NCX-10 machining center and the heat-treat oven, both of which run at capacity and have queues piling up in front of them.

Chapter 16: The Bottlenecks

Chapters sixteen through twenty are the bottleneck-management sequence. The NCX-10 is the plant's most efficient machine, cost-justified in the previous decade to replace three older machines that were retired. Its setup time is long, its run rate is fast, and it is the only machine that can perform the operations for the parts it handles. The heat-treat oven runs at capacity because expeditors push small urgent batches through, and because small parts that do not fill the oven waste heat cycles. Neither bottleneck can be replaced. Each has to be exploited.

- Rule one, an hour lost at a bottleneck is an hour lost for the entire system. Rule two, an hour saved at a non-bottleneck is a mirage. The two rules govern where improvement effort should go.
- Exploit the NCX-10. Dedicate two setup operators so the machine never waits for a changeover. Keep the machine running through breaks. Perform quality control before the NCX-10 rather than after, so the bottleneck never processes parts that will be scrapped. Bring back the retired older machines to offload volume the NCX-10 does not have to run.
- Exploit the heat-treat oven. Combine batches that share a temperature. Perform quality inspection before the oven so bad parts do not consume oven time. Sequence part sizes so the oven's volume is used.

- The eighteen-month program Alex's team had been running to install the NCX-10 was cost-justified on efficiencies. It made the plant less able to produce, because it concentrated variability on a single station. Goldratt's implicit warning is that the capital proposals with the strongest ROI on paper often move the constraint to a place the plant cannot recover from.

Chapter 21: Subordinating The System

Chapters twenty-one through twenty-five are the subordination sequence. Once the bottlenecks are identified and being exploited, the rest of the plant has to be subordinated to them. Alex's team introduces a color-tag system, red tags for parts that go through a bottleneck and green tags for parts that do not. Red tags jump queues everywhere. Green tags wait. Material release is throttled so the plant does not pile inventory in front of the NCX-10 and the heat-treat oven.

- Subordination is the step that feels wrong and matters most. A non-bottleneck should run at the pace of the bottleneck, not at its own maximum. Underutilization of a non-bottleneck is the correct operating state.
- The color-tag system is a practical form of drum-buffer-rope. The bottleneck is the drum. Its output rate paces the plant. Buffer inventory ahead of the drum protects against starvation. The rope is the signal from the drum back to material release, which triggers new material only when the drum can consume it.
- Batch sizes should be smaller than traditional cost-accounting rules would set them. Smaller batches reduce work-in-process, shorten flow time, and free capacity at non-bottlenecks. The cost-accounting objection is that setup time rises. The Goldratt response is that setup time at a non-bottleneck is free.
- Traditional operations research emphasizes balanced-line design, in which each station has capacity approximately equal to average demand. Goldratt argues this is exactly the wrong design, because the statistical fluctuations then compound and no station has protective capacity. Excess capacity at non-bottlenecks is a feature.

Chapter 26: Meeting The Goal

Chapters twenty-six through thirty are the payoff sequence. The plant meets its three-month goal and then some. Order backlog shrinks. On-time delivery rises. Inventory falls. The plant begins winning new business that it could not previously compete for, including a large order from a customer whose previous supplier could not meet a lead-time commitment. Bill Peach withdraws the threat of closure. Alex's domestic plot line, which has been running in parallel, reaches its own resolution as Julie returns from her parents' house and Alex begins applying the Socratic method to the marriage.

- The plant's throughput rises measurably within one quarter after the bottleneck changes are installed. Inventory falls in parallel. Operational expense stays roughly constant. All three system measures move in the right direction.
- The improvements come from operating changes, not from capital investment. The retired older machines that come out of storage are essentially free. The changes to the heat-treat sequencing cost nothing. The color-tag system is administrative.
- The plant's success creates its next problem. As the NCX-10 constraint is elevated, the constraint moves. Alex's team discovers that the plant's next bottleneck is not in the plant. It is the market. The plant can now produce faster than the sales team can sell. Goldratt uses this shift to make the point that the constraint is dynamic.
- The domestic plot line resolves through the same Socratic method Alex has been learning at the plant. The novel treats the family and the plant as two settings for the same operating discipline. Both require the leader to stop optimizing local measures and start asking what the system is for.

