Working Backwards

Insights, Stories, and Secrets from Inside Amazon

Colin Bryar
and Bill Carr

Working Backwards

Insights, Stories, and Secrets from Inside Amazon

Colin Bryar and Bill Carr

St. Martin's Press
New York

4

Communicating

Narratives and the Six-Pager

The eerie silence at the beginning of Amazon meetings. The ban on PowerPoint and the shift to narratives. How narratives produce clear thinking and stimulate valuable discussion. How to write an effective six-pager. The payoff: the "narrative information multiplier."

* * *

If you were to ask recently hired Amazon employees about what has surprised them most in their time at the company so far, one response would certainly top the list:

"The eerie silence in the first 20 minutes of many meetings."

At Amazon, after a brief exchange of greetings and chitchat, everyone sits at the table, and the room goes completely silent. Silent, as in not a word. The reason for the silence? A six-page document that everyone must read before discussion begins.

Amazon relies far more on the written word to develop and communicate ideas than most companies, and this difference makes for a huge competitive advantage. In this chapter we'll talk about how and why Amazon made the transition from the use of PowerPoint (or any other presentation software) to written narratives, and how it has benefited the company—and can benefit yours too.

Amazon uses two main forms of narrative. The first is known as the "six-pager." It is used to describe, review, or propose just about any type of idea, process, or business. The second narrative form is the PR/FAQ. This one is specifically linked to the Working Backwards process for new product development. In this chapter, we'll focus on the six-pager and in the following chapter we'll look at the PR/FAQ.

The End of PowerPoint at S-Team Meetings

One of my (Colin's) roles as Jeff's shadow in the early days of the company was to manage the agenda of the weekly S-Team meeting, which took place every Tuesday and typically ran for four hours. Roughly 80 percent of the time was focused on execution, namely how the company was making progress toward achieving the S-Team goals. In the S-Team meeting, we would select between two and four S-Team goals and do a deep dive on their progress. The meeting was expensive: between preparation and attendance, it consumed at least half a day each week for the top leaders in the company. Given the types of decisions made in the meeting, the stakes were high.

In those early days, each deep dive would begin with a presentation by the relevant team on the status of their work toward the goal. Typically, this involved an oral presentation by one or more of the team members backed up by PowerPoint slides. Too often, we found, the presentations did not serve the purpose for which they were intended. The format often made it difficult to evaluate the actual progress and prevented the presentations from proceeding as planned. The deep dives were, in short, frustrating, inefficient, and error prone for both the presenter and the audience.

Jeff and I often discussed ways to improve the S-Team meetings. Shortly after a particularly difficult presentation in early 2004, we had some downtime on a business flight (no Wi-Fi yet on planes), so we read and discussed an essay called "The Cognitive Style of PowerPoint: Pitching Out Corrupts Within," by Edward Tufte, a Yale professor who is an authority on the visualization of information.[1] Tufte identified in one sentence the problem we'd been experiencing: "As analysis becomes more causal, multivariate, comparative, evidence based, and resolution-intense," he writes, "the more damaging the bullet list becomes." That description fit our discussions at the S-Team meetings: complex, interconnected, requiring plenty of information to explore, with greater and greater consequences connected to decisions. Such analysis is not well served by a linear progression of slides that makes it difficult to refer one idea to another, sparsely worded bits of text that don't fully express an idea, and visual effects that are more distracting than enlightening. Rather than making

things clear and simple, PowerPoint can strip the discussion of important nuance. In our meetings, even when a presenter included supporting information in the notes or accompanying audio, the PowerPoint presentation was never enough.

Besides, the Amazon audience of tightly scheduled, experienced executives was eager to get to the heart of the matter as quickly as possible. They would pepper the presenter with questions and push to get to the punch line, regardless of the flow of slides. Sometimes the questions did not serve to clarify a point or move the presentation along but would instead lead the entire group away from the main argument. Or some questions might be premature and would be answered in a later slide, thus forcing the presenter to go over the same ground twice.

In his essay, Tufte proposed a solution. "For serious presentations," he wrote, "it will be useful to replace PowerPoint slides with paper handouts showing words, numbers, data graphics, images together. High-resolution handouts allow viewers to contextualize, compare, narrate, and recast evidence. In contrast, data-thin, forgetful displays tend to make audiences ignorant and passive, and also to diminish the credibility of the presenter."

Tufte offered wise advice on how to get started. "Making this transition in large organizations requires a straightforward executive order: *From now on your presentation software is Microsoft Word, not PowerPoint. Get used to it*." That is essentially what we did.

While Tufte's essay wasn't the sole impetus behind the move to narratives, it crystallized our thinking. On June 9, 2004, the members of the S-Team received an email with the following subject line: "No PowerPoint presentations from now on at S-Team."[2] The message was simple, direct, and earthshaking: from that day forward, S-Team members would be required to write short narratives describing their ideas for presentation at S-Team meetings. PowerPoint was henceforth banned.

I (still Colin) was the one who sent the email—at Jeff's direction, of course, as he was the only person in the company who could mandate such a significant change. I felt great after sending it. We had finally found a way to meaningfully improve the effectiveness of the S-Team meetings, so I thought the email would be well received. Boy was I wrong. The email whipped through the Amazon management ranks, and the almost instantaneous and near-universal reaction was basically, "You must be

kidding." That evening and for the next few days I fielded a flurry of phone calls and a deluge of emails asking about the change. The outcry was particularly intense from the S-Team members who were scheduled to present within the next two weeks. They had to quickly understand the new narrative process and learn to effectively use the tools at their disposal. And the fate of a new idea that may have been months in development was riding on the outcome of the meeting.

We probably should not have been surprised by that reaction. Until that June day in 2004, PowerPoint had been the default tool for communication of ideas in many meetings at Amazon, just as it was and still is at many companies. Everybody knew its delights and perils. What could be more exhilarating than listening to a charismatic executive deliver a rousing presentation backed up by snappy phrases, dancing clip art, and cool slide transitions? So what if you couldn't remember the details a few days later? And what could be worse than suffering through a badly organized presentation using a drab template and tons of text in a font too small to read? Or, worse still, squirming as a nervous presenter stumbled and faltered through slide after slide?

The real risk with using PowerPoint in the manner we did, however, was the effect it could have on decision-making. A dynamic presenter could lead a group to approve a dismal idea. A poorly organized presentation could confuse people, produce discussion that was rambling and unfocused, and rob good ideas of the serious consideration they deserved. A boring presentation could numb the brain so completely that people tuned out or started checking their email, thereby missing the good idea lurking beneath the droning voice and uninspiring visuals.

It would take time for people to get the hang of the narrative form. First, there were no codified rules about what the narrative should be, and Jeff offered a short explanation of the reason behind the change.

> The reason writing a good 4 page memo is harder than "writing" a 20 page powerpoint is because the narrative structure of a good memo forces better thought and better understanding of what's more important than what, and how things are related.
>
> Powerpoint-style presentations somehow give permission to gloss over ideas, flatten out any sense of relative importance, and ignore the

interconnectedness of ideas.[3]

The first few narratives were laughably poor when evaluated by today's standards. Some teams ignored the length limit, which was meant to keep the narratives brief enough so they could be read in the meeting itself. Enthusiastic teams, who felt their idea could not be adequately expressed in such a limited space, came in with 30 or 40 pages of prose. When authors learned that we were serious about a page limit, some squeezed as much as text onto a page as possible, using tiny fonts, reducing the width of the margins, and single-spacing the text. We wanted to go back to the benefits of writing, but not to the look of a sixteenth-century document.

Gradually, we settled on a standard format. Maximum length: six pages, no desperate tricks in formatting please. Appendices with further information or supporting detail could be attached, but would not be required reading in the meeting itself.

How to Write an Effective Six-Pager

Six-pagers vary widely, so rather than attempting a complete style guide (impossible), we've written one in a style we might submit today, if we were recommending for the first time that we use narratives instead of PowerPoint at S-Team meetings—a six-pager about six-pagers. Some of this is a pared-down version of what you've just read, which may help you see how we squeeze big ideas into the format of a true six-pager. (Note: this example would fit easily onto six pages of 8.5 x 11–inch paper, single-spaced in 11-point type, but reproduction in this book may run longer due to formatting differences.)

Dear PowerPoint: It's Not You, It's Us

Our decision-making process simply has not kept up with the rapid growth in the size and complexity of our business. We therefore advocate that, effective immediately, we stop using PowerPoint at S-Team meetings and start using six-page narratives instead.

What's Wrong with Using PowerPoint?

S-Team meetings typically begin with a PowerPoint (PP) presentation that describes some proposal or business analysis for consideration. The style of the deck varies from team to team, but all share the constraints imposed by the PowerPoint format. No matter how complex or nuanced the underlying concepts, they are presented as a series of small blocks of text, short bullet-pointed lists, or graphics.

Even the most ardent PP fans acknowledge that too much information actually spoils the deck. Amazon's bestselling book on PowerPoint describes three categories of slides:

1. 75 words or more: A dense discussion document or white paper that is not suitable for a presentation—it's better distributed in advance and read before the meeting.
2. 50 words or so: A crutch for the presenter who uses it as a teleprompter, often turning away from an audience while reading aloud.
3. Even fewer words: A proper presentation slide, used to visually reinforce primarily spoken content. The presenter must invest time to develop and rehearse this type of content.*

One widely accepted rule of thumb, the so-called 6x6 Rule, sets a maximum of six bullet points, each with no more than six words. Other guidelines suggest limiting text to no more than 40 words per slide, and presentations to no more than 20 slides. The specific numbers vary, but the theme—limiting information density—is a constant. Taken as a whole, these practices point to a consensus: there's only so much information one can fit into a PP deck without confusing, or losing, one's audience. The format forces presenters to condense their ideas so far that important information is omitted.

Pressed against this functional ceiling, yet needing to convey the depth and breadth of their team's underlying work, a presenter—having spent considerable time pruning away content until it fits the PP format—***fills it back in, verbally***. As a result, the public speaking skills of the presenter, and the graphics arts expertise behind their slide deck, have an undue—and highly variable—effect on how well their ideas are understood. No matter how much work a team invests in developing a proposal or business analysis, its ultimate success can therefore hinge upon factors irrelevant to the issue at hand.

We've all seen presenters interrupted and questioned mid-presentation, then struggle to regain their balance by saying things like, "We'll address that in a few slides." The flow becomes turbulent, the audience frustrated, the presenter flustered. We all want to deep dive on important points but have to wait through the whole presentation before being satisfied that our questions won't be answered somewhere later on. In virtually every PP presentation, we have to take handwritten notes throughout in order to record the verbal give-and-take that actually supplies the bulk of the information we need. The slide deck alone is usually insufficient to convey or serve as a record of the complete argument at hand.

Our Inspiration

Most of us are familiar with Edward Tufte, author of the seminal (and Amazon bestselling) book *The Visual Display of Quantitative Information*. In an essay titled "The Cognitive Style of PowerPoint: Pitching Out Corrupts Within," Tufte encapsulates our difficulties precisely:

> As analysis becomes more causal, multivariate, comparative, evidence based, and resolution-intense, the more damaging the bullet list becomes.

This certainly describes S-Team meetings: complex, interconnected, requiring plenty of information to explore, with greater and greater consequences connected to decisions. Such analysis is not well served by a linear progression of slides, a presentation style that makes it difficult to refer one idea to another, to fully express an idea in sparsely worded bits of text, and to enlighten instead of distract with visual effects. Rather than making things clear and simple, PowerPoint is stripping our discussions of important nuance.

Tufte's essay proposes a solution. "For serious presentations," he writes, "it will be useful to replace PowerPoint slides with paper handouts showing words, numbers, data graphics, images together. High-resolution handouts allow viewers to contextualize, compare, narrate, and recast evidence. In contrast, data-thin, forgetful displays tend to make audiences ignorant and passive, and also to diminish the credibility of the presenter."

He goes on: "For serious presentations, replace PP with word-processing or page-layout software. Making this transition in large organizations requires a straightforward executive order: From now on your presentation software is Microsoft Word, not PowerPoint. Get used to it." We've taken this recommendation to heart, and we now propose to follow his advice.

Our Proposal: Banish PP in Favor of Narratives

We propose that we stop using PowerPoint in S-Team meetings immediately and replace it with a single narrative document. These narratives may sometimes include graphs and bulleted lists, which are essential to brevity and clarity, but it must be emphasized:

merely reproducing a PP deck in written form will NOT be acceptable. The goal is to introduce the kind of complete and self-contained presentation that only the narrative form makes possible. **Embrace it**.

Our Tenet: Ideas, Not Presenters, Matter Most

A switch to narratives places the team's ideas and reasoning center stage, leveling the playing field by removing the natural variance in speaking skills and graphic design expertise that today plays too great a role in the success of presentations. The entire team can contribute to the crafting of a strong narrative, reviewing and revising it until it's at its very best. It should go without saying—sound decisions draw from ideas, not individual performance skills.

The time now spent upon crafting gorgeous, graphically elegant slide presentations can be recaptured and used for more important things. We can give back the time and energy now wasted on rehearsing one's time at the podium and relieve a major, unnecessary stressor for many team leaders. It won't matter whether the presenter is a great salesperson, a complete introvert, a new hire out of college, or a VP with 20 years of experience; what matters will be found on the page.

Last, the narrative document is infinitely portable and scalable. It is easy to circulate. Anyone can read it at any time. You don't need handwritten notes or a vocal track recorded during the big presentation to understand its contents. Anyone can edit or make comments on the document, and they are easily shared in the cloud. The document serves as its own record.

The Readers' Advantage: Information Density and Interconnection of Ideas

One useful metric for comparison is what we call the *Narrative Information Multiplier* (tip of the hat to former Amazon VP Jim Freeman for coining this term). A typical Word document, with text in Arial 11-point font, contains 3,000–4,000 characters per page. For comparison, we analyzed the last 50 S-Team PowerPoint slide presentations and found that they contained an average of just 440 characters per page. This means a written narrative would contain ***seven to nine times the information density*** of our typical PowerPoint presentation. If you take into account some of the other PowerPoint limitations discussed above, this multiplier only increases.

Tufte estimates that people read three times faster than the typical presenter can talk, meaning that they can absorb that much more information in a given time while reading a narrative than while listening to a PP presentation. A narrative therefore delivers much more information in a much shorter time.

The Narrative Information Multiplier is itself multiplied when one considers how many such meetings S-Team members attend in a single day. A switch to this denser format will allow key decision-makers to consume much more information in a given period of time than with the PowerPoint approach.

Narratives also allow for nonlinear, interconnected arguments to unfold naturally—something that the rigid linearity of PP does not permit. Such interconnectedness defines many of our most important business opportunities. Moreover, better-informed people make higher-quality decisions, and can deliver better, more detailed feedback on the presenting teams' tactical and strategic plans. If our executives are better informed, at a deeper level, on a wider array of important company initiatives, we will gain a substantial competitive advantage over executives elsewhere who rely on traditional low-bandwidth methods of communication (e.g., PP).

