

MEASURING AND CHANGING CULTURE

It is practically a truism in DevOps circles that culture is of huge importance. However, culture is intangible; there exist many definitions and models of culture. Our challenge was to find a model of culture that was well-defined in the scientific literature, could be measured effectively, and would have predictive power in our domain. Not only did we achieve these objectives, we also discovered that it is possible to influence and improve culture by implementing DevOps practices.

MODELING AND MEASURING CULTURE

There are many approaches to modeling *culture* in the literature. You can choose to look at national culture—for example, what country one belongs to. You may also talk about what organizational cultural values are enacted that influence the way teams behave. And even within organizational culture, there are several ways to define and model “culture.” Organizational culture

can exist at three levels in organizations: basic assumptions, values, and artifacts (Schein 1985). At the first level, basic assumptions are formed over time as members of a group or organization make sense of relationships, events, and activities. These interpretations are the least “visible” of the levels—and are the things that we just “know,” and may find difficult to articulate, after we have been long enough in a team.

The second level of organizational culture are values, which are more “visible” to group members as these collective values and norms can be discussed and even debated by those who are aware of them. Values provide a lens through which group members view and interpret the relationships, events, and activities around them. Values also influence group interactions and activities by establishing social norms, which shape the actions of group members and provide contextual rules (Bansal 2003). These are quite often the “culture” we think of when we talk about the culture of a team and an organization.

The third level of organizational culture is the most visible and can be observed in artifacts. These artifacts can include written mission statements or creeds, technology, formal procedures, or even heroes and rituals (Pettigrew 1979).

Based on discussions in DevOps circles and the importance of “organizational culture” at the second level, we decided to select a model defined by sociologist Ron Westrum. Westrum had been researching human factors in system safety, particularly in the context of accidents in technological domains that were highly complex and risky, such as aviation and healthcare. In 1988, he developed a typology of organizational cultures (Westrum 2014):

- **Pathological** (power-oriented) organizations are characterized by large amounts of fear and threat. People often hoard information or withhold it for political reasons, or distort it to make themselves look better.
- **Bureaucratic** (rule-oriented) organizations protect departments. Those in the department want to maintain their “turf,” insist on their own rules, and generally do things by the book—*their* book.
- **Generative** (performance-oriented) organizations focus on the mission. How do we accomplish our goal? Everything is subordinated to good performance, to doing what we are supposed to do.

Westrum’s further insight was that the organizational culture predicts the way information flows through an organization. Westrum provides three characteristics of good information:

1. It provides answers to the questions that the receiver needs answered.
2. It is timely.
3. It is presented in such a way that it can be effectively used by the receiver.

Good information flow is critical to the safe and effective operation of high-tempo and high-consequence environments, including technology organizations. Westrum describes the characteristics of organizations that fall into his three types in Table 3.1.

An additional insight from Westrum was that this definition of

organizational culture predicts performance outcomes. We keyed in on this in particular, because we hear so often that culture is important in DevOps, and we were interested in understanding if culture could predict software delivery performance.

Table 3.1 Westrums Typology of Organizational Culture.

Pathological (Power-Oriented)	Bureaucratic (Rule-Oriented)	Generative (Performance-Oriented)
Low cooperation	Modest cooperation	High cooperation
Messengers “shot”	Messengers neglected	Messengers trained
Responsibilities shirked	Narrow responsibilities	Risks are shared
Bridging discouraged	Bridging tolerated	Bridging encouraged
Failure leads to scapegoating	Failure leads to justice	Failure leads to inquiry
Novelty crushed	Novelty leads to problems	Novelty implemented

MEASURING CULTURE

In order to measure the culture of organizations, we take advantage of the fact that these types form “points on a scale . . . a ‘Westrum continuum’” (Westrum 2014). This makes it an excellent candidate for Likert-type questions. In psychometrics, the Likert scale is used to measure people’s perceptions by asking them to rate how strongly they agree or disagree with a statement. When people answer a Likert-type question, we assign the answer a value on a scale from 1 to 7, where 1 means “Strongly disagree”

and 7 means “Strongly agree.”

For this approach to work, the statement must be worded strongly, so that people can strongly agree or disagree (or indeed feel neutral) about it. You can see a re-creation from the survey showing the statements we created from Westrum’s model, along with the Likert scale, in Figure 3.1.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Information is actively sought.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Messengers are not punished when they deliver news of failures or other bad news.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Responsibilities are shared.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cross-functional collaboration is encouraged and rewarded.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Failure causes inquiry.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
New ideas are welcomed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Failures are treated primarily as opportunities to improve the system.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 3.1: Likert-Type Questions for Measuring Culture

Once we have the responses to these questions from several people (often dozens or hundreds of people), we need to determine if our measure of organizational culture is valid and reliable from a statistical point of view. That is, we need to find out if the questions are being understood similarly by all people taking the survey and if, taken together, they are actually measuring organizational culture. If analyses using several statistical tests confirm these properties, we call what we have measured a “construct” (in this case, our construct would be “Westrum organizational culture”), and we can then use this measure in further research.