Chapter 31: The Five Focusing Steps

The final ten chapters, thirty-one through forty, promote Alex to division manager and force him to codify what he has learned so that he can teach it to other plant managers. Jonah leaves the story. Alex and Lou develop the Five Focusing Steps and a second sequence, the three questions of change: what to change, what to change to, and how to cause the change. The 30th Anniversary Edition adds a section that codifies the Five Focusing Steps and includes an essay by Goldratt, *Standing on the Shoulders of Giants*, which places the Theory of Constraints alongside Ford, Toyota, and other operating traditions.

- The Five Focusing Steps. First, identify the system's constraint. Second, decide how to exploit the constraint. Third, subordinate everything else to the exploitation decision. Fourth, elevate the constraint through investment if the first three steps have been exhausted. Fifth, return to step one, and do not let inertia turn the fix into a new constraint.
- The three questions of change. What to change. What to change to. How to cause the change to happen. Goldratt treats the third question as the hardest, and the one his later books focus on. The scientific method identifies what to change. The Theory of Constraints identifies what to change to. Change management is a separate discipline.
- The constraint is dynamic. Once a bottleneck is elevated, the constraint moves. In Alex's case the plant constraint moves to the market. In other systems it moves to another operation, to a policy, or to a mindset. The step-five warning against inertia is that the operating rules built for the previous constraint often become the next constraint.
- The novel form matters. Goldratt argued that a textbook of the Theory of Constraints would not have reached factory managers. The novel embeds the concepts in a specific operating story, and the reader learns the concepts by watching them work in the plant. Forty years after publication the book has sold millions of copies and is still the most-assigned operations text in business schools.

About the Authors

Eliyahu M. Goldratt, born in Israel in nineteen forty-seven and died in twenty eleven, was a physicist by training and the developer of the Theory of Constraints. His first commercial product was OPT, a scheduling software he built in the nineteen seventies for manufacturing companies. When the software's principles proved difficult to teach through consulting alone, Goldratt collaborated with Jeff Cox to write *The Goal* in nineteen eighty-four. Cox is a professional business novelist who has co-written several books that translate operating concepts into narrative form. Goldratt went on to write six more Theory of Constraints books, including *It's Not Luck*, *Critical Chain*, and *Necessary but Not Sufficient*, and founded the Goldratt Consulting group.

The Goal has been through five editions since nineteen eighty-four. The Fourth Revised, Thirtieth Anniversary Edition of twenty fourteen is the current authoritative version and differs meaningfully from earlier editions. It adds a codified statement of the Five Focusing Steps that appeared piecemeal in earlier printings, adds the *Standing on the Shoulders of Giants* essay in which Goldratt places the Theory of Constraints alongside Ford's flow line and Toyota's just-in-time system, and moves the *Fortune Small Business* case studies that had been in the twenty oh four edition to the publisher's website. *Time* named the book one of the twenty-five most influential business management books in twenty eleven.

Key Takeaways

1) The goal is to make money, and that changes what you measure.

The book's central operating claim is that most operating dysfunction comes from teams optimizing measures that are not the goal. If the goal is to make money, the three measures that matter are throughput, inventory, and operational expense. Every operating decision moves at least one of them. The traditional cost-accounting measures, machine efficiency, cost per part, direct labor utilization, are local measures that describe stations rather than the system.

- Throughput is the rate at which the system generates money through sales. Not production. Not shipments. Money in the door.
- Inventory is money currently tied up in the system. Raw materials, work-in-process, and finished goods are all inventory in this frame.

- Operational expense is the money spent to turn inventory into throughput. Payroll, utilities, depreciation, and consumables.

2) Every system has a constraint, and the whole system's output moves only when the constraint moves.

The Theory of Constraints's operating claim is that the throughput of any system is governed by one resource, one policy, or one behavior. Improvements everywhere else produce inventory and no throughput. The Bearington plant's NCX-10 machining center and its heat-treat oven were the constraints. Improvements at every other station made no difference to on-time delivery until the two constraints were exploited.

- Rule one, an hour lost at a bottleneck is an hour lost for the whole system. Rule two, an hour saved at a non-bottleneck is a mirage.
- The constraint is where work piles up in front and starves behind. In office and service work it is the step where the queue is longest and the downstream team waits.
- Improvements at non-constraints are attractive to management because they are visible and measurable in local terms. The Goldratt argument is that these improvements pay off only when the freed capacity feeds the constraint.

3) Balancing capacity everywhere is exactly the wrong design.