The Presenters' Advantage: Forces Greater Clarity of Thought

We know that writing narratives will likely prove to be harder work than creating the PP presentations that they will replace; this is actually positive. The act of writing will force the writer to think and synthesize more deeply than they would in the act of crafting a PP deck; the idea on paper will be better thought out, especially after the author's entire team has reviewed it and offered feedback. It's a daunting task to get all the relevant facts and all one's salient arguments into a coherent, understandable document—**and it should be.**

Our goal as presenters is not to merely introduce an idea but to demonstrate that it's been carefully weighed and thoroughly analyzed. Unlike a PP deck, a solid narrative can—and must—demonstrate how its many, often disparate, facts and analyses are interconnected. While an ideal PP presentation can do this, experience has shown that they rarely do in practice.

A complete narrative should also anticipate the likely objections, concerns, and alternate points of view that we expect our team to deliver. Writers will be forced to anticipate smart questions, reasonable objections, even common misunderstandings—and to address them proactively in their narrative document. You simply cannot gloss over an important topic in a narrative presentation, especially when you know it's going to be dissected by an audience full of critical thinkers. While this may seem a bit intimidating at first, it merely reflects our long-standing commitment to thinking deeply and correctly about our opportunities.

The old essay-writing adage "State, support, conclude" forms the basis for putting a convincing argument forward. Successful narratives will connect the dots for the reader and thus create a persuasive argument, rather than presenting a disconnected stream of bullet points and graphics that leave the audience to do all the work. Writing persuasively requires and enforces clarity of thought that's even more vital when multiple teams collaborate on an idea. The narrative form demands that teams be in sync or, if they are not, that they clearly state in the document where they are not yet aligned.

Edward Tufte sums up the benefits of narratives over PP with his own blunt clarity: "PowerPoint becomes ugly and inaccurate because our thoughts are foolish, but the slovenliness of PowerPoint makes it easier for us to have foolish thoughts."

How to Conduct a Meeting in This New Format

Narratives would be distributed at the start of each meeting and read by all in attendance during the time normally taken up by the slide deck—approximately the first 20 minutes. Many will want to take notes, or annotate their copy, during this time. Once everybody signals their readiness, conversation about the document begins.

We know that people read complex information at the rough average of three minutes per page, which in turn defines the functional length of a written narrative as about six pages for a 60-minute meeting. Our recommendation is therefore that teams respect the six-page maximum. There will no doubt be times when it feels difficult to condense a complete presentation into this size, but the same limitation—which is really one of meeting lengths—faces PP presenters as well. We believe that six pages should be enough, but we will review over time and revise if necessary.

Conclusion

PowerPoint could only carry us so far, and we're thankful for its service, but the time has come to move on. Written narratives will convey our ideas in a deeper, stronger,

more capable fashion while adding a key additional benefit: they will act as a forcing function that shapes sharper, more complete analysis. Six-page narratives are also incredibly inclusive communication, precisely because the interaction between the presenter and audience is zero during reading. No biases matter other than the clarity of reasoning. This change will strengthen not just the pitch, but the product—and the company—as well.

FAQ

Q: Most other companies of our size use PowerPoint. Why do we need to be different, and what if this switch turns out to be the wrong move?

A: In simplest terms, we see a better way. Amazon differs from other major companies in ways that help us stand out, including our willingness to go where the data lead and seek better ways of doing familiar things. If this move doesn't work out, we'll do what we always do—iterate and refine, or roll it back entirely if that's what the results show us is best.

Q: Why not distribute the narrative ahead of the meeting so we're ready?

A: The short time between distribution and the meeting might not give all attendees sufficient time for that task. Also, since the document replaces the deck, no time is lost by dedicating this phase of the meeting to a silent reading that brings everybody up to speed before Q&A begins. Last but certainly not least, this gives each presenting team the most possible time to complete and refine their presentation.

Q: My team has proven to be very good at PP presentations—do we HAVE to switch?

A: YES. One danger of an unusually strong PP presentation is that the stage presence or charm of the presenter can sometimes unintentionally blind the audience to key questions or concerns. Slick graphics can distract equally well. Most importantly, we've shown that even the best use of PP simply cannot deliver the completeness and sophistication that narratives can.

Q: What if we put our PP deck into printed form and add some extended comments to strengthen and extend the information content?

A: NO. Reproducing PP on paper also reproduces its weaknesses. There's nothing one can do in PP that cannot be done more thoroughly, though sometimes less attractively, in a narrative.

Q: Can we still use graphs or charts in our narratives?

A: YES. Most complex issues derive key insights from data and we expect that some of that data may be best represented in the form of a chart or graph. However, we do not expect that graphics alone can make the compelling and complete case we expect from a true written narrative. Include them if you must, but don't let graphics predominate.

Q: Six pages feels short. How much can we fit onto a page?

A: The six-page limit acts as a valuable forcing function that ensures we only discuss the most important issues. We also set aside 20 minutes for reading and expect that every attendee can read the entire thing during that time. Please don't fall prey to the temptation to fiddle with margins or font size to squeeze more into the document. Adding density to stay under the six-page limit works against this goal and tempts writers to stray into less important areas of consideration.

Q: How will we measure the success of this change?

A: Great question. We have not been able to identify a quantitative way to measure the quality of a series of S-Team decisions today, nor are we proposing a metric at this

time. Comparing the two approaches will be a qualitative exercise. We propose implementing narratives for the next three months and then polling the S-Team to ask if they're making better-informed decisions.

Six-Pagers Vary in Structure and Content

In the mock-up six-pager above, we've included two optional sections that many presenters at Amazon have found helpful. The first is to call out one or more key tenets that our proposal relies upon—a foundational element of the reasoning that led us to make this recommendation. Tenets give the reader an anchor point from which to evaluate the rest. If the tenet itself is in dispute, it's easier to address that directly rather than take on all the logical steps that derive from that position.

The second optional section, perhaps more commonly used, is the inclusion of an FAQ. Strong six-pagers don't just make their case, they anticipate counterarguments, points of contention, or statements that might be easily misinterpreted. Adding the FAQ to address these saves time and gives the reader a useful focal point for checking the thoroughness of the authors' thinking. (See appendix B for additional FAQ and tenet examples.)

We should also note that some six-pagers are longer than six pages, because they include supporting data or documentation in appendices—data that's not usually read during the meeting.

Six-page narratives can take many forms. Our mock-up provides one example, laid out specifically for our topic. We wouldn't typically expect to see a section titled "Our Inspiration," for instance, even though it serves a useful purpose in this narrative. Headings and subheadings, graphs or data tables, and other design elements will be specific to the individual narrative.

An Amazon quarterly business review, for instance, might be broken down like this instead:

Introduction
Tenets
Accomplishments
Misses
Proposals for Next Period
Headcount
P&L
FAQ
Appendices (includes things like supporting data in the form of spreadsheets, tables and charts, mock-ups)

The six-pager can be used to explore any argument or idea you want to present to a group of people—an investment, a potential acquisition, a new product or feature, a monthly or quarterly business update, an operating plan, or even an idea on how to improve the food at the company cafeteria. It takes practice to master the discipline of writing these narratives. First-time writers will do well to review and learn from successful examples.

The New Meeting Format

When the meeting topic is covered by a narrative, it works best if the entire audience reads the narrative, in the room, at the beginning of the meeting. The silence can be unsettling at first, but after you've been through the process a few times, it becomes routine. Even though you cannot hear it, with a well-written narrative there is a massive amount of useful information that is being transferred in those 20 minutes.

We mentioned earlier the estimated reading speed of three minutes per page, which led to the six-page limit. If yours is a 30-minute meeting, a three-page narrative would therefore be more appropriate. Our goal has been to leave two-thirds of the meeting time for discussing what we've read.

Still, people read at different speeds. Some will review the appendices, some won't. Some attendees will make comments in a shared online document, like Bill does, so that all meeting participants can see everyone's comments. I (Colin) prefer the old-fashioned way, making comments on paper so I can lose myself in the document. This also helps me avoid the confirmation bias that might arise were I to read the real-time comments others were adding to the shared document. Besides, I know I'll hear everybody's view soon enough.

When everyone has read the document, the presenter takes the floor. First-time presenters often start by saying, "Let me orally walk you through the document." **Resist that temptation**; it will likely be a waste of time. The whole point of the written document is to clearly present the reasoning and to avoid the hazards of live presentation. The attendees have already walked themselves through the argument.

Some groups at Amazon go around the room, ask for high-level feedback, then pore over the document line by line. Other groups ask a single individual to give all their feedback on the entire document, then ask the next person in the audience to do the same. Just pick a method that works for you—there's no single correct approach.

Then the discussion begins, which essentially means that the audience members ask questions of the presenting team. They seek clarification, probe intentions, offer insights, and suggest refinements or alternatives. The

presenting team has put great care and thought into the narrative, and the audience members have a responsibility to take it seriously. The key goal of the meeting, after all, is to seek the truth about the proposed idea or topic. We want that idea to become the best it can possibly be as a result of any adjustments we make along with the presenting team.

During the discussion stage, it's also important that notes be taken on behalf of the entire audience, preferably by someone knowledgeable about the subject who is not the primary presenter. The presenter is generally too involved in answering questions to capture effective notes at the same time. If I don't see anyone taking notes at the discussion stage, I will politely pause the meeting and ask who is going to do so. It's vital that we capture and record the salient points of the ensuing discussion, as those comments become part of the output of the narrative process.

Feedback as Collaboration

Providing valuable feedback and insight can prove to be as difficult as writing the narrative itself. Two of the most cherished gifts I (Colin) received in my career are pens, given to me by people whose narratives I had read and commented on. (I would typically give a printout of the narrative with my handwritten notes on it to the presenters after the meeting.) Both people told me that my comments had played a key role in making their businesses successful. I say this not to boast but to provide evidence that when the reader takes the narrative process just as seriously as the writer does, the comments can have real, significant, and long-lasting impact. You are not just commenting on a document, you're helping to shape an idea, and thereby becoming a key team member for that business.

Because examples of excellent six-page narratives are disseminated throughout the company, and because expectations about their nature and quality are so well understood by employees, it rarely happens that a team presents a substandard narrative at a meeting. I did once receive a six-pager that was not up to snuff. The team who wrote it was glossing over hard problems with platitudes. I politely handed it back to them, said it wasn't ready to be discussed, adjourned the meeting, and suggested they use the time to work on improving the narrative. But, as I said, those scenarios are extremely rare. Mostly it's about supporting the team by giving robust feedback. Jeff has an uncanny ability to read a narrative and consistently arrive at insights that no one else did, even though we were all reading the same narrative. After one meeting, I asked him how he was able to do that. He responded with a simple and useful tip that I have not forgotten: he assumes each sentence he reads is wrong until he can prove otherwise. He's challenging the content of the sentence, not the motive of the writer. Jeff, by the way, was usually among the last to finish reading.

This approach to critical thinking challenges the team to question whether the current narrative has it right or if there are additional fundamental truths to uncover, and if they are aligned with the Amazon Leadership Principles. For example, say a narrative reads, "Our customer-friendly returns policy allows returns up to 60 days from the time of purchase compared to the 30 days typically offered by our competitors." A

busy executive doing a cursory read and already thinking about their next meeting may be content with that statement and move on. However, a critical reader would challenge the implicit assumption being made, namely, that the longer allowable return duration makes the policy customer friendly. The policy may be better than a competitor's, but is it *actually* customer friendly? Then during the discussion, the critical reader may ask, "If Amazon is really customer obsessed, why do we penalize the 99 percent of customers who are honest and want to return an item by making them wait until our returns department receives the item to make sure it's the right item and that it's not damaged?" This type of thinking—in which you assume there is something wrong with the sentence—led Amazon to create the no-hassle return policy, which specifies that the customer should get a refund even before Amazon receives the returned goods. (The refund is reversed for the small percentage of people who do not send back the item.) Here is another instance where you don't need to "have a Jeff" in order to apply this exacting style of critical thinking to ideas at your company.

Final Thoughts About Narratives

Narratives are designed to increase the quantity and quality of effective communication in your organization—by an order of magnitude over traditional methods. Creating such solid narratives requires hard work and some risk-taking. Good ones take many days to write. The team writing the narrative toils over the topic, writes its first draft, circulates and reviews and iterates and repeats, then finally takes the vulnerable step of saying to their management and their peers, "Here's our best effort. Tell us where we fell short." At first this openness can prove intimidating.

But as we've seen, this model imposes duties and expectations upon the audience as well. They must objectively and thoroughly evaluate the idea, not the team or the pitch, and suggest ways to improve it. The work product of the meeting is ultimately a joint effort of the presenter and their audience—thinking that they can all stand behind. Silence in the discussion stage is the equivalent of agreement with what is presented, but it carries the same weight as a full-blown critique.

In this way, the presenter and audience become integrally linked to the subsequent success or failure of the initiative, or the correctness or incorrectness of a team's business analysis. When looking at any of Amazon's big wins, remember that every major success has gone through multiple narrative reviews; it's likely there were meaningful contributions from the audience as well as the team. On the other hand, for every failed initiative or analysis that fell short, there were senior leaders who looked at it and thought, "This makes sense," or, "Yes, this should work." Either way, if the narrative process works to its fullest potential, you're all in it together.

5

Working Backwards

Start with the Desired Customer Experience

Start with the customer and work backwards—harder than it sounds, but a clear path to innovating and delighting customers. A useful Working Backwards tool: writing the press release and FAQ before you build the product.

* * *

Most of Amazon's major products and initiatives since 2004 have one very Amazonian thing in common—they were created through a process called Working Backwards. It is so central to the company's success that we used it as the title for our book. Working Backwards is a systematic way to vet ideas and create new products. Its key tenet is to start by defining the customer experience, then iteratively work backwards from that point until the team achieves clarity of thought around what to build. Its principal tool is a second form of written narrative called the PR/FAQ, short for press release/frequently asked questions.

We both witnessed its birth. Colin was in his tenure as Jeff's shadow when the Working Backwards process was launched and he participated in every Working Backwards review presented to Jeff in the twelve months thereafter. And Bill's experience was forged by applying and refining the Working Backwards concept in the early stages of the process that led to the development of every digital media product.

Trial and Error, Then Success

Working as Jeff's shadow was a bit like drinking from a fire hose. One surprising challenge of the job I (Colin) noticed early on was just how much context switching went on each day. Every week Jeff—and therefore I—had three recurring meetings: the four-hour S-Team meeting discussed in the previous chapter, a Weekly Business Review (chapter six), and an informal Monday-morning S-Team breakfast near the office. In addition to those, on any given day we'd usually meet with two to four product teams, where we'd spend between one and two hours doing a deep dive on new products and features. Throw in the occasional retail, finance, and operations updates, plus a fire drill or two requiring immediate attention, and you have a typical week.