Analyzing Constructs

Prior to conducting any analysis between our measures—for example, does organizational culture impact software delivery performance?—we must analyze the data and measures themselves. When using robust survey measures, we use *constructs*.

In this first step, we conducted several analyses to ensure our survey measures were valid and reliable. These analyses included tests for discriminant validity, convergent validity, and reliability.

- **Discriminant validity:** making sure that items that are not supposed to be related are actually unrelated (e.g., that items that we believe are not capturing organizational culture are not, in fact, related to it).
- **Convergent validity:** making sure that items that are supposed to be related are actually related (e.g., if measures are supposed to measure organizational culture, they do measure it).
- **Reliability:** making sure the items are read and interpreted similarly by those who take the survey. This is also referred to as internal consistency.

Taken together, validity and reliability analyses confirm our measures and come before any additional analyses to test for relationships, like correlation or prediction. For more on validity and reliability, refer to Chapter 13. Additional information about the statistical tests used to confirm validity and reliability

can be found in Appendix C.

Our research has consistently found our Westrum construct—an indicator of the level of organizational culture that prioritizes trust and collaboration in the team—to be both valid and reliable.¹ This means you can use these questions in your surveys too. To calculate the “score” for each survey response, take the numerical value (1-7) corresponding to the answer to each question and calculate the mean across all questions. Then you can perform statistical analysis on the responses as a whole.

Culture enables information processing through three mechanisms. First, in organizations with a generative culture, people collaborate more effectively and there is a higher level of trust both across the organization and up and down the hierarchy. Second, “generative culture emphasizes the mission, an emphasis that allows people involved to put aside their personal issues and also the departmental issues that are so evident in bureaucratic organizations. The mission is primary. And third, generativity encourages a ‘level playing field,’ in which hierarchy plays less of a role” (Westrum 2014, p. 61).

We should emphasize that bureaucracy is not necessarily bad. As Mark Schwartz points out in *The Art of Business Value*, the goal of bureaucracy is to “ensure fairness by applying rules to administrative behavior. The rules would be the same for all cases—no one would receive preferential or discriminatory treatment. Not only that, but the rules would represent the best products of the accumulated knowledge of the organization: Formulated by

bureaucrats who were experts in their fields, the rules would impose efficient structures and processes while guaranteeing fairness and eliminating arbitrariness” (Schwartz 2016, p. 56).

Westrum’s description of a rule-oriented culture is perhaps best thought of as one where following the rules is considered more important than achieving the mission—and we have worked with teams in the US Federal Government we would have no issue describing as generative, as well as startups that are clearly pathological.

WHAT DOES WESTRUM ORGANIZATIONAL CULTURE PREDICT?

Westrum’s theory posits that organizations with better information flow function more effectively. According to Westrum, this type of organizational culture has several important prerequisites, which means that it is a good proxy for the characteristics described by these prerequisites.

First, a good culture requires trust and cooperation between people across the organization, so it reflects the level of collaboration and trust inside the organization.

Second, better organizational culture can indicate higher quality decision-making. In a team with this type of culture, not only is better information available for making decisions, but those decisions are more easily reversed if they turn out to be wrong because the team is more likely to be open and transparent rather than closed and hierarchical.

Finally, teams with these cultural norms are likely to do a better job with their people, since problems are more rapidly discovered and addressed.

We hypothesized that culture would predict both software delivery performance and organizational performance. We also predicted that it would lead to higher levels of job satisfaction.² Both of these hypotheses proved to be true. We show these relationships in Figure 3.2.

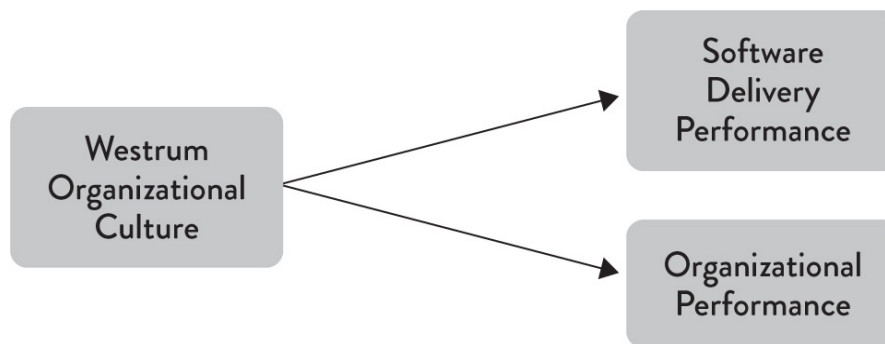


Figure 3.2: Westrum Organizational Culture's Outcomes

CONSEQUENCES OF WESTRUM'S THEORY FOR TECHNOLOGY ORGANIZATIONS

For modern organizations that hope to thrive in the face of increasingly rapid technological and economic change, both resilience and the ability to innovate through responding to this change are essential. Our research into the application of Westrum's theory to technology shows that these two characteristics are connected. Initially developed to predict safety outcomes, our research shows it also predicts both software delivery and organizational performance. This makes sense,

because safety outcomes are performance outcomes in a healthcare setting. By extending this to technology, we expected this type of organizational culture to positively impact software delivery and organizational performance. This mirrors research performed by Google into how to create high-performing teams.