Traditional plant design and traditional operations research reward matching capacity at every station to average demand. Goldratt's contribution is to demonstrate that a balanced line cannot deliver average throughput, because dependent events with statistical fluctuations compound. The scouts on the hike are the model. Herbie's average pace is the same as everyone else's, but because his slow steps cannot be made up by faster scouts ahead of him in the line, the whole troop is paced by Herbie.

- Excess capacity at non-constraints is a feature, not waste. It absorbs the variability that would otherwise starve the constraint.
- The correct design has one constraint with protective capacity everywhere else. Cost-per-part measures will condemn this design. The alternative is to condemn the measures.
- The consequence for capital investment is that ROI on paper is not the same as ROI in the plant. An investment that raises a non-constraint's efficiency reduces cost per part

locally and adds nothing to system throughput.

4) Drum-buffer-rope is the scheduling application.

The drum is the constraint, whose output rate paces the whole system. The buffer is protective inventory placed just ahead of the constraint so a hiccup upstream does not starve it. The rope is the signal from the constraint back to material release, so new material enters the system only at the rate the constraint can consume. The practical effect is that inventory falls, cycle time shortens, and the constraint runs continuously. The plant's throughput becomes reliable rather than variable.

- The color-tag system Alex's team introduces at Bearington, red tags for parts that pass through the constraints and green tags for parts that do not, is a physical implementation of drum-buffer-rope.
- Batch sizes should be smaller than cost-accounting rules would set them. The setup-cost objection is that setup time rises. The Goldratt reply is that setup time at a non-bottleneck is free because the non-bottleneck has protective capacity.
- Material release rate is a policy variable, not a fixed operational necessity. Slowing material release usually raises throughput, because it prevents the plant from piling inventory in front of a constraint that cannot consume it faster.

5) The Five Focusing Steps are the discipline.

Identify the constraint. Exploit it. Subordinate everything else to it. Elevate it. Return to step one and do not let inertia turn the fix into a new constraint. The steps run in order, and the order matters. Elevate before exploit is the mistake most companies make. Investing in more capacity at the constraint before the current constraint has been squeezed produces expensive marginal capacity when free capacity was still available.

- Step two, exploit, is usually free or nearly free. Squeezing the constraint's current utilization to one hundred percent through better setup, quality-check placement, and batch design typically produces double-digit throughput gains without capital.
- Step three, subordinate, is the counterintuitive step. Non-constraints must run at the constraint's pace, which means they will appear underutilized in cost-accounting reports. This is correct.

- Step five, do not let inertia set the next constraint, warns that the operating rules built for the previous constraint often become the next one. Policies outlast the conditions that created them.

6) The constraint is dynamic, and the market becomes the constraint most often.

Once a plant constraint is elevated, the constraint moves. In Alex's case the constraint moves from the NCX-10 to the market. The plant can now produce faster than the sales team can sell. Goldratt's later work extends the point. For most growth-stage companies the operating constraints are exploited quickly through process design, and the durable constraint becomes demand generation, pricing, or market access. The Theory of Constraints applies to marketing, sales, and product-development flow as directly as it applies to manufacturing.

- The pattern of operating constraint into market constraint is now the common shape of growth-company scaling. A team that has run the Five Focusing Steps in operations should run them again in commercial functions.
- Sales pipelines have constraints. Product development has constraints. Customer onboarding has constraints. The identification method is the same. Find where the queue is longest and the downstream is waiting.
- The framework does not conflict with strategic positioning or product-market fit. It names the operational binding constraint on which throughput currently depends, and leaves the strategic question to a different set of tools.

7) Change management is the third question and the hardest one.

Goldratt names three questions of change. What to change. What to change to. How to cause the change. The scientific method addresses the first. The Theory of Constraints addresses the second. The third question, how to actually make the change happen in an organization full of people whose measures reward the old behavior, is where most implementations fail. Alex's team spends chapters twenty-one through twenty-five as much on convincing the plant to adopt the new priorities as on designing them. The union representative, the accounting function, and the corporate measurement system all resist for different reasons.

- The measurement system is usually the largest source of resistance. If a plant manager is measured on machine efficiency, telling him to run non-constraints below capacity will cost him his job.
- The Socratic method is Goldratt's preferred change mechanism. Jonah does not tell Alex the answer. He asks questions until Alex arrives at it. The mechanism is slower than instruction and produces more durable behavior change.
- The corporate function most often resistant to the operating changes is the finance and accounting function, because the measures it produces are the ones the operating changes contradict. Aligning the measures with the new operating model is a precondition for the changes sticking.