The product team meetings usually took up a plurality of the available hours in the week. Jeff and I would need to get up to speed on where we left off with any given team, so the first part of each product meeting could be viewed as setup cost. Then we'd discuss the progress made since our last meeting, ask and answer questions, discuss new issues or problems, and agree on next steps that needed to be addressed before we met with the team again. Despite everyone's best intentions, the meetings were often error prone and inefficient. Sometimes the "setup time" would consume too much of the meeting: teams, rightfully proud of their recent accomplishments, wanted to talk about them at the expense of the important decisions we needed to know about, so by the time the team recapped their progress, there was not enough time left for what actually needed to get done. Other times we'd discover, too late, that the team was not aligned with Jeff and had veered off path from the previous meeting. When that happened, it was extremely frustrating for everyone, not to mention a waste of valuable time.

As I mentioned earlier, part of my role as Jeff's shadow was to help him be as effective as possible. We needed to improve each stage of these product meetings. We needed to begin by quickly and accurately caching the right information during the setup portion of the meeting. Then we had to focus on the most important issues moving forward. Finally, we had to map out a clear trajectory for the teams to follow between the current

meeting and the next one. If we could do all that, it would be a huge win for everyone. We'd be more efficient in addressing the hard problems, which would help us make higher-quality decisions more quickly. With that increase in the speed of good decision-making, Jeff would be able to connect deeply with a greater number of teams.

While I was trying to help sort all this out, Jeff was spending a disproportionate amount of his time on Amazon's digital transformation, and on what would eventually become the first set of Amazon's cloud computing services.

So my goal was not easy to reach. It required trial and error over the course of many months. Jeff tried many different ideas, some of them seemingly crazy, like starting a project proposal by writing a user manual or a technical API guide, relying solely on mock-ups, and other approaches to visualizing the outcome of a project. I remember getting frantic calls from nontechnical product managers saying, "Colin, I'm supposed to meet with Jeff next week. Can you send me a good example of a user manual? Also, I'm supposed to write something called an API guide but have no idea what that is!" We were not committed to any of these experimental formats and stopped using them when we realized they were counterproductive.

In the end, what turned out to work best was relying on the core Amazon principle of customer obsession and a simple yet flexible way of writing narrative documents. These two elements form the Working Backwards process—starting from the customer experience and working backwards from that by writing a press release that literally announces the product as if it were ready to launch and an FAQ anticipating the tough questions. While this next section describes the evolution of Working Backwards as seen through the experience of the digital team, a handful of other teams went through a similar process. Bringing together the experience of these teams enabled us to hone and refine Working Backwards into its final form.

Where Are the Mock-Ups? Bill and the Launch of Digital

In 2004, I (Bill) was one of the leaders selected to create and lead Amazon's digital media organization. I was itching to launch new stores for digital music, movies, and TV shows. I also needed to revamp our e-book store, which had gone online in 2000 and was then a tiny business because books could only be read on a PC and were more expensive than the print edition.

I assumed that the launch process for digital media would essentially be the same as it was for other new Amazon businesses—toys, electronics, and tools, for example—which were known as "category expansions." For those launches, the process had been straightforward. The team would gather the data to build a catalog of items, establish relationships with vendors to source them, set prices, build content for category pages, and then launch. It wasn't easy, but we weren't inventing a new store or customer experience from scratch.

As I was to learn, the process for creating the digital media business would be quite different because there was so much more to creating a great digital media customer experience than simply adding the next retail category to the Amazon website.

The first part of the process went as normal. Our team of three or four people developed plans using the tried-and-true MBA-style methods of the time. We gathered data about the size of the market opportunity. We constructed financial models projecting our annual sales in each category, assuming, of course, an ever-increasing share of digital sales. We calculated gross margin assuming a certain cost of goods from our suppliers. We projected an operating margin based on the size of the team we would need to support the business. We outlined the deals we would make with media companies. We sketched out pricing parameters. We described how the service would work for customers. We put it all together in crisp-looking PowerPoint slides (this was still several months before the switch to narratives) and comprehensive Excel spreadsheets.

We had several meetings with Jeff to present our ideas. At each one, he would listen carefully to what we had to say. He would ask probing questions and study the financials. But he never seemed satisfied or convinced. He found our proposals light on the details as to how the service

would work for customers. Finally, inevitably, he would ask, "Where are the mock-ups?"

Jeff was referring to the visual representations that would show exactly how the new service would look on the Amazon website. Mock-ups should be detailed, showing the entire customer experience from landing page to purchase—screen design, buttons, text, the sequence of clicks, everything. To create a meaningful and informative mock-up you have to think through every element of what the service will offer, what the experience will be for the customer, how all the features will work on the page. It requires a ton of work to think through the whole business and a ton more work to create and refine the visuals.

We didn't have any mock-ups. We just wanted to sell Jeff on the opportunity, show him that these digital media businesses could be large, set a budget, and get the green light to start building the team. We would deal with the customer experience and other details once we got his go-ahead.

But if Jeff wants to see mock-ups, you had better make mock-ups.

A few weeks later we were back with rough mock-ups in hand. Jeff listened carefully to our presentation and then began asking detailed questions about every button, word, link, and color. For music, he asked how our service would be better than iTunes. For e-books, he wanted to know how much the e-books would cost. He asked if people would be able to read their e-books on a tablet or a phone as well as their PC.

We answered as we had before. We hadn't figured out all that stuff! We just needed his basic approval so we could hire the team, start negotiating deals with media companies, and get something launched. That answer did not go over well. At all. Jeff wanted to know exactly what we were going to build and how it would be better for customers than the competition. He wanted us to agree on those details before we started hiring a team or establishing vendor relationships or building anything.

It was clear that half-baked mock-ups were no better, perhaps worse, than no mock-ups at all. To Jeff, a half-baked mock-up was evidence of half-baked thinking. And he was quick to say so, often using strong language to make his point inescapably clear. Jeff wanted us to know that we couldn't just charge down the first available and most convenient path to chase after this opportunity. We needed to think through our plan in detail.

We went back to work. The deeper we dug, the clearer it became that digital media was going to be unlike any other Amazon business. The obvious difference was that we would not be shipping brown boxes to customers but rather delivering digital bits over wires. That was the least complicated part. There also had to be a great way for the customer to manage, read, listen to, or watch those bits once they had them. This would require custom apps and hardware.

As we continued to meet with Jeff, we tried various kinds of spreadsheets and PowerPoint slides to present and explore our ideas, none of which seemed to be particularly effective. At some point, I don't remember exactly when, Jeff suggested a different approach for the next meeting. Forget the spreadsheets and slides, he said. Instead, each team member would write a narrative document. In it, they would describe their best idea for a device or service for the digital media business.

The next meeting arrived, and we all showed up with our narratives. (As mentioned, ours was one of several teams involved in the early experimentation with narratives at the company. They were not yet official Amazon policy.) We distributed them and read them to ourselves and then discussed them, one after another. One proposed an e-book reader that would use new E Ink screen technology. Another described a new take on the MP3 player. Jeff wrote his own narrative about a device he called the Amazon Puck. It would sit on your countertop and could respond to voice commands like, "Puck. Please order a gallon of milk." Puck would then place the order with Amazon.

The great revelation of this process was not any one of the product ideas. As we've described in chapter four, the breakthrough was the document itself. We had freed ourselves of the quantitative demands of Excel, the visual seduction of PowerPoint, and the distracting effect of personal performance. The idea had to be in the writing.

Writing up our ideas was hard work. It required us to be thorough and precise. We had to describe features, pricing, how the service would work, why consumers would want it. Half-baked thinking was harder to disguise on the written page than in PowerPoint slides. It could not be glossed over through personal charm in the presentation.

After we started using the documents, our meetings changed. There was more meat and more detail to discuss, so the sessions were livelier and

longer. We weren't so focused on the pro forma P&L and projected market segment share. We talked at length about the service itself, the experience, and which products and services we thought would appeal most to the customer.

After a lot of trial and error, and incremental moves in this direction among many teams involved in the narrative experiment, Jeff then pushed the idea further. What if we thought of the product concept narrative as a press release? Usually, in a conventional organization, a press release is written at the end of the product development process. The engineers and product managers finish their work, then "throw it over the wall" to the marketing and sales people, who look at the product from the customer point of view, often for the first time. They're the ones who write the press release, which describes the killer features and fantastic benefits and is designed to create buzz, capture attention, and, above all, get customers to leap out of their chairs to buy.

In this standard process, the company works forward. The leaders come up with a product or business that is great for the company, and then they try to shoehorn it into meeting previously unmet customer needs.

That approach can lead to some undesirable results, Jeff believed. To make his point, he used Sony as a hypothetical example. Suppose Sony decides to introduce a new TV. The sales and marketing group has done its research into customer preferences and market trends (but not necessarily the customer experience) and has determined that Sony should offer a 44-inch TV at a price point of $1,999. The engineering team, however, has been working on the new TV for quite some time, and their focus has been on picture quality, which means higher resolution, and they have not been especially concerned about price point. The TV they come up with will cost $2,000 just to manufacture. So there is no way that the retail price can be $1,999.

If the two organizations had started the process by writing a press release, they would have had to agree on the features, cost, customer experience, and price. Then they could have worked backwards to figure out what to build, thereby surfacing the challenges they would face in product development and manufacturing.

The Kindle Press Release

Kindle was the first product offered by the digital media group, and it, along with several AWS products, was among the first at Amazon to be created using the press release approach.

Kindle was a breakthrough in multiple dimensions. It used an E Ink display. The customer could shop for, buy, and download books directly from the device—no need to connect to a PC or to Wi-Fi. Kindle offered more e-books than any other device or service available at the time and the price was lower. Today, that set of features sounds absolutely standard. In 2007, it was pioneering.

But Kindle had not started out that way. In the early stages of its development—before we got started on the press release approach and when we were still using PowerPoint and Excel—we had not described a device that could do all these things from the customer perspective. We had focused on the technology challenges, business constraints, sales and financial projections, and marketing opportunities. We were working forward, trying to invent a product that would be good for Amazon, the company, not the customer.

When we wrote a Kindle press release and started working backwards, everything changed. We focused instead on what would be great for customers. An excellent screen for a great reading experience. An ordering process that would make buying and downloading books easy. A huge selection of titles. Low prices. We would never have had the breakthroughs necessary to achieve that customer experience were it not for the press release process, which forced the team to invent multiple solutions to customer problems. (We tell the whole Kindle story in chapter seven.)

As we got more adept at using the Working Backwards process, we refined the press release document and added a second element: the FAQ, frequently asked questions, with, of course, answers.

The FAQ section, as it developed, included both external and internal questions. External FAQs are the ones you would expect to hear from the press or customers. “Where can I purchase a new Amazon Echo?” or “How does Alexa work?”

Internal FAQs are the questions that your team and the executive leadership will ask. “How can we make a 44-inch TV with an HD display that can retail for $1,999 at a 25 percent gross margin?” or “How will we make a Kindle reader that connects to carrier networks to download books without customers having to sign a contract with a carrier?” or “How many new software engineers and data scientists do we need to hire for this new initiative?”

In other words, the FAQ section is where the writer shares the details of the plan from a consumer point of view and addresses the various risks and challenges from internal operations, technical, product, marketing, legal, business development, and financial points of view.

The Working Backwards document became known as the PR/FAQ.

The Features and Benefits of the PR/FAQ

The primary point of the process is to shift from an internal/company perspective to a customer perspective. Customers are pitched new products constantly. Why will this new product be compelling enough for customers to take action and buy it? A common question asked by executives when reviewing the product features in the PR is "so what?" If the press release doesn't describe a product that is meaningfully better (faster, easier, cheaper) than what is already out there, or results in some stepwise change in customer experience, then it isn't worth building.

The PR gives the reader the highlights of the customer experience. The FAQ provides all the salient details of the customer experience as well as a clear-eyed and thorough assessment of how expensive and challenging it will be for the company to build the product or create the service. That's why it's not unusual for an Amazon team to write ten drafts of the PR/FAQ or more, and to meet with their senior leaders five times or more to iterate, debate, and refine the idea.

The PR/FAQ process creates a framework for rapidly iterating and incorporating feedback and reinforces a detailed, data-oriented, and fact-based method of decision-making. We found that it can be used to develop ideas and initiatives—a new compensation policy, for example—as well as products and services. Once your organization learns how to use this valuable tool, it is addicting. People start to use it for everything.

Over time, we refined and normalized the specifications for the PR/FAQ. The press release (PR) portion is a few paragraphs, always less than one page. The frequently asked questions (FAQ) should be five pages or less. There are no awards for extra pages or more words. The goal isn't to explain all the excellent work you have done but rather to share the distilled thinking that has come from that work.

People who write press releases for a living, or indeed anyone who has been professionally edited, knows the importance of boiling things down as much as possible, but the people in product development don't always understand this. In the early days of the PR/FAQ, a common mistake people made was to assume that more means better. They'd produce long documents, attach page after page of narrative, insert charts and tables in an

appendix. The virtue of this approach, at least from the perspective of the writer, is that it shows all their work and allows them to avoid hard decisions about what's important and what's not—leaving those for the group. However, restricting the length of the document is, to use a term that came up when describing the narratives, a forcing function—we have seen that it develops better thinkers and communicators.

The creation of the PR/FAQ starts with the person who originated either the idea or the project writing a draft. When it's in shareable condition, that person sets up a one-hour meeting with stakeholders to review the document and get feedback. At the meeting, they distribute the PR/FAQ in either soft or hard copy, and everyone reads it to themselves. When they have finished, the writer asks for general feedback. The most senior attendees tend to speak last, to avoid influencing others.

Once everyone has given their high-level responses, the writer asks for specific comments, line by line, paragraph by paragraph. This discussion of the details is the critical part of the meeting. People ask hard questions. They engage in intense debate and discussion of the key ideas and the way they are expressed. They point out things that should be omitted or things that are missing.

After the meeting, the writer distributes meeting minutes to all the attendees, including notes on the feedback. Then they get to work on the revision, incorporating responses to the feedback. When it is polished, they present it to the executive leaders in the company. There will be more feedback and discussion. More revision and more meetings may be required.

The PR/FAQ review process can be stressful, no matter how constructive and unbiased the feedback. Gaps will be found! A PR/FAQ under serious consideration for implementation will typically require multiple drafts and meetings with the leadership. Senior managers, directors, and executive leaders who oversee the authors of PR/FAQs become skilled evaluators and contributors to the process. The more PR/FAQs they read, and the more products they build and launch using the PR/FAQ process, the more capable they become at identifying the omissions and flaws in the author's thinking. And so the process itself creates a tier of master evaluators as it vets and strengthens the idea and aligns everyone involved in the project, from individual contributor to

CEO. It also increases the likelihood that a project will be approved and funded. You should plan on making many revisions to the PR/FAQ document, even after the project has formally started, to reflect changes and new elements.

Example: Blue Corp. Announces the Launch of Melinda, the Smart Mailbox

Melinda is the physical mailbox designed to securely receive and keep safe all your e-commerce and grocery deliveries.