The Delivery Performance Construct

In Chapter 2, we said that delivery performance combines four metrics: lead time, release frequency, time to restore service, and change fail rate. When performing cluster analysis, all four metrics together meaningfully classify and discriminate among our high, medium, and low performers. That is, all four measures are good at categorizing teams. However, when we tried to turn these four metrics into a *construct*, we ran into a problem: the four measures don't pass all of the statistical tests of validity and reliability. Analysis showed that only lead time, release frequency, and time to restore together form a valid and reliable construct. Thus, in the rest of book, when we talk about *software delivery performance* it is defined using only the combination of those three metrics. Also, when software delivery performance is shown to correlate with some other construct, or when we talk about predictions involving software delivery performance, we're only talking about the construct as defined and measured this way.

Note, however, that change fail rate is strongly correlated with the software delivery performance construct, which means that in most cases, things correlated with the software delivery

performance construct are also correlated with change fail rate.

Google wanted to discover if there were any common factors among its best-performing teams. They started a two-year research project to investigate what made Google teams effective, conducting “200+ interviews with . . . employees and [looking] at more than 250 attributes of 180+ active Google teams” (Google 2015). They expected to find a combination of individual traits and skills that would be key ingredients of high-performing teams. What they found instead was that “who is on a team matters less than how the team members interact, structure their work, and view their contributions” (Google 2015). In other words, it all comes down to team dynamics.

How organizations deal with failures or accidents is particularly instructive. Pathological organizations look for a “throat to choke”: Investigations aim to find the person or persons “responsible” for the problem, and then punish or blame them. But in complex adaptive systems, accidents are almost never the fault of a single person who saw clearly what was going to happen and then ran toward it or failed to act to prevent it. Rather, accidents typically emerge from a complex interplay of contributing factors. Failure in complex systems is, like other types of behavior in such systems, emergent (Perrow 2011).

Thus, accident investigations that stop at “human error” are not just bad but dangerous. Human error should, instead, be the *start* of the investigation. Our goal should be to discover how we could improve information flow so that people have better or

more timely information, or to find better tools to help prevent catastrophic failures following apparently mundane operations.

HOW DO WE CHANGE CULTURE?

John Shook, describing his experiences transforming the culture of the teams at the Fremont, California, car manufacturing plant that was the genesis of the Lean manufacturing movement in the US, wrote, “what my . . . experience taught me that was so powerful was that the way to change culture is not to first change how people think, but instead to start by changing how people behave—what they do” (Shook 2010).³

Thus we hypothesize that, following the theory developed by the Lean and Agile movements, implementing the practices of these movements can have an effect on culture. We set out to look at both technical and management practices, and to measure their impact on culture. Our research shows that Lean management, along with a set of other technical practices known collectively as *continuous delivery* (Humble and Farley 2010), do in fact impact culture, as shown in Figure 3.3.

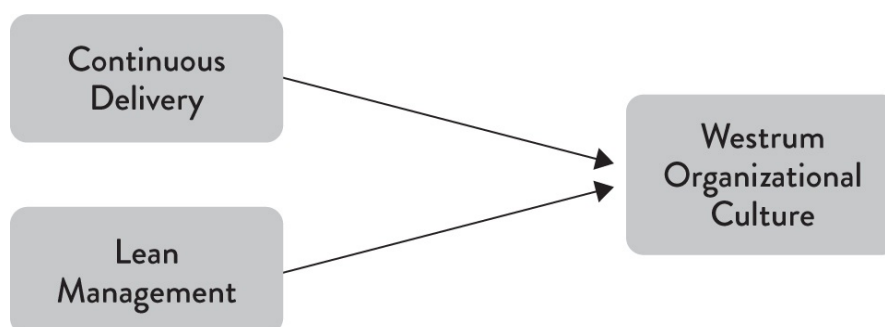


Figure 3.3: Westrum Organizational Culture’s Drivers

You can act your way to a better culture by implementing these practices in technology organizations, just as you can in manufacturing. In the next chapter we'll examine the technical practices, and then in Chapters 7 and 8 we'll discuss management practices.

¹ In 2016, 31% of respondents were classified as pathological, 48% bureaucratic, and 21% generative.

² These hypotheses are based on previous research and existing theories, and bolstered by our own experiences and the experiences we see and hear from others in the industry. Our research hypotheses are all built this way. This is an example of inferential predictive research, which you can read more about in Chapter 12.

³ The story of this transformation is told in episode 561 of the WBEZ radio show *This American Life* (This American Life 2015).