Strengths And Limits

The book's strength is the novel form itself. Forty years and roughly seven million copies in, *The Goal* is still the most-assigned operations text in business schools because the narrative embeds the concepts in a specific operating story that the reader can follow without prior training in operations research. The Herbie hike, the NCX-10 machining center, and the heat-treat oven are memorable enough to survive re-tellings, which means the concepts survive with them. The Five Focusing Steps and the drum-buffer-rope scheduling application are precise enough to argue with and general enough to apply outside manufacturing. The framework has since been extended to project management through Critical Chain, to distribution and supply chain, and to service and knowledge-work flows. Goldratt refuses the compromise position on measurement. He argues that cost-accounting measures actively obstruct the improvements the operating model needs, and that the finance function's role has to change alongside the operations. Most operations writing hedges on this point. Goldratt does not.

The limits are real. The book is about a factory, and the factory is a nineteen eighties factory. Readers in software, professional services, or high-mix low-volume operations have to translate the material into their own operating shape, and the translation is not always obvious. The domestic subplot is dated and reads as a distraction to some modern readers, though the Socratic-method parallel it supports is central to the argument. The book is thin on the mechanics of running Theory of Constraints in a

multi-plant or multi-product organization, and Goldratt's later books, particularly *It's Not Luck* and *Critical Chain*, are the necessary complements for readers whose operations are more complex than a single plant. Finally, the book's presentation of cost accounting is deliberately unsympathetic. Cost accountants who have grown up with Activity-Based Costing and modern managerial-accounting practice will find the caricature dated. The underlying argument, that local measures obscure system performance, still holds. The caricature is the weakest part of the delivery.

Bruce's Take

This is the book I hand to any founder-CEO who tells me their operations team is busy and their output is flat. The diagnosis is exact and it has aged better than almost anything else written in the nineteen eighties. Every growth-stage company I work with has a Bearington plant somewhere, and it is almost never in manufacturing. It is in the sales pipeline, in the onboarding sequence, in the product-release cycle. The team is busy, the local dashboards look fine, and system output is not moving. Goldratt's frame gives the CEO a way to name the problem and a place to point the improvement effort. That combination is rare.

The context sensitivity is worth naming. In a hardware or industrial business the book maps almost directly, and the Five Focusing Steps and drum-buffer-rope translate into the current operation with minimal adaptation. In a SaaS or professional services business the translation is harder. The constraint is often a person rather than a machine, the buffer is a queue of engagements rather than a pile of work-in-process, and the political cost of subordinating a highly paid non-constraint is different from asking a machine to sit idle. The book works in these contexts, but the reader has to do the translation. Where I see founders get stuck is trying to apply the mechanics before they have accepted the underlying shift from local efficiency to system throughput.

The most credible outside objection comes from the Lean and Toyota Production System camp, articulated by writers like James Womack and Daniel Jones in *Lean Thinking*. Their argument is that Goldratt over-privileges the constraint and under-privileges continuous flow, and that a fully lean system with small batches, quick changeovers, and pulled demand does not need drum-buffer-rope because it does not accumulate the inventory that drum-buffer-rope manages. My read is that the objection lands in mature repetitive manufacturing where lean has room to run. For

most of my growth-stage clients, whose operations are high-mix and less mature, Goldratt's constraint frame is easier to install and produces faster results. Lean is the long game. Theory of Constraints is the fast diagnostic.

The book that completes what The Goal leaves incomplete is The Phoenix Project by Gene Kim, Kevin Behr, and George Spafford. It is the deliberate homage to Goldratt written for IT and DevOps, and it does more than translate the metaphors. It shows the Five Focusing Steps running inside a knowledge-work organization where the constraint is a single overworked engineer rather than a machine, and where the political mechanics of subordination look like reassigning executive attention rather than idling equipment. Read The Goal first for the argument. Read The Phoenix Project second for the operating manual in a business that is not a factory.

Discussion Questions

For CEOs

- What is the current binding constraint on my company's throughput, and can I name it specifically enough that my leadership team would give the same answer?
- Which of my top five improvement initiatives is aimed at the current constraint, and which are aimed at non-constraints that will produce local efficiency without changing system output?
- For each of my operating measures, machine efficiency, cost per unit, utilization, is the measure aligned with throughput, inventory, and operational expense, or does it reward behavior that opposes those?
- Where in the organization are we investing in capacity elevation before we have exhausted exploitation of the current capacity, and what would exploit look like specifically?
- If my operating team ran the Five Focusing Steps on our commercial flow, sales pipeline, customer onboarding, product development, what would the constraint be, and what have we been doing about it?
- How much of my inventory, in the accounting sense and in the broader sense of work-in-process across all functions, is money tied up in work that has not yet become throughput?