PR Newswire, Atlanta, GA, November 5, 2019

Today Blue Corp. announced the launch of Melinda, a smart mailbox that ensures secure and properly chilled delivery and storage for your online purchases and groceries. With Melinda, you no longer need to worry about getting your deliveries stolen from your doorstep or spoiled groceries. Plus, you're notified as soon as your packages are delivered. Packed with smart technology, Melinda costs just $299.

Today, 23 percent of online shoppers report having packages stolen from their front porch, and 19 percent complain of grocery deliveries being spoiled. With no easy solution to these problems, customers give up and stop ordering online.

Melinda, with its smart technology and insulation, makes stolen packages and spoiled groceries a thing of the past. Each Melinda includes a camera and a speaker. When a delivery courier arrives at your home, Melinda tells the courier to scan the package barcode by holding it up to the camera. If the code is valid, the front door opens and Melinda instructs the courier to place the package inside and close the door securely. The built-in scale in the base of each Melinda verifies that the weight of the delivery matches the weight of the item(s) you ordered. The courier receives a voice confirmation, and your purchase is safe and secure. Melinda sends you a text letting you know that your item arrived along with a video of the courier making the delivery.

When you return home and are ready to retrieve your delivery, just use the built-in fingerprint reader to unlock the door. Melinda can store and recognize up to ten saved fingerprints so that all members of your family can access Melinda.

Do you use Instacart, Amazon, or Walmart for online grocery delivery? If so, are you tired of spoiled groceries in the hot sun? Melinda keeps your chilled and frozen food cold. The walls of Melinda are two inches thick and made with the same pressure-injected foam used in the best coolers, keeping your groceries cool for up to twelve hours.

Melinda fits easily on your porch or stoop, taking up just a few feet of space, and you can choose from a variety of colors and finishes to make Melinda an attractive addition to the appearance of your home.

"Melinda is a breakthrough in safety and convenience for online shoppers," says Lisa Morris, CEO of Blue Corp. "In creating Melinda we combined a number of the latest technologies at the low price of just $299."

"Melinda is a lifesaver," said Janet Thomas, a frequent online shopper and customer of Instacart. "It is so frustrating when one of my packages is stolen from my front porch, and it can be time-consuming to work with customer support to get a refund. I use Instacart every week for grocery delivery, and many times I am not home when my groceries arrive. I love knowing that they are kept cool and secure in my Melinda. I selected the natural teak finish for my Melinda—it looks great on my front porch."

To order your Melinda, simply visit keepitcoolmelinda.com, or visit amazon.com, walmart.com, Walmart stores, and other leading retailers.

Internal FAQs

Q: How large is the estimated consumer demand for Melinda?

A: Based on our research, we estimate that ten million households in the United States, Europe, and Asia would want to buy Melinda at a $299 price point.

Q: Why is $299 the right price point?

A: There are no directly comparable products in the marketplace today. One similar product is Amazon Key, which allows couriers access to your home, garage, or car using smart lock technology. Another similar product is Ring Doorbell, which ranges in price from $99 to $499. We based our price on customer surveys and focus groups combined with the price needed to ensure profitability.

Q: How does Melinda recognize barcodes on packages?

A: We will license barcode-scanning technology from Green Corp. at a cost of $100K per year. In addition, we need to develop an API that will allow us to link a Melinda customer account with any e-commerce provider (Amazon, Walmart, eBay, OfferUp, etc.), which provides us with the item tracking number from the e-commerce or delivery merchant. This way we can recognize the barcode with the package tracking number and know either the exact or an estimated weight for each item.

Q: What if a customer receives an order from an e-commerce provider and they haven't linked their account yet?

A: We make it easy for customers to link their orders because we will offer a browser plug-in for Melinda customers that detects when they place an order with an e-commerce provider, which then links their account and the order details to their Melinda.

Q: Why will e-commerce providers like Amazon and Walmart be willing to share these package delivery details with us? What is in it for them?

A: We believe we can convince them that the customer experience benefits will enable them to increase their sales. In addition, we will work closely with their business and legal groups to ensure that we handle their customer data in ways that meet their stringent requirements. Alternatively, we will offer a simple UI for customers to copy and paste each tracking number from their e-commerce provider to the Melinda app.

Q: What happens if a customer gets more than one delivery in a day?

A: Melinda can accept multiple deliveries each day until the unit is full.

Q: What if the package is too big for Melinda?

A: Packages exceeding 2′x2′x4′ won't fit in Melinda. Melinda can still record the delivery person and scan the barcode, but the item is stored outside Melinda.

Q: How does Melinda prevent a courier from stealing items that are already in Melinda from a prior order?

A: There are several ways. The first is that the forward-facing camera records any activity or access to Melinda. The second is that there is a scale at the base of the unit

that detects the weight of the shipment and verifies that this matches the item(s) ordered. If a second delivery is made in one day, Melinda knows the weight of the first delivery and the estimated weight of the second delivery, so if the net weight is lower, Melinda knows that the courier has removed something and will sound an alarm.

Q: What is the estimated bill of materials (BOM) or cost to manufacture each Melinda, and how much profit will we make per unit?

A: The estimated BOM is $250 for each Melinda, meaning that our gross profit per unit is $49. The most expensive parts in Melinda are the shell and insulation ($115), the fingerprint reader ($49), and the scale.

Q: What is the power source for Melinda?

A: Melinda requires a standard AC outlet.

Q: What size team is required to build Melinda?

A: We estimate that we need a team of 77 at an annualized cost of $15 million. There are several teams required to build Melinda, but these can be broken down into hardware and software teams. On the hardware side, we need a team for each of the following:

- The physical shell, color choices, and finishes (6)
- Integration of the various smart and mechanical components, including the fingerprint reader, the camera, the automatic (open/close) door, the speaker, and the camera (12)

On the software side, we will need a team for each of the new services. Below is our current assessment of what teams will be required and how many people should be on each team, including product managers, engineers, designers, and so on:

- Voice commands to couriers (10)
- Fingerprint capture and storing (8)
- Package tracking and item weight details (11)
- Barcode reader (7)
- API to link e-commerce accounts to Melinda (12)
- Browser plug-in/web interface for account linking (5)
- Melinda app for iOS and Android (6)

★ ★ ★

This fictitious PR/FAQ is designed to illustrate the kinds of thinking and problems that the author and readers of a PR/FAQ should consider.

The product itself is both realistic and unrealistic. The customer problem of stolen packages and melting groceries is very real (although the research/stats here are phony), and the various components and technologies all exist. The Melinda, as described, is not realistic in that the costs are almost certainly underestimated (the product is overly complex), and the total addressable market for the product is probably very small.

However, the example enables us to illustrate the ways in which the PR/FAQ process helps authors assess the viability of any new product by

forcing them to consider and document all elements and constraints, including (but not limited to) the consumer needs and total addressable market, the per-unit economics and P&L, key dependencies, and the feasibility (how challenging it is to build the product). A good PR/FAQ is one in which the author has clearly considered and grappled with each of these issues, seeking truth and clarity on each.

Press Release Components

These are the key elements of the press release:

Heading: Name the product in a way the reader (i.e., your target customers) will understand. One sentence under the title.

"*Blue Corp. announces the launch of Melinda, the smart mailbox*."

Subheading: Describe the customer for the product and what benefits they will gain from using it. One sentence only underneath the heading.

"Melinda is the physical mailbox designed to securely receive and keep safe all your e-commerce and grocery deliveries."

Summary Paragraph: Begin with the city, media outlet, and your proposed launch date. Give a summary of the product and the benefit.

"PR Newswire, Atlanta, GA, November 5, 2019. Today Blue Corp. announced the launch of Melinda, a smart mailbox that ensures secure and properly chilled delivery and storage for your online purchases and groceries."

Problem Paragraph: This is where you describe the problem that your product is designed to solve. Make sure that you write this paragraph from the customer's point of view.

"Today, 23 percent of online shoppers report having packages stolen from their front porch, and 19 percent complain of grocery deliveries being spoiled."

Solution Paragraph(s): Describe your product in some detail and how it simply and easily solves the customer's problem. For more complex products, you may need more than one paragraph.

"With Melinda, you no longer need to worry about getting your online purchases and deliveries stolen…"

Quotes and Getting Started: Add one quote from you or your company's spokesperson and a second quote from a hypothetical customer in which they describe the benefit they are getting from using your new product. Describe how easy it is to get started, and provide a link to your website where customers can get more information and purchase the product.

"Melinda is a breakthrough in safety and convenience for online shoppers…"

FAQ Components

Unlike the PR, the FAQ section has a more free-form feel to it—there are no mandatory FAQs. The PR section does not typically include visuals, but it is more than appropriate to include tables, graphs, and charts in the FAQ. You must include things like your pro forma P&L for a new business or product. If you have high-quality mock-ups or wireframes, they can be included as an appendix.

Often FAQs are divided into external (customer focused) and internal (focused on your company). The external FAQs are those that customers and/or the press will ask you about the product. These will include more detailed questions about how the product works, how much it costs, and how/where to buy it. Because these questions are product specific, they are unique to an individual PR/FAQ. For internal FAQs, there is a more standardized list of topics you will need to cover. Here are some of the typical areas to address.

Consumer Needs and Total Addressable Market (TAM)

- How many consumers have this need or problem?
- How big is the need?
- For how many consumers is this problem big enough that they are willing to spend money to do something about it?
- If so, how much money would they be willing to spend?
- How many of these consumers have the characteristics/capabilities/constraints necessary to make use of the product?

These consumer questions will enable you to identify the core customers by filtering out those who don't meet the product constraints. In the case of Melinda, for example, you would eliminate people who:

- don't have enough space on their front porch for this product
- don't have a front porch or similar outdoor area with access to the street at all (e.g., most apartment dwellers)
- don't have a suitable source of electricity

- wouldn't be pleased to have a large storage/mailbox on their front porch
- don't receive many deliveries or deliveries that need refrigeration
- don't live in areas where package theft is a problem
- don't have interest or ability to pay $299 to answer the need

Only a discrete number of people will pass through all these filters and be identified as belonging to the total addressable market.

Research into these questions (e.g., how many detached homes are there in a given area?) can help you estimate the total addressable market (TAM), but like any research, there will be a wide error bar. The author and readers of the PR/FAQ will ultimately have to decide on the size of the TAM based on the data gathered and their judgment about its relevance. With Melinda, this process would likely lead to the conclusion that the TAM is in fact pretty small.

Economics and P&L

- What are the per-unit economics of the device? That is, what is the expected gross profit and contribution profit per unit?
- What is the rationale for the price point you have chosen for the product?
- How much will we have to invest up front to build this product in terms of people, technology, inventory, warehouse space, and so on?

For this section of the PR/FAQ, ideally one or more members of your finance team will work with you to understand and capture these costs so you can include a simplified table of the per-unit economics and a mini P&L in the document. A resourceful entrepreneur or product manager can do this work themselves if they do not have a finance manager or team.

For new products, the up-front investment is a major consideration. In the case of Melinda, there is a requirement for 77 people to work on the hardware and software, for an annualized cost of roughly $15 million. This means that the product idea needs to have the potential to earn well in excess of $15 million per year in gross profit to be worth building.

The consumer questions and economic analysis both have an effect on the product price point, and that price point, in turn, has an effect on the size of the total addressable market.

Price is a key variable in the authoring of your PR/FAQ. There may be special assumptions or considerations that have informed your calculation of the price point—perhaps making it relatively low or unexpectedly high—that need to be called out and explained. Some of the best new product proposals set a not-to-exceed price point because it forces the team to innovate within that constraint and face the tough trade-offs early on. The problem(s) associated with achieving that price point should be fully explained and explored in the FAQ. Suppose your research into Melinda leads you to conclude that to realize the largest possible TAM, you need to offer the product at no more than $99. The bill of materials (BOM), however, comes to $250. Now you have two choices to suggest. First, alter the specs, strip out features, or take other actions that will reduce the BOM to below $99. Second, construct a financial plan that shows heavy losses in the early days of release, but also shows that the losses can eventually be mitigated with BOM reductions as the product achieves scale or can be enhanced with some additional source of revenue (e.g., an associated service or subscription).

Dependencies

- How will we convince couriers (USPS, UPS, FedEx, Amazon Fulfillment, Instacart, etc.) to actually use this device instead of their current/standard delivery methods?
- How will we ensure that couriers (who don't work for you and over whom you have no control) will use the Melinda UI properly and bother to actually put packages in it instead of just leaving the package by the front door like they typically do?
- Won't it take more time (which is precious) for them to make a delivery than it does today?
- What third-party technologies are we dependent on for Melinda to function as promised?

A common mistake among less-seasoned product managers is to not fully consider how third parties who have their own agendas and incentives

will interact with their product idea, or what potential regulatory or legal issues might arise.

The role of third parties is a major issue with Melinda, whose success largely depends on their involvement and proper execution. Without the correct package tracking data or the cooperation of the companies that own that data and the couriers who deliver the packages, Melinda (as described) would be useless. The only alternative would be for customers to manually enter their tracking information for every single delivery into the Melinda app, which they are unlikely to do—and even if they did, it would still require couriers to be willing and able to use it. A good PR/FAQ honestly and accurately assesses these dependencies and describes the specific concepts or plans for the product to solve them.

Feasibility

- What are the challenging product engineering problems we will need to solve?
- What are the challenging customer UI problems we will need to solve?
- What are the third-party dependencies we will need to solve?
- How will we manage the risk of the up-front investment required?

These questions are intended to help the author clarify to the reader what level of invention is required and what kind of challenges are involved in building this new product. These criteria vary from product to product, and there are different types of challenges ranging from technical to legal to financial to third-party partnerships and customer UI or acceptance.

With Melinda, the engineering challenges are probably quite manageable, since no new technologies need to be developed or employed. The user interface is also familiar. The third-party dependencies present the greatest challenge to making Melinda work.

Go Ahead?

It is important to note that, during our time with Amazon, most PR/FAQs never made it to a stage where they were launched as actual products. What this means is that a product manager will put in a lot of time exploring product ideas that never get to market. This may be because of the intense competition for resources and capital among the hundreds of PR/FAQs that are authored and presented each year within the company. Only the very best will rise to the top of the stack and get prioritized and resourced, whether the pool of capital comes from within a large company like Amazon or from a startup investor. The fact that most PR/FAQs don't get approved is a feature, not a bug. Spending time up front to think through all the details of a product, and to determine—without committing precious software development resources—which products not to build, preserves your company's resources to build products that will yield the highest impact for customers and your business.

Another one of the biggest benefits of a written PR/FAQ is that it enables the team to truly understand the specific constraints and problems that would prevent a new product idea from being viable and aligning on them. At that point, the product or leadership team must decide if they will keep working on the product, addressing the problems and constraints surfaced by the PR/FAQ and developing solutions that will potentially make the product viable, or if they will set it aside.

In the case of Melinda, the author and team would certainly come to the conclusion that this isn't a viable product for many reasons. The TAM may be just too small, no matter what the product price point. The product might be too bothersome to use, even if the functionality itself is familiar to most customers. It may be unrealistic that Amazon and Walmart would provide a data feed or that couriers would bother to use the product. The device may simply be too expensive to build and profitably sell for $299, no matter how big the TAM might grow to be.

This process enables a product team and the company leadership to gain a thorough understanding of the opportunity and the constraints. Leadership and management are often about deciding what *not* to do rather

than what *to* do. Bringing clarity to why you aren't doing something is often as important as having clarity about what you are doing.

If, after the PR/FAQ process, the leadership team still believes in the product and wants it to become a reality, the process will have given them a thorough understanding of the problems that would need to be solved in order to move forward with it. Perhaps a problem can be solved through an acquisition or a partnership. Perhaps it can be solved with the passage of time—new technologies may become available, or the costs of the technology might come down. Perhaps the company decides that the problem or constraint is solvable, that the solution will require risk and cost, and that they are willing to assume that risk and cost because the TAM is large and therefore the potential rewards are great.

This last consideration came up frequently in reviews with Jeff, as we would wrestle with product ideas using the PR/FAQ process. A team might identify a hard problem during a review that we did not know how to solve, and didn't know if we *could* solve. Jeff would say something to the effect of, "We shouldn't be afraid of taking on hard problems if solving them would unlock substantial value."

Above all, keep in mind that the PR/FAQ is a living document. Once it is approved by the leadership team, it will almost certainly still be edited and changed (a process that should be directed by or reviewed with the leadership team). There is no guarantee that an idea expressed in an excellent PR/FAQ will move forward and become a product. As we've said, only a small percentage will get the green light. But this is not a drawback. It is, in fact, a huge benefit of the process—a considered, thorough, data-driven method for deciding when and how to invest development resources. Generating and evaluating great ideas is the real benefit of the Working Backwards process.

6

Metrics

Manage Your Inputs, Not Your Outputs

Why metrics become more important as a company grows. The metrics life cycle. The difference between input metrics and output metrics. Making sure your metrics are unbiased. Using metrics at business reviews. The key pitfalls of the review meeting.

* * *

Jeff and I (Colin) once visited a Fortune 500 company to meet privately with the CEO in his office. During our meeting, an assistant dashed in and handed the boss a sheet of paper. The CEO glanced at it, waved it at us, and proudly said, "Our stock is up 30 cents this morning!" His mood brightened, as if he had personally caused the rise.

As we drove to our next meeting, Jeff said, "There's nothing that CEO did to cause that 30-cent blip in the stock price." I agreed, and added that I wouldn't be surprised if the assistant had thrown multiple printouts in the recycle bin that morning when the blip wasn't so big. Would the same scene have played out if the share price had dropped 30 cents? The deeper lesson, one that we'll explore in this chapter, is this: share price is what Amazon calls an "output metric." The CEO, and companies in general, have very little ability to directly control output metrics. What's really important is to focus on the "controllable input metrics," the activities you directly control, which ultimately affect output metrics such as share price.

All too often, companies pay attention to the wrong signals, or lack the ability to see into key business trends, even while they feel positively awash in data. In this chapter, we'll show you how to select and measure metrics that will enable you to focus on which activities will drive your business in a meaningful and positive direction. We'll look at how Amazon chooses its metrics by focusing on controllable input metrics, which are the drivers that, when managed well, can lead to profitable growth. We'll talk about how we present and interpret data, and how rigorous metrics ownership drives accountability. We'll also share some hard lessons we learned when

optimizing the wrong metrics, and why we struggled at times to put even the best of our data to good use. We'll show what can happen if your company focuses its attention on the wrong kind of data trends, and we'll describe some common pitfalls.

Unlike the topics covered in previous chapters, there is no single playbook or written set of rules for how Amazon uses metrics to run its businesses. The material we'll discuss is based on our own Amazon experiences as well as discussions we've had with other current and past senior Amazon leaders.

Staying Close to the Business

We've alluded to Amazon's growing pains. Not long into the company's trajectory, there reached a point when Jeff could no longer see each part of the process with his own eyes. Firsthand experience and direct observation were replaced by the proxies of management layers and canned reports. Some business-critical information, such as number of new customers and sales by category, was simply there for the taking and easy to collect. But there were other kinds of information that we could only produce with a series of bespoke ad hoc reports. It was difficult to reliably and quickly answer the question, "How is the business trending?"

This early history is fascinating, and the development of every metric has its own story, but let's skip ahead to 2000 when Amazon recognized $2.76 billion in annual revenue and its famously data-driven culture was prevalent throughout the company. During the fourth quarter—in which our net sales ended up increasing by 44 percent over Q4 of the previous year—there was a daily "war room" meeting where the senior Amazon leaders would analyze a three-page metrics deck and figure out what actions we'd have to take to successfully respond to the demands of what was shaping up to be a record-breaking holiday season. A key component of the deck was the backlog, which was a tally of the orders we had taken minus the shipments we had made. The backlog indicated the amount of work we'd need to do to make sure our customers received their gifts before the holidays. It would take a massive, concentrated effort. Many corporate employees were conscripted for work in the fulfillment centers and customer service. Colin worked the night shift from 7 p.m. to 5:30 a.m. in the Campbellsville, Kentucky, fulfillment center and telecommuted from the Best Western hotel to stay on top of his day job. Bill stayed in Seattle to keep the Video store running smoothly during the day and traveled south 2.5 miles each night to work in the Seattle fulfillment center.

It was touch and go for a while. If we overpromised, we'd ruin a customer's holiday. If we underpromised and stopped accepting orders, we were basically telling our customers to go elsewhere for their holiday needs.

It was close, but we made it. Shortly after that holiday season we held a postmortem, out of which was born the Weekly Business Review (WBR).

The purpose of the WBR was to provide a more comprehensive lens through which to see the business.

The WBR has proved very useful over the years and is widely adopted throughout the company. We'll show how the WBR is constructed and implemented so the company can improve each and every week. It has a fractal nature that allows us to easily adapt to different situations, from small groups to billion-dollar businesses. Small teams, business category lines, and the entire online retail business all have their own WBRs. In addition to our discussion of the benefits of the WBR, we'll point out some common mistakes in their design and execution, including a few big ones we made ourselves. Though we focus on the WBR in this chapter, the same principles and techniques can be applied wherever you need to look at data to help make informed decisions.

The Metrics Life Cycle

When the retail, operations, and finance teams began to construct the initial Amazon WBR, they turned to a well-known Six Sigma process improvement method called DMAIC, an acronym for Define-Measure-Analyze-Improve-Control.[1] Should you decide to implement a Weekly Business Review for your business, we recommend following the DMAIC steps as well. The order of the steps matters. Progressing through this metrics life cycle in this order can prevent a lot of frustration and rework, allowing you to achieve your goals faster.

Define

First, you need to select and define the metrics you want to measure. The right choice of metrics will deliver clear, actionable guidance. A poor choice will result in a statement of the obvious, a nonspecific presentation of everything your company is doing. Donald Wheeler, in his book *Understanding Variation*, explains:

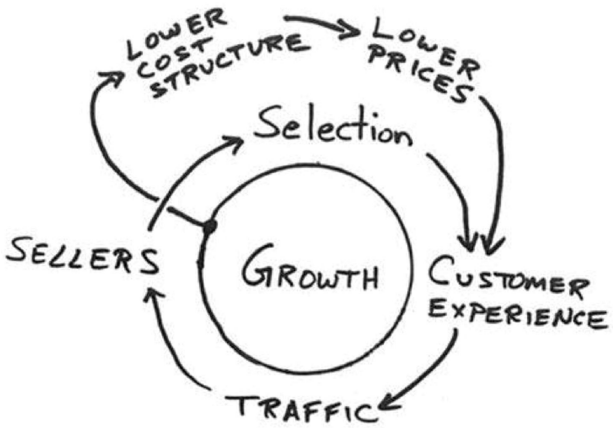
> Before you can improve any system … you must understand how the inputs affect the outputs of the system. You must be able to change the inputs (and possibly the system) in order to achieve the desired results. This will require a sustained effort, constancy of purpose, and an environment where continual improvement is the operating philosophy.[2]

Amazon takes this philosophy to heart, focusing most of its effort on leading indicators (we call these "controllable input metrics") rather than lagging indicators ("output metrics"). Input metrics track things like selection, price, or convenience—factors that Amazon can control through actions such as adding items to the catalog, lowering cost so prices can be lowered, or positioning inventory to facilitate faster delivery to customers. Output metrics—things like orders, revenue, and profit—are important, but they generally can't be directly manipulated in a sustainable manner over the long term. Input metrics measure things that, done right, bring about the desired results in your output metrics.

We can't tell you how many times we've heard people say, when talking about a recently launched Amazon initiative, "You can do that at Amazon because you don't care about profits." That simply isn't true. Profits are just as important to Amazon as to any other major company. Other output metrics like weekly revenue, total customers, Prime subscribers, and (over the long term) stock price—or more accurately, free cash flow per share—matter very much to Amazon. Early detractors mistook Amazon's emphasis on input metrics for a lack of interest in profits and pronounced the company doomed, only to be stunned by its growth over the ensuing years.

1. The Flywheel: Input Metrics Lead to Output Metrics and Back Again

In 2001 Jeff drew the simple diagram below on a napkin to illustrate Amazon's virtuous cycle, also called the "Amazon flywheel." This sketch, inspired by the flywheel concept in Jim Collins's book *Good to Great*, is a model of how a set of controllable input metrics drives a single key output metric—in this case, growth. In this closed-loop system, as you inject energy into any one element, or all of them, the flywheel spins faster:



Since it's a cycle, you can start at any input. The metrics for Customer Experience, for example, could include speed of shipping, breadth of selection, richness of product information, ease of use, and so forth. Watch what happens when we improve customer experience:

- Better customer experience leads to *more traffic*.
- More traffic attracts *more sellers* seeking those buyers.
- More sellers lead to *wider selection*.
- Wider selection enhances *customer experience,* completing the circle.
- The cycle drives *growth,* which in turn *lowers cost structure*.
- Lower costs lead to *lower prices*, improving *customer experience*, and the flywheel spins faster.

The Amazon flywheel captures the major aspect of what makes Amazon's retail business successful. Therefore, it should be no surprise that

almost all the metrics discussed in the WBR can be categorized into one of the flywheel elements. In fact, the first page of the WBR deck has a picture of the very same flywheel above.

2. Identify the Correct, Controllable Input Metrics

This step sounds easy but can be deceptively tricky, and the details matter. One mistake we made at Amazon as we started expanding from books into other categories was choosing input metrics focused around selection, that is, how many items Amazon offered for sale. Each item is described on a "detail page" that includes a description of the item, images, customer reviews, availability (e.g., ships in 24 hours), price, and the "buy" box or button. One of the metrics we initially chose for selection was the number of new detail pages created, on the assumption that more pages meant better selection.

Once we identified this metric, it had an immediate effect on the actions of the retail teams. They became excessively focused on adding new detail pages—each team added tens, hundreds, even thousands of items to their categories that had not previously been available on Amazon. For some items, the teams had to establish relationships with new manufacturers and would often buy inventory that had to be housed in the fulfillment centers.

We soon saw that an increase in the number of detail pages, while seeming to improve selection, did not produce a rise in sales, the output metric. Analysis showed that the teams, while chasing an increase in the number of items, had sometimes purchased products that were not in high demand. This activity did cause a bump in a different output metric—the cost of holding inventory—and the low-demand items took up valuable space in fulfillment centers that should have been reserved for items that were in high demand.

When we realized that the teams had chosen the wrong input metric—which was revealed via the WBR process—we changed the metric to reflect consumer demand instead. Over multiple WBR meetings, we asked ourselves, "If we work to change this selection metric, as currently defined, will it result in the desired output?" As we gathered more data and observed the business, this particular selection metric evolved over time from

- number of detail pages, which we refined to

- number of detail page views (you don't get credit for a new detail page if customers don't view it), which then became
- the percentage of detail page views where the products were in stock (you don't get credit if you add items but can't keep them in stock), which was ultimately finalized as
- the percentage of detail page views where the products were in stock and immediately ready for two-day shipping, which ended up being called Fast Track In Stock.

You'll notice a pattern of trial and error with metrics in the points above, and this is an essential part of the process. The key is to persistently test and debate as you go. For example, Jeff was concerned that the Fast Track In Stock metric was too narrow. Jeff Wilke argued that the metric would yield broad systematic improvements across the retail business. They agreed to stick with it for a while, and it worked out just as Jeff Wilke had anticipated.

Fast Track In Stock, combined with inventory holding cost, provided the teams with an actionable and correct set of input metrics to add selection in a way that would profitably drive sales. Once you have metrics solidified, you can then set a standard and measure teams against that standard. For instance, we decided that in each category, we wanted 95 percent of detail page views to display a product that was in stock and ready for immediate shipping.

These new input metrics created a substantial change in the work and behavior of the category teams. Their focus shifted to reviewing other websites and retail stores and combing through Amazon search logs to determine what items people were searching for in each category but weren't finding on Amazon. From this they could develop a "stack-ranked" or prioritized list of manufacturers to approach and items to acquire that mattered most to consumers. Rather than focusing on the sheer number of items added, they could instead add the items that would make the biggest impact on sales. Sounds simple, but with the wrong input metrics or an input metric that is too crude, your efforts may not be rewarded with an improvement in your output metrics. The right input metrics get the entire organization focused on the things that matter most. Finding exactly the

right one is an iterative process that needs to happen with every input metric.

Note: Most of the examples we give in this chapter are of large companies with substantial resources. But DMAIC and the WBR process is eminently scalable. Your level of investment should be on par with the resources you have.

If you are a nonprofit, figure out a modest number of key metrics that reliably show how well you are doing. For example, how often do you contact your donor base, and how does that frequency affect your funding?

A big mistake people make is not getting started. Most WBRs have humble beginnings and undergo substantial changes and improvement over time.

Measure

Building tools to collect the metrics data you need may sound rather simple, but—like choosing the metrics themselves—we've found that it takes time and concerted effort to get the collection tools right. In chapter two, we discussed how important it is to understand and remove bias in the interview process. Removing bias is just as important in metrics. Each of Jeff's direct reports who ran a business unit had an inherent bias to choose metrics and collect data that would show that their units were trending positive. It's just human nature to want to succeed.

In the early 2000s, Jeff and CFO Warren Jenson—who was succeeded in 2002 by Tom Szkutak—stated explicitly how critical it was for the finance team to uncover and report the unbiased truth. Jeff, Warren, and Tom all insisted that, regardless of whether the business was going well or poorly, the finance team should "have no skin in the game other than to call it like they see it," based on what the data revealed. This truth-seeking mentality permeated the entire finance team and was critical because it ensured that company leaders would have unvarnished, unbiased information available to them as they made important decisions. Having an independent person or team involved with measurement can help you seek out and eliminate biases in your data.

The next step after determining which tools to use is to collect the data and present it in a usable format. Often the data you want will be scattered

across different systems and may take some serious software resources to compile, aggregate, and display correctly. Do not compromise here. Make the investment. If you don't, you may find that you are flying blind with respect to some important aspect of the business.