- When my leadership team identifies a bottleneck and proposes to elevate it, who is responsible for asking whether exploit and subordinate have been completed first, and what happens if the answer is no?
- If the operating constraint moved to the market next quarter, would my sales and marketing team recognize itself as the new constraint, and what change would that recognition require?

For Teams

- For our primary operating flow, where is the queue currently longest and the downstream team most often waiting, and does that match what we would call our biggest problem?
- Which of our current improvement projects is aimed at the constraint, and which are aimed elsewhere? What system throughput improvement does each promise?
- Where are we currently rewarding local efficiency at non-constraints, and what would change if we let those stations run below capacity to feed the constraint?
- How does our current material-release or work-intake policy compare to a drum-buffer-rope design that would throttle input to match the constraint's output?
- What is our current batch size, in operations, in sales, in project intake, and would smaller batches release constraint capacity at acceptable setup cost?
- For the last three capital investments we made, which raised capacity at the current constraint and which raised capacity elsewhere?
- Which of the Five Focusing Steps do we currently run to completion, and which do we skip under pressure?
- If we finished exploiting our current operating constraint next quarter and it moved to another function, would that function's team recognize itself as the new constraint, or would the improvement stall there?

Practical Tools

1. Constraint Identification

The first of the Five Focusing Steps is identify, and it is the step most operating teams skip. Naming the constraint precisely enough that everyone gives the same answer is a prerequisite for exploit and subordinate to produce anything. Use it at the start of any process-improvement effort, in an operating-model review, or when the team is running multiple improvement projects and the results have not compounded.

How it works:

- Choose one flow. It can be a production process, a sales pipeline, a customer onboarding sequence, or a product-development lifecycle.
- Map the flow as a sequence of steps. Include the queue that sits in front of each step, and the throughput rate the step delivers.
- For each step, note where the queue is longest and where the downstream step waits. The step whose queue is longest and whose downstream is starving is the constraint.
- Ask three team members to name the constraint independently. If the answers differ, the constraint is not yet clear and the mapping needs more detail.
- State the constraint explicitly in one sentence. Publish it. Every subsequent improvement decision refers back to this sentence.

The team leaves with a single named constraint that every member can point to, and the improvement backlog can be resorted around it. Multiple ongoing projects that were producing incremental local gains can now be evaluated for their contribution to lifting the named constraint.

2. Exploit Before Elevate

The order of the Five Focusing Steps matters. Elevating a constraint before exploiting it produces expensive marginal capacity while free capacity is still available. Squeezing the current constraint through better sequencing, quality control before the constraint, and reduced setup or handoff time is the highest-return operating move most companies overlook. Use it whenever a capital investment or hiring request is proposed to raise capacity at the constraint. Run the exploit exercise before the investment is approved.

How it works:

- State the current constraint. Measure its current utilization. Include time lost to setup, changeover, quality inspection, defective input, and idle time.
- For each source of lost time, identify a fix that does not require capital. Setup operators dedicated to the constraint. Quality inspection before the constraint. Batch sequencing to reduce changeover. Feeding the constraint during breaks.
- Estimate the throughput gain from each fix. Rank the fixes by effort and return.
- Implement the top two or three fixes. Measure the constraint's utilization and the system's throughput before and after.
- Reassess whether elevation is still needed. In many cases the exploit sequence produces enough capacity that the original investment case no longer holds.

The constraint's utilization rises measurably, system throughput rises with it, and the team gains a template for future exploit exercises. The habit of running exploit before elevate reduces capital requests over time.

3. Subordinate The Non-Constraints

Subordination is the step that feels wrong to plant managers and to modern knowledge-work leaders alike. A non-constraint should run at the constraint's pace, not at its own maximum. Cost-accounting measures resist this, because non-constraints will appear underutilized. Use it once exploit is in progress at the constraint, and any time a team member raises a concern that a non-constraint station is being under-worked.