As you develop the collection tools, make sure they are measuring what you think they are measuring. Diving deep to understand exactly how the data is collected helps spot potential problems. Consider the metric "in stock," which attempts to answer the question, "What percentage of my products are immediately available to purchase and ship?" There are many ways to define and collect data about in-stock items—for example:

- We take a snapshot of our catalog each night at 11 p.m., determine which items are in stock, and weight each item by trailing 30-day product sales. That is, if product A has sold 30 units in the past month and product B has sold 10 units in the past month, and they are both out of stock at the time the in-stock measurement is recorded, product A will impact the in-stock metric three times more than product B will.
- We add software to the product pages that performs the following actions. Every time a product page is displayed, we add one to the metric "Total Number of Product Pages Displayed." If that product is in stock when it is displayed, we add one to the metric "Total Number of In-Stock Product Pages Displayed." At the end of the day, we divide the "Total Number of In-Stock Product Pages Displayed" by the "Total Number of Product Pages Displayed," to get our overall in-stock metric for the day. For example, suppose you displayed one million detail pages total across every product in your catalog, and 850,000 of those product pages displayed a product that was in stock. Then your demand-weighted in-stock percentage for that day would be 85 percent. Products that customers view more have a greater impact on this metric than products that are viewed rarely.

Each of these metrics measures in-stock in a different way and can yield quite a different result for the same business on the same day. The first metric may skew the data depending on the time of day the company

receives the bulk of its inventory. If most of the inventory comes in at night, the item could have been out of stock for most of the day but replenished just before the in-stock data is collected. The result will be that the in-stock performance will look better to the company than what the bulk of the customers actually experienced that day. And if a popular item is out of stock for a long period of time, it will have less of an impact on the metric each day since the metric is weighted by trailing 30-day sales of that item.

The second metric, while more expensive to collect (at least in the short run), is a more accurate representation of what customers experienced that day. It captures, from a customer point of view, what percent of the time they experienced that Amazon was in-stock on the item(s) they viewed. The first metric is inward-facing and operations-centric, while the second metric is outward-facing and customer-centric. Start with the customer and work backwards by aligning your metrics with the customer experience.

One often-overlooked piece of the puzzle is determining how to audit metrics. Unless you have a regular process to independently validate the metric, assume that over time something will cause it to drift and skew the numbers. If the metric is important, find out a way to do a separate measurement or gather customer anecdotes and see if the information trues up with the metric you're looking at. So, a recent example would be testing for COVID-19 by region. It is not enough to look at the number of positive tests in your region as compared to another region with a population of a similar size. You must also look at the number of tests per capita performed in each region. Since both the number of positive tests and the number of tests per capita in each location will keep changing, you will need to keep updating your audit of the measurements.*

Analyze

This stage has been given many different labels by different teams—reducing variance, making the process predictable, getting the process under control, to name a few. But the Analyze stage is all about developing a comprehensive understanding of what drives your metrics. Until you know all the external factors that impact the process, it will be difficult to implement positive changes.

The objective in this stage is separating signals from noise in data and then identifying and addressing root causes. Why is it we can pick 100 items per hour in a fulfillment center on one shift and 30 items per hour on another? Why are we able to display pages in under 100 milliseconds most of the time yet some pages take 10 seconds to display? Why are customer service contacts per order always higher on Mondays than on other days of the week?

When Amazon teams come across a surprise or a perplexing problem with the data, they are relentless until they discover the root cause. Perhaps the most widely used technique at Amazon for these situations is the Correction of Errors (COE) process, based upon the "Five Whys" method developed at Toyota and used by many companies worldwide. When you see an anomaly, ask why it happened and iterate with another "Why?" until you get to the underlying factor that was the real culprit. This COE process requires the team who had a significant error or problem to write a document describing the problem or error, and to drill down on what caused it by asking and answering "Why?" five times in order to get to the true root cause.

Charlie Bell, an SVP in AWS and a great operational guru at Amazon, put it aptly when he said, "When you encounter a problem, the probability you're actually looking at the actual root cause of the problem in the initial 24 hours is pretty close to zero, because it turns out that behind every issue there's a very interesting story."

In the end, if you stick with identifying the true root causes of variation and eliminating them, you'll have a predictable, in-control process that you can optimize.

Improve

Once you have developed a solid understanding of how your process works along with a robust set of metrics, you can devote energy to improving the process. For instance, if you reach the point where you can reliably achieve a weekly 95 percent in-stock rate, you can then ask, "What changes do we need to make to get to 98 percent?"

If you have progressed through the prior three steps (Define, Measure, and Analyze), then your actions to improve the metric will have a higher

chance of succeeding because you'll be responding to signals instead of noise. If you immediately jump to the Improve stage, you'll be working with imperfect information on a process you likely don't fully understand yet, and the actions you take will be much less likely to generate desired results. In the forthcoming example, we'll show how a large Amazon department neglected to complete the first three steps, which caused lots of thrash and yielded no meaningful results.

After you have been operating a WBR for a while, you may notice that a metric is no longer yielding useful information. In that case, it's okay to prune it from the deck.

Control

This final stage is all about ensuring that your processes are operating normally and performance is not degrading over time. As your fundamental understanding of what drives the business improves, it's common for the WBR to become an exception-based meeting rather than a regular one for discussing each and every metric.

Another thing that can happen in this stage is that you'll identify processes that can be automated. Once a process is well understood and the decision-making logic can be encoded in software or hardware, it's a potential candidate for automation. Forecasting and purchasing are two examples of processes that were eventually automated at Amazon. It took years of collaborative effort among category buyers and software engineers—and involving plenty of trial and error—to automate the forecasting and purchasing decisions across the hundreds of millions of products in Amazon's catalog. But it's now done with greater accuracy than even a large team of buyers could do manually.

The WBR: Metrics at Work

At Amazon, the Weekly Business Review (WBR) is the place where metrics are put into action. We'll talk first about how data presentation (mostly graphic) is designed to draw attention where it's most needed. Second, we'll describe the meeting itself, how it's structured to maximize results, and some cautionary notes about how it can fail.

The Deck

Each meeting begins with the virtual or printed distribution of the data package, which contains the weekly snapshot of graphs, tables, and occasional explanatory notes for all your metrics. In this book, we use the term "the deck" to refer to this overall data package. Data visualization software has advanced greatly since the advent of the WBR deck. There are many excellent choices ranging in price from free to modest for smaller organizations, while more advanced tools are available for large enterprises. In practice, many of today's organizations do not assemble a single WBR deck. Instead, separate departments rely on a virtual deck where they can access these data visualization tools to generate the information for their own area. We'll show you a few example graphs in the pages that follow, but first let's review some of the Amazon deck's distinctive features:

The deck represents a data-driven, end-to-end view of the business. While departments shown on org charts are simple and separate, business activities usually are not. The deck presents a consistent, end-to-end review of the business each week that is designed to follow the customer experience with Amazon. This flow from topic to topic can reveal the interconnectedness of seemingly independent activities.

It's mostly charts, graphs, and data tables. With so many metrics to review, written narrative or explanatory notes would undercut the efficiency of the read-through. One notable exception we'll discuss below is how to deal with anecdotes.

How many metrics should you review? There is no magic number or formula. Coming up with the right metrics takes time, and you should seek to improve them continuously. Over time you and your team

should modify, add, and remove metrics based on the strength and quality of the signal each emits.

Emerging patterns are a key point of focus. Individual data points can tell useful stories, especially when compared to other time periods. In the WBR, Amazon analyzes trend lines to highlight challenges as they emerge rather than waiting for them to be summed up in quarterly or yearly results.

Graphs plot results against comparable prior periods. Metrics are intended to trend better over time. Care is taken to ensure that prior periods are structured to provide apples-to-apples comparisons so as not to highlight false variances due to predictable things like holidays or weekends.

Graphs show two or more timelines, for example, trailing 6-week and trailing 12-month. Trend lines for the short term can magnify small but important issues that are hard to spot when averaged out over longer periods.

Anecdotes and exception reporting are woven into the deck. One trait of an Amazon WBR deck that people often remark upon is the liberal use of two tools: anecdotes and exception reporting—that is, the description of an element that falls outside some standard or usual situation. Both tools enable you to dive into examples that contain something that doesn't follow the natural or accustomed patterns and that can sometimes, but not always, reveal a defect, a broken process, or a problem with system logic. The use of anecdotes and exception reporting has enabled leaders to audit at scale in a very detailed way. This ability to flag, evaluate, examine, dig deep, and seek specific solutions for a wide range of issues in a very large organization is something we've noticed is uniquely Amazonian, and it's helpful for small and large businesses alike. We'll provide some examples.

The Meeting

What happens inside the WBR is critical execution not normally visible outside the company. A well-run WBR meeting is defined by intense customer focus, deep dives into complex challenges, and insistence on high standards and operational excellence. One may wonder, at what level is it

appropriate for executives to shift focus to output metrics? After all, companies and their senior executives are routinely judged by output metrics like revenue and profit. Jeff knows this well, in part based on his time spent working at a Wall Street investment firm. The simple answer is that the focus does not shift at any level of management. Yes, executives know their output metrics backward and forward. But if they don't continue to focus on inputs, they lose control over and visibility into the tools that generate output results. Therefore, at Amazon, everyone from the individual contributor to the CEO must have detailed knowledge of input metrics to know whether the organization is maximizing outputs.

The deck is usually owned by someone in finance. Or more accurately, the data in the deck are certified as accurate by finance. However, because multiple people in the room are responsible for each section of the deck, no one "runs" the meeting per se. For most companies, excluding large companies with tens of billions in revenue and multiple big divisions, the audience for the WBR is the CEO and CFO. The meeting attendees should include the executive team and their direct reports as well as anyone who owns or is speaking to any specific section in the deck. Because technology now enables virtual meetings, it is possible to include many more people in the meeting. Adding more junior members of the company to the WBR can increase their engagement in the business and further their growth and development—by allowing them to observe the discussions and thinking of more seasoned leaders.

It is worth noting here that, at Amazon, even the most senior executives review the full WBR deck of metrics, including all the inputs and outputs. Metrics—as well as anecdotes about the customer experience—are the area where the leadership principle Dive Deep is most clearly demonstrated by senior leaders. They carefully examine the trends and changes in the metrics; audit incidents, failures, and customer anecdotes; and consider whether the input metrics should be updated in some way to improve the outputs.

The WBR is an important embodiment of how metrics are put into action at Amazon, but it isn't the only one. Metrics dashboards and reports are established by every engineering, operations, and business unit at the company. In many cases metrics are monitored in real time, and each critical technical and operational service receives an "alarm" to ensure that

failures and outages are identified instantly. In other cases, teams rely on dashboards that are updated hourly or daily for their metrics. The WBR meeting and process is distinctive in how it has enabled Amazon to drive the flywheel faster every year, which in turn has yielded exceptional results.

We use consistent and familiar formatting to speed interpretation

A good deck uses a consistent format throughout—the graph design, time periods covered, color palette, symbol set (for current year/prior year/goal), and the same number of charts on every page wherever possible. Some data naturally lend themselves to different presentations, but the default is to display in the standard format.

Amazon thereby looks at the same set of data every week, in the same order, and gets a holistic view of the business. The team builds up expertise in spotting trends and picks up the rhythm of the review; anomalies stand out more distinctly, and the meeting runs more efficiently.

We focus on variances and don't waste time on the expected

People like talking about their area, especially when they're delivering as expected, and even more so when they exceed expectations, but WBR time is precious. If things are operating normally, say "Nothing to see here" and move along. The goal of the meeting is to discuss exceptions and what is being done about them. The status quo needs no elaboration.

Our business owners own metrics and are prepared to explain variances

Amazon business owners are responsible for tracking the success of their area as defined by their metrics. In the weekly review, the owners, not the finance team, are expected to provide a crisp explanation for variances against expectations. As a result, business owners quickly become adept at spotting trends. Every week they review the deck before the WBR and respond by discussing what action they plan to take to address the variances.

This is a hard-earned lesson; we've seen a metric owner display their metrics in front of a group where it's obviously the first time that person has seen the data. That's a big mistake, a waste of everyone else's time, and will most definitely result in a kerfuffle with the senior leader in the room. By

the time the WBR meeting occurs, each metric owner should have thoroughly analyzed the metrics they own.

Sometimes even the well prepared are hit with a question to which the right answer isn't immediately apparent. In that case, the owner is expected to say something like, "I don't know. We are still analyzing the data and will get back to you." This is preferable to guessing, or worse, making something up on the fly.

We keep operational and strategic discussions separate

The WBR is a tactical operational meeting to analyze performance trends of the prior week. At Amazon, it was not the time to discuss new strategies, project updates, or upcoming product releases.

We try not to browbeat (it's not the Inquisition)

It's okay to dig into a meaningful variation that needs more attention, and to point out when high standards have not been met. Still, success demands an environment where people don't feel intimidated when talking about something that went wrong in their area. Some Amazon teams were better at exemplifying this than others were, and, quite honestly, it's an area where the company could improve. Sometimes WBRs can devolve into downright hostile environments, especially at times when a major slip-up caused the comments to focus more on the presenter than on the issue. While fear may be a good short-term motivator, it will ultimately cause more problems than it solves.

Mistakes should be a learning experience for all. If people become afraid of pointing out their own mistakes because they will feel humiliated in front of their peers, it's human nature for them to do whatever they can to hide those mistakes in future meetings. Variances that get glossed over are lost learning opportunities for everybody. To prevent this, mistakes should be acknowledged as a chance to take ownership, understand the root cause, and learn from the experience. Some tension is unavoidable and appropriate, but we think it's better to establish a culture where it's not just okay, it's actually encouraged to openly discuss mistakes.

We make transitions easy

We've attended many executive meetings where the most expensive team in the company wastes valuable time, for example, fumbling the handoff of

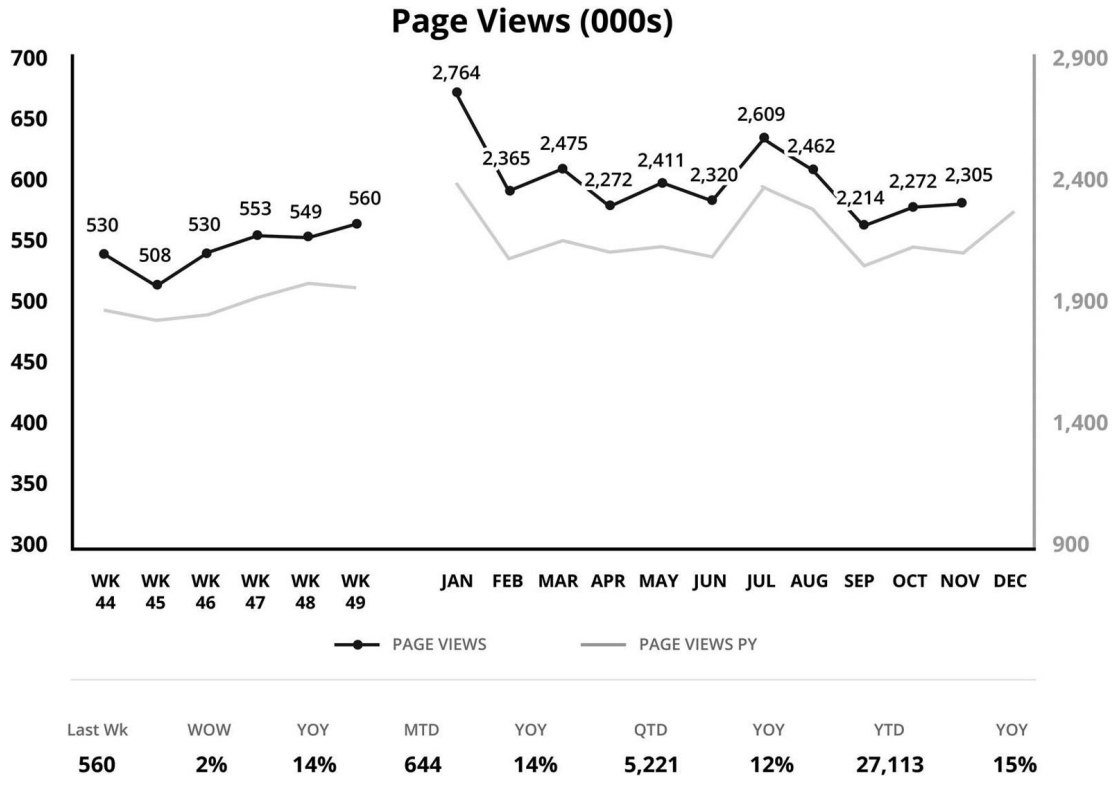
the presentation from one person to the next because the second person's dashboard doesn't load easily, or what have you. To make these transitions quick and seamless, you have to put in the up-front work. The WBR is Amazon's most expensive and impactful weekly meeting, and every second counts—plan ahead and run the meeting efficiently.

Anatomy of a Metrics Chart

A WBR can easily include charts that number in the hundreds, and that much data benefits enormously from consistency of presentation. In our sample charts below, we've included different types of metrics from different functional business areas to illustrate the flexibility of this approach.

Zooming In: Weekly and Monthly Metrics on a Single Graph

As we noted above, at Amazon we routinely place our trailing 6 weeks and trailing 12 months side by side on the same x-axis. The effect is like adding a "zoom" function to a static graph that gives you a snapshot of a shorter time period, with the added bonus that you're seeing both the monthly graph and the "zoomed-in" version of it simultaneously. Here we provide an example of what that dual view looks like in practice. (The following charts do not contain actual data; they are for illustration purposes only.)



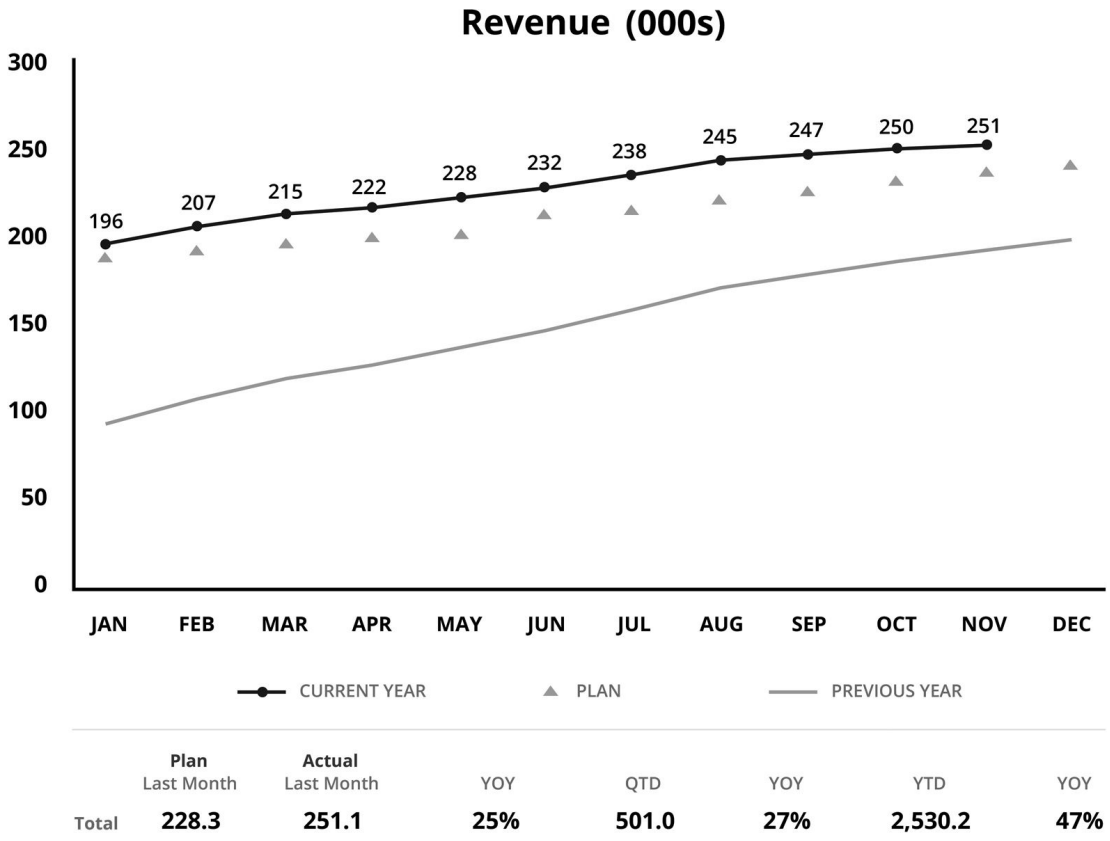
This graphic measures page views for a business, and conveys a lot of data in a small space:

- The gray line is prior year, the black line is current year
- The left graph, those first 6 data points, shows the trailing 6 weeks
- The right graph, with 12 data points, shows the entire trailing year month by month
- This built-in "zoom" adds clarity by magnifying the most recent data, which the 12-month graph puts into context.

At the bottom of the chart, we call out additional key data points, most of which compare one period to another.

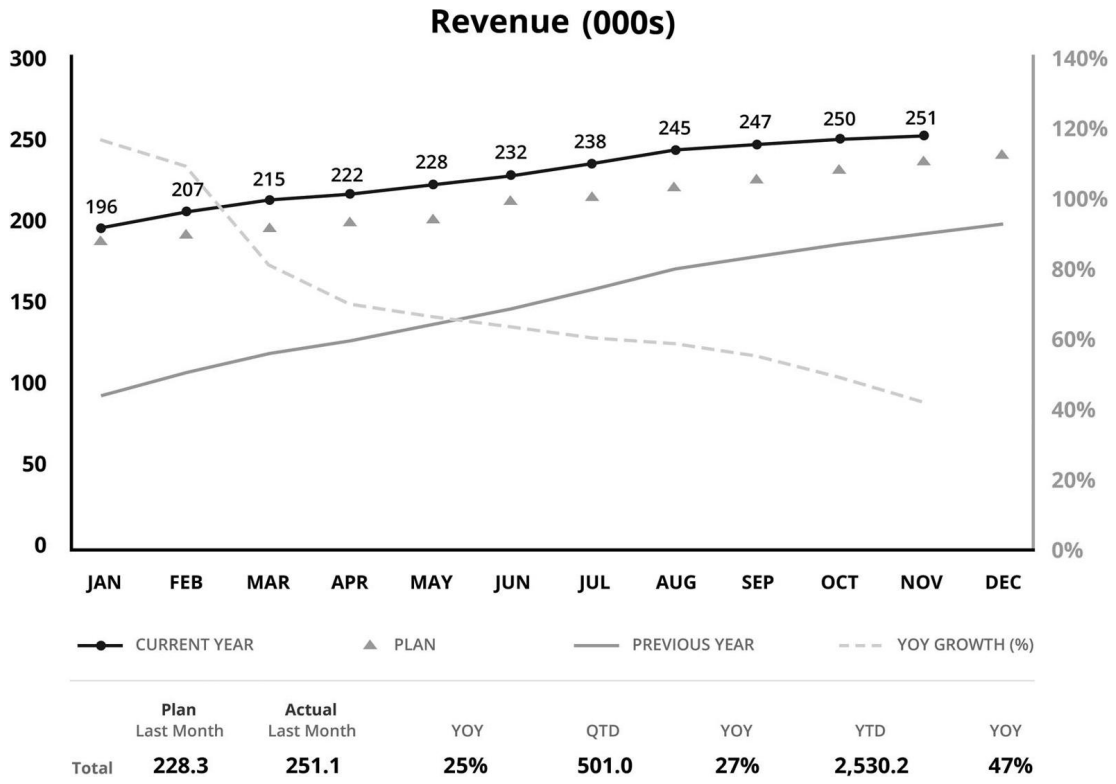
Why We Watch Year-over-Year (YOY) Trends

This example graph, similar to ones you may see in a typical monthly business review (the monthly version of the WBR), compares actual monthly revenue against both planned revenue and prior year revenue. As you can see below, it looks like we are beating plan and growing at a decent clip year over year:



	Plan Last Month	Actual Last Month	YOY	QTD	YOY	YTD	YOY
Total	228.3	251.1	25%	501.0	27%	2,530.2	47%

There's nothing to see here, so let's move on—right? Maybe not. Here's the same graph with one additional trend line: YOY growth rates plotted with a dotted line against a secondary y-axis:



Without the dotted line, you might not have noticed the rate at which the current and projected year trends are slowly converging. Adding YOY *growth rates* in addition to the underlying metric you are measuring is a great way to spot trends. In this example, YOY growth has actually decelerated 67 percent since January with no signs of plateauing. The business may look healthy at a cursory glance, but trouble is looming on the horizon. The enhanced graphic reveals the need for action that the simpler graph obscures.

Output Metrics Show Results. Input Metrics Provide Guidance.

There's another familiar lesson in this graph: output metrics—the data we graphed above—are far poorer indicators of trend causes than input metrics. It turned out in this case that the cause of our decelerating growth was a reduction in the rate of acquiring new customers—but nothing in these graphs gives any clue to that cause. With a sizable existing business, if you only pay attention to the output metric "revenue," you typically won't see the effects of new customer deceleration for quite some time. However, if you look at input metrics instead—things like "new customers," "new

customer revenue," and "existing customer revenue"—you will detect the signal much earlier, and with a much clearer call to action.

Not Every Chart Compares Against Goals

Some WBR charts don't include goals, but that's often appropriate. If the goal of the metric is to spot trends or highlight when a process is out of control, or if we don't have a target at all (e.g., percentage of Android vs iOS mobile users), plotting against a goal is not necessary.

Data Combined with Anecdote to Tell the Whole Story

Numerical data become more powerful when combined with real-life customer stories. The Dive Deep leadership principle states, "Leaders operate at all levels, stay connected to the details, audit frequently, and are skeptical when metrics and anecdotes differ. No task is beneath them."

Amazon employs many techniques to ensure that anecdotes reach the teams that own and operate a service. One example is a program called the Voice of the Customer. The customer service department routinely collects and summarizes customer feedback and presents it during the WBR, though not necessarily every week. The chosen feedback does not always reflect the most commonly received complaint, and the CS department has wide latitude on what to present. When the stories are read at the WBR, they are often painful to hear because they highlight just how much we let customers down. But they always provide a learning experience and an opportunity for us to improve.

One Voice of the Customer story was about an incident when our software barraged a few credit cards with repeated $1.00 pre-authorizations that normally happen only once per order. The customers weren't charged, and such pre-authorizations expire after a few days, but while they were pending, they counted against credit limits. Usually, this would not have much of an effect on the customer. But one customer wrote to say that just after buying an item on Amazon, she went to buy medicine for her child, and her card was declined. She asked that we help resolve the issue so she could purchase the medicine her child needed. At first, an investigation into her complaint revealed that an edge-case bug—another way of saying a rare occurrence—had bumped her card balance over the limit. Many companies

would dismiss such cases as outliers, and thus not worthy of attention, on the assumption that they rarely happen and are too expensive to fix. At Amazon, such cases were regularly attended to because they *would* happen again and because the investigation often revealed adjacent problems that needed to be solved. What at first looked to be just an edge case turned out to be more significant. The bug had caused problems in other areas that we did not initially notice. We quickly fixed the problem for her and for all other impacted customers.

These stories remind us that the work we do has direct impact on customers' lives. There are comparable programs that capture similar anecdotes for Amazon's third-party sellers and corporate AWS customers.

Exception reports come in many flavors, but the following Contribution Profit (CP) example should illustrate the basic concept and its usefulness. CP is defined as the incremental money generated after selling an item and deducting the variable costs associated with that item. It's essentially the money the company has left over after the sale of the item, which goes to pay for the fixed costs of the business and, ideally after that, contributes a profit. There is a CP Exception report that lists the top ten CP negative products (ones that did not generate a profit) within a category for the previous week. Doing a deep dive into these ten products, which often vary from week to week, can reveal very useful information about problems that require action. Here are a few findings that could result from reviewing a top-ten CP negative report:

- CP was negative due to price markdowns that were necessary because we had purchased too many of a certain item, which took up valuable fulfillment center space and capital. The purchase was initiated by the automated purchasing system that had been fed faulty input data. Action: investigate the source of the faulty input data to correct the system.
- CP was negative due to price markdowns originating from a manual purchase order error. The order quantity the buyer entered on the purchase order was too large, and they did not follow the correct process due to a lack of training. Action: use the incident as a teaching moment.

- CP was negative due to faulty cost allocation. The finance system was not allocating costs correctly for a certain class of item. Action: fix the cost allocation system.
- CP was negative because the logistics provider charged more than double the appropriate fee for shipping a particular item. The provider charged the higher fee based on incorrect size and weight information listed for the item in the catalog. Action: fix the catalog data and come up with a plan to put a mechanism in place to prevent the same error from happening for other items in the catalog.
- CP was negative because the item is sold at a low price but is expensive to ship. Whiteboards and yard rakes are examples of products that can fall into this category. Action: evaluate whether these items should be stocked and sold or some other change should be made, such as changing suppliers or changing the default shipping method.

Data and anecdotes make a powerful combination when they're in sync, and they are a valuable check on one another when they are not.

Perhaps the most powerful anecdote in this regard features Jeff himself. Though it happened outside the WBR, it's worth mentioning here. Amazon has a program called Customer Connection, which is mandatory for corporate employees above a certain level. While the details have changed over the years, the premise has remained the same. Every two years the corporate employee is required to become a customer service agent for a few days. The employee gets some basic refresher training from a CS agent, listens in on calls, watches email/chat interactions, and then handles some customer contacts directly. Once they learn the tools and policies, they perform some or all of those tasks under the supervision of a CS agent. (One of my own calls was from a customer whose neighbor's dog had eaten his Amazon package. He offered to send us the uneaten bits to prove his case.)