How it works:

- For each non-constraint station in the flow, note its current utilization and its current output rate.
- Calculate the constraint's pace. Set each non-constraint's target pace to match the constraint's, plus a small buffer to protect against variability.
- Explicitly authorize the non-constraints to run below their maximum. Note the cost-accounting implications for the operating-review meeting that will follow.
- In the operating review, defend the subordination decision with the throughput data rather than with the utilization data. Show the finance team that system throughput is up.
- Update the measurement system so that the non-constraints are not penalized for the subordination. Constraint utilization and system throughput become the measures.

Non-constraints run at the constraint's pace, work-in-process falls between stations, and system throughput rises. The finance team's monthly report begins reflecting the new measures rather than the old ones.

4. Drum-Buffer-Rope Design

Drum-buffer-rope is the Theory of Constraints's scheduling application. It ties the system's input rate to the constraint's output rate through a buffer that protects the constraint from starvation. The design applies to manufacturing scheduling, to professional-services intake, to sales-pipeline pacing, and to product-development sprint planning. Use it when the current work-intake or scheduling policy is producing high work-in-process inventory, long cycle times, or unpredictable delivery.

How it works:

- Identify the drum. This is the constraint. Its throughput rate paces the whole flow.
- Design the buffer. Place a protective queue immediately in front of the drum, measured in time or units. Its size covers the variability upstream. Too small starves the drum. Too large hides upstream problems.
- Design the rope. This is the signal from the drum back to the intake point. New work enters the flow at the drum's consumption rate, not faster.

- Publish the drum, buffer, and rope explicitly. Everyone in the flow should know where the drum is, what the buffer size is, and what the intake rule is.
- Monitor the buffer. When the buffer runs low, expedite upstream. When it grows, throttle intake. The buffer becomes the operational control point, not the drum itself.

Work-in-process falls. Cycle time shortens. On-time delivery becomes reliable. The drum runs continuously, which is the throughput of the whole system.

5. Measurement Redesign

The largest source of resistance to Theory of Constraints implementations is the measurement system. If plant managers are measured on machine efficiency, sales leaders on pipeline volume, and engineering teams on story points delivered, they will optimize those measures regardless of what the strategy says. Use it before announcing any Theory of Constraints change, or immediately when a previous implementation has stalled and the team is reverting to old behaviors.

How it works:

- List every operating measure currently reviewed monthly by the leadership team. Include finance, operations, sales, and product measures.
- For each measure, ask whether it rewards system throughput, inventory reduction, or operational-expense discipline, or whether it rewards a local behavior that opposes one of these.
- For the measures that oppose the system goals, redesign the measure or remove it. Machine efficiency becomes constraint utilization plus system throughput. Pipeline volume becomes closed-won revenue with cycle time.
- Publish the new measurement dashboard before any operating change is announced. Get the finance function to own the new measures.
- Review after one quarter. Measures that operating teams have quietly ignored will reappear in team behavior. Kill them.

The operating team's daily and weekly attention shifts to the new measures. Individual efficiency measures fade. System throughput becomes the shared measure. The behavior change lasts, because the measurement change lasts.

6. Post-Improvement Constraint Search

The fifth focusing step warns that once a constraint is elevated, the constraint moves, and inertia often turns the previous constraint's operating rules into the new constraint. The search prevents the pattern by explicitly relocating the constraint after any successful improvement cycle. Use it as soon as the previous constraint stops being the binding limit on system throughput, in practice whenever a bottleneck-relieving investment or process change has produced its expected throughput gain.

How it works:

- State the previous constraint and the change that elevated it. Confirm that the change has produced the expected throughput gain.
- Repeat the constraint-identification exercise. Where is the new queue longest and the downstream now waiting? The new constraint is often not in the operating flow.
- For each candidate for the new constraint, ask whether it is another operating step, a policy carried over from the previous constraint's operating rules, or a commercial or market limit.
- Name the new constraint explicitly. Reset the improvement backlog to focus on it. Retire the operating rules that were designed for the previous constraint and are now producing local optimization at the new constraint.
- Communicate the shift to the leadership team. The team's attention must move with the constraint. If it does not, the team will keep exploiting the previous constraint that no longer binds.

The organization runs the Five Focusing Steps continuously rather than as a one-time exercise. Each improvement cycle lifts the constraint, discovers the next one, and resets the operating priorities. The system becomes a process of ongoing improvement, which is the book's subtitle.

7. Three Questions Of Change

Goldratt's three questions of change, what to change, what to change to, and how to cause the change, name the three separate skills required for a Theory of Constraints implementation to hold. Most stalls happen at the third question, because the operating team knows what to change and what to change to but cannot get the measurement system, the incentives, or the culture to accept the new operating rules. Use it when a Theory of Constraints change was designed correctly and is failing to produce the expected system throughput gain, or when a previous improvement is reverting.

How it works:

- State clearly what was to change and what it was to change to. Confirm the design is still correct.
 - For the change to happen, list every operating rule, measure, incentive, or cultural assumption in the organization that the change contradicts.
 - For each contradiction, choose one of three responses. Redesign the operating rule. Redesign the incentive. Or explain the cultural assumption openly and give the team the vocabulary to override it.
 - Publish the response plan alongside the operating change itself. Do not announce the change without the accompanying rule, incentive, and cultural adjustments.
 - Review after one quarter. Track which contradictions have been resolved and which the team has quietly reverted to. The reversions name the next round of change work.
- The operating change holds through at least one full operating cycle. The measures, the incentives, and the cultural vocabulary all support the new behavior. Reversion is rare, and when it happens, the team can name the specific contradiction and fix it.

Further Research

- Eliyahu Goldratt's Critical Chain, which applies the Theory of Constraints to project management and buffer scheduling. The natural sequel for readers whose primary flows are project-based rather than manufacturing-based. ([link](#))

- Eliyahu Goldratt's *It's Not Luck*, the direct novel sequel to *The Goal*, which extends the Thinking Processes to broader business problems including marketing, distribution, and corporate strategy. ([link](#))
- Eliyahu Goldratt's *Necessary but Not Sufficient*, a business novel that treats ERP implementation through the Theory of Constraints lens and argues that most large software implementations fail because they change the technology without changing the operating measures. ([link](#))
- James Womack and Daniel Jones's *Lean Thinking*, which develops a parallel operations framework from the Toyota Production System tradition. Compatible with the Theory of Constraints on flow and on the primacy of throughput, different on batch design and on the role of the constraint. ([link](#))
- H. Thomas Johnson and Anthony Atkinson's *Relevance Regained*, the reference academic critique of traditional cost accounting that Goldratt's operating argument depends on. Provides the accounting-side case for the measurement redesign the operating team needs to see. ([link](#))

Related Books

- **Critical Chain**, Eliyahu M. Goldratt. The Theory of Constraints applied to project management, with buffer scheduling that solves the multi-project resource contention problem. The natural sequel for readers whose primary flows are project-based rather than manufacturing-based.
- **It's Not Luck**, Eliyahu M. Goldratt. The direct novel sequel to *The Goal*, which extends Alex Rogo's story into a corporate role and introduces the Thinking Processes, a broader set of tools for identifying and resolving conflicts in business strategy.
- **The Phoenix Project**, Gene Kim, Kevin Behr, and George Spafford. A novel about IT operations that uses the same narrative form and the same Theory of Constraints framework to teach DevOps. Written explicitly as an homage to *The Goal*, and the best translation of Goldratt's ideas into knowledge work.

- **Lean Thinking**, James Womack and Daniel Jones. The Toyota Production System framework, complementary to Theory of Constraints on flow and throughput. Different on batch design, on the role of the constraint, and on the path from current state to future state.
- **The Machine That Changed the World**, James Womack, Daniel Jones, and Daniel Roos. The MIT study of the automotive industry that launched the Western lean movement. Companion reading to The Goal for readers who want the academic evidence for the operating claims Goldratt makes narratively.
- **The Toyota Way**, Jeffrey Liker. The management-principle statement of the Toyota Production System, with fourteen principles organized around long-term philosophy, process, people, and problem-solving. Reads well alongside Goldratt for a leader who wants both traditions.
- **Reengineering the Corporation**, Michael Hammer and James Champy. The nineteen ninety-three business-process reengineering statement, useful for context because it is the alternative operating framework The Goal was competing with in that decade. Different premise, and Goldratt's work has held up better.

These summaries present Bruce Eckfeldt's analysis of business and leadership books for founder-CEOs and leadership-team members of growth companies. Each one identifies the ideas most relevant to decision quality, strategy, and leadership capacity, notes where the evidence is weak, and suggests practical ways to test what is worth testing. All ideas, stories, and quotations belong to the original author and to the people they cite. The assessments and recommendations are Bruce's. It is a companion to the book, not a substitute for reading it.

To learn more about Bruce's programs, request other book summaries, or get in touch, contact booksummaries@eckfeldt.com.