Jeff is not exempted from this program. While I was working as his shadow, it came time for his Customer Connection recertification, and we dutifully traveled an hour each day to the customer service center in Tacoma, Washington. Jeff was particularly good with customers over the

phone, though he was sometimes overly generous. He gave one customer a full product refund when the policy was to refund the shipping cost only.

On the first day of training, we listened to the CS agent handle a few calls. On one call, the customer complained that her lawn furniture had arrived damaged. The CS agent asked for the product number. As the customer was looking for it, the CS agent muted the call, and said to us, "I bet she's referring to this lawn chair," and pointed to the product on the Amazon site. Sure enough, when the customer read out the number from the packing slip it was the one that the CS had predicted it would be. Jeff and I raised our eyebrows in surprise but didn't want to interrupt the call.

After the issue had been resolved and the call ended, Jeff asked, "How did you know the customer was going to say this?" The CS agent responded that it happened quite often with this newly listed product. The packaging was inadequate, and the furniture often got banged or bruised in transit.

Jeff had recently been learning about how Toyota approached quality control and continuous improvement. One technique they used in their automobile assembly line was the Andon Cord. The car-in-progress moves along the line, and each employee adds a part or performs a task. When any worker notices a quality problem, they are authorized to pull a cord that stops the entire assembly line. A team of specialists swarms to the cord-puller's station, troubleshoots the issue, and develops a fix so the error never happens again.

Here was a similar situation at Amazon, except without the Andon Cord. The CS agent knew of a problem but had no way to improve the process. All the agent could do was offer a concession, make an apology, and ship a new product. We did have a process where each category manager looked at their monthly performance, including products with higher return rates and customer service inquiries. So this issue would have eventually been detected and fixed. But it likely would have taken several weeks and too many additional dissatisfied customers before that would happen.

As we considered the problem of the damaged lawn furniture before the next call arrived, Jeff blurted out, "We need an Andon Cord for customer service." There was no assembly line to halt, but the CS agent would be given the authority to click on what we called "the big red button" on their control screen. Once that button was clicked, two things happened:

the “Add to Cart” and “1-Click” buttons would disappear from the product page so no customers could buy that product, and the category manager would immediately be notified that purchasing for one of their products had been disabled until they could investigate and fix the issue.

It took some time to put Jeff’s idea into operation. We had to build the tools that would remove the Buy Now or Add to Cart button and alert the appropriate internal teams, put together the necessary reporting infrastructure, and train the CS team on how and when to press the big red button. There was some concern that the big red button would be pressed too often. Selling products on a regular basis, after all, was quite important for the health of the company.

That concern proved to be unwarranted—the CS agents were not overly zealous in pressing that button. It turns out that the Amazon version of the Andon Cord empowered the right people, those on the front lines who were talking directly to customers. It surfaced serious issues as soon as they were noticed. It proved once again that giving employees the right tools to solve problems and relying on their good judgment is a powerful combination. It is used widely across Amazon.

That story has been told many times and proves the power of anecdote to illuminate data and make it memorable.

★ ★ ★

Now, as effective as the WBR process can be, it can also go astray in several ways, including poor meeting management, focusing on normal variations rather than signals, and looking at the right data but in the wrong way.

Pitfall 1: Disaster Meetings

One large software group, run by a senior leader who is no longer at Amazon, had memorably rough WBR meetings. Learning and taking ownership of problems and their solutions were two important goals of the WBR process, and on that front these meetings were a huge missed opportunity. They wasted a lot of everyone's time.

One issue was that the attendee list got more and more bloated, and we had to keep finding bigger conference rooms to fit everyone. Likewise, the number of metrics we were trying to track kept ballooning—sometimes for the better, but more often for the worse.

The meetings were also just really unpleasant. There was a lack of ground rules and decorum, with quite a bit of interruption and sniping. Any anomaly was blood in the water, with accusatory questions fired at the presenter. The conversation would quickly regress, as multiple people, usually with little to add, would chime in—seemingly to show off, or to curry favor. Worse yet, some of these lengthy asides seemed to be aimed at running out the clock—with the speaker extending unproductive conversations before their own area would come under fire.

Meetings like that were painful to attend. The Earn Trust leadership principle exists in part to prevent this behavior from occurring. It states, "Leaders listen attentively, speak candidly, and treat others respectfully. They are vocally self-critical, even when doing so is awkward or embarrassing. Leaders do not believe their or their team's body odor smells of perfume. They benchmark themselves and their teams against the best." But these meetings, in the early days, clearly exemplified where we failed to live up to that principle. The original, well-intentioned meeting was set up to improve the software systems from one week to the next. But it gained a life of its own, and sometimes turned a roomful of smart people with probing questions into an angry mob, devouring those who could make a difference and robbing them of their very will to succeed.

What should we have done? Even though, as we've mentioned, there is no one person running the meeting—different people take over for different slides—the most senior person should be responsible for setting the tone and ground rules every week. That person should also, in this case, have

limited attendance to owners and key stakeholders, and also limited the metrics to be reviewed to a specific, essential set: irrelevant metrics should have been deleted from the deck. All of us leaders of that software group, not just that one individual, should have examined the meeting as relentlessly as the participants were examining one another. Collectively, we should have recognized that many of the areas being measured were not yet operationally under control and predictable. Many of the teams had skipped the first three DMAIC steps—define, measure, analyze—in an attempt to operate at the Improve stage. They ended up chasing blips on a graph with not much to show for their effort. We should have—politely and constructively!—recommended they do the necessary legwork to convert their metrics from noise to signal.

Last, we should have recognized that implementing a WBR for this new group for the first time was bound to be messy, requiring trial and error. In the end, we should have ensured that attendees felt free to talk about their mistakes and were actively encouraged to do so, allowing others to learn from them. The key to these meetings is to create a balance between extremely high standards and an atmosphere where people feel comfortable talking about mistakes.

One Amazonian still recalls those disaster meetings even though they occurred more than 15 years ago. He said,

> You're really looking for teams to be willing to take themselves apart, to become naked in front of everyone, to say: "I screwed up. This wasn't right. Here's where it broke." But I remember one particular leader who said instead, "Who is the person with poor judgment who did that?"
>
> The problem with statements like that is that people are basically convicted and sentenced before they've even responded. The leader should have reserved judgment instead of attacking, then begun to understand what actually happened. People are only trying to do the right thing; they're not trying to sabotage the business, and they don't hate customers. They feel tremendous responsibility for what they build.
>
> Since then, we've gotten to be more mature, grounded in freedom from fear. Every time we do the awesome thing, of course we try to

reward it. And the more a team eviscerates itself, being vocally self-critical, the more we try to reward that too. If a team is papering things over, and hasn't looked at the customer experience, *then* you might ask hard questions.

Two things are striking about this recollection. The first is how vivid it remains so many years after the fact—evidence that a punishing environment can leave indelible marks. The second is that this team learned from these early missteps, made adjustments, and eventually built a better process.

Pitfall 2: Noise Obscuring the Signal

Contradictory as this may sound, variation in data is normal. And unavoidable. It's therefore critical to differentiate normal variation (noise) from some fundamental change or defect in a process (signal). Trying to attach meaning to variations within normal bounds is at best a waste of effort and at worst dangerously misleading. It's bad enough when someone proudly explains how their herculean efforts moved their key metric up by 0.1 percent this week, taking precious time away from more important things. Worse, if that same metric went *down* by 0.1 percent, you could easily waste time chasing down the root cause and "fixing" an issue that's really nothing more than normal variation.

At Amazon, understanding what's normal is the responsibility of the metrics owner, whether that's an individual contributor or a manager of thousands. Many statistical methods, such as XMR control charts,[3] can highlight when a process is out of control. For us, however, experience and a deep understanding of the customer most often turned out to be the best way to filter out the signal from the background noise. For the most part, metrics are reviewed daily by their owners and weekly in the WBR, so that expected fluctuations become familiar and exceptions stand out.

★ ★ ★

Amazon's approach to metrics embodies the Customer Obsession leadership principle. The relevance of Customer Obsession becomes evident in the company's focus on input versus output metrics. If you look at the input metrics for Amazon, they often describe things customers care about, such as low prices, lots of available products, fast shipping, few customer service contacts, and a speedy website or app. A lot of the output metrics, such as revenue and free cash flow, are what you'd typically see in a company's financial report. Customers don't care about those. But as we stated at the beginning of this book, Amazon has an unshakable conviction that the long-term interests of shareowners are perfectly aligned with the interests of customers. Controllable input metrics are a quantitative (diving deep with data) and qualitative (anecdotes) way of measuring how well the

organization is satisfying these customer interests so that the output metrics trend the way the company desires.

Properly evaluating your business and striving to improve each week requires a willingness to openly discuss failures, learn from them, and always look for inventions that will delight customers even more.

Appendix B

Sample Narrative Tenets and FAQs

Dave Glick, a former Amazon VP, was the first person to use tenets in the six-pager. Dave had a series of narrative review meetings with Jeff that did not go well. Dave said,

> We had gotten through those bad meetings and to a place where we could have a discussion about our strategy. At the end of the discussion, we had agreement on the strategy, and we summarized it in five bullet points. Jeff said, "You should write these down and put them at the top of your document every month, so we remember what we decided last time." And thus, tenets were born. The next month I showed up with my document with the tenets front and center. It helped us all reload the cache, and made the rest of the meeting productive since we didn't have to rehash our previous decisions.[1]

One of the many benefits tenets can bring is strong alignment among everyone involved. They also provide a set of guiding principles to rely on to help with decision-making. Jeff liked the tenets so much that he asked other teams to incorporate them into their narratives. Formulating a tenet is difficult, and subtle nuances of meaning can sometimes have a large downstream impact on a project.

Tenets help organizations make hard choices and trade-offs. A tenet breaks the tie between two benefits, values, or outcomes where there is a natural tension between them. It is often the case that individuals or departments find themselves in conflict over the two outcomes because there is a legitimate argument for both outcomes. A simple example is speed vs. quality. Obviously both are desirable, and certain teams or individuals may be more focused on speed while others are more focused on quality.

Sample Tenets

Simple example tenet (this is not an Amazon tenet): Speed and quality are always important, but, when forced to make a choice between the two, we will always prioritize quality.

In this tenet, either answer (speed or quality) is valid. When the leadership team of your company aligns on a tenet like this one, refers to it consistently in meetings, and insists that it appear in relevant six-pagers, you will be amazed by how effective this is in aligning and enabling your organization.

Amazon had been working with tenets before we adopted the six-pager narrative approach. Jeff, for example, often discussed the following tenet with various internal audiences.

Tenet: We don't make money when we sell things. We make money when we help customers make purchase decisions.[2]

This guided some challenging and controversial decisions in Amazon's early days, one of which was about product reviews posted on our website. Negative reviews could potentially discourage a customer from buying a product and thus reduce revenue. So, if we are in the business to make money, why would we post negative reviews? But the tenet states that we make money not by selling things, but by helping customers make purchase decisions. The tenet instantly makes our obligation obvious. The customer needs information, positive and negative, to make an informed decision. We continued to post negative customer reviews.

Tenet: When forced to choose between building something that's convenient for customers or convenient for ourselves, we'll choose the former.

Seems like an obvious one, but companies don't always follow this tenet. Packaging, for example. Have you ever experienced the joy of opening a box that contains that product you've desperately been waiting for, only to have your joy turn to despair because the product is encased in a clamshell container of military-grade plastic? That packaging was most definitely created for the convenience of the company—easier to ship, easier to display in a store, harder for customers to steal.

Before we articulated this tenet, Amazon made this very mistake. We developed packaging designed to make it cheap and easy to wrap our books and sturdy enough to prevent damage in shipping. In 1999, Jeff received an email from an elderly woman who wrote that she loved Amazon's service, except for one problem: she had to wait for her nephew to come over to break through the packaging.[3] After receiving that email, Jeff asked the team to invent a new design that would have all the characteristics the company needed and that would also be easy for customers to open. Amazon extended that concept ten years later to other product lines with its Frustration-Free Packaging Initiative.[4]

Tenet: We don't let defects travel downstream. When we notice a defect, we will not rely on good intentions to solve the problem. We'll invent and build systematic methods to eliminate that defect.

This tenet is useful in any continuous improvement environment such as the fulfillment centers and the customer service operation. In order to prevent a defect from traveling downstream, you may need to build systems to detect and measure the defect and create a feedback loop to make sure the defect doesn't happen again. The problem will not be solved by encouraging people to try harder or relying on the good intentions of customer service people. The heartfelt "I'm sorry you had this problem, we will try harder to meet your needs in the future" does not result in the improvement of a flawed system.

One well-known defect in a fulfillment center is the "switcheroo"—when the actual weight of a package ready for loading onto a delivery truck does not match the expected weight of the products that should be in the box (plus the weight of the packaging). This is an indication that something is

wrong with the order—maybe the wrong item was packed or the order is incomplete. When the weights don't match, the package is flagged and a person has to open it and inspect what's inside. This sounds pretty simple, but, in aggregate, it's a massive endeavor. You have to have precise data on the weight of tens of millions of items from millions of manufacturers, merchants, and sellers. Your weighing scales have to be extremely accurate, or they may detect a mismatch when there isn't one.

But what happens if a package goes out with the defect undetected? The customer may get something different from what was ordered. This does not make for a good customer experience.

The tenet says that we will "eliminate the defect." That's an aggressive goal, and it cannot be achieved immediately. It serves as a powerful advocate for the customer, and it has led to the development of many systems and processes to prevent and eliminate defects. As we've described, one of the best-known of these processes is the Andon Cord, which was adapted from the Toyota Production System: factory workers can pull a physical cord to halt the assembly line when they spot a defect. At Amazon, the customer service people have a virtual cord—actually a button—that they can push when a defect is noticed. It instantly prevents Amazon from selling any more of the affected product until the customer issue is resolved.

This tenet appeared in so many narratives and was so useful in advocating for the customer that Amazon incorporated it into the Leadership Principles as Insist on the Highest Standards.

Sample FAQs

An FAQ is a good way to tee up issues for discussion or highlight important points or risks in your argument. Such FAQs allow the author to take control over the discussion and steer it to productive areas for dialogue. An honest, objective, and nonemotional tone tends to work best when answering these questions. There's no point in sugarcoating things, and it helps to state the tough issues up front. Amazon's Earn Trust leadership principle states, "Leaders listen attentively, speak candidly, and treat others respectfully. They are vocally self-critical, even when doing so is awkward or embarrassing. Leaders do not believe their or their team's body odor smells of perfume. They benchmark themselves and their teams against the best." Here are some FAQs that we have found useful:

What were the biggest mistakes we have made last period, and what have we learned from them?
What are the key inputs for this business?
What is the single biggest thing we can do to move the needle in this business, and how will we organize to do just that?
What are the top reasons we should not do what we're proposing today?
When push comes to shove, what are the things we won't compromise on?
What's hard about the problem we are trying to solve?
If our team had X more people or Y more dollars, how would we deploy those resources?
What are the top three new initiatives, products, or experiments our team has launched in the past X months, and what did we learn from them?
What dependencies do we have in our area today over which we wish we had control?