Achieving Decentralized Control

Militaries have spent centuries wrestling with the balance between centralized and decentralized control, with the ideas regularly tested in battle. Moreover, the modern military, exemplified by the US Marine Corps, relies upon the initiative of subordinates. Marines view war as a domain of inherent uncertainty, where the side that can best exploit uncertainty will win.

Product developers often track deviations from plan and invest effort into correcting these deviations. The starting assumption is that deviations are bad and conformance to plan is ideal. In contrast, Marines believe that warfare constantly presents unforeseen obstacles and unexpected opportunities. They recognize that this emerging information is visible first to the front lines, so control is pushed down to front line troops who are trained to react quickly.

Balancing Centralization and Decentralization

Modern militaries focus on **decentralized execution supported by centralized coordination**.

D1: The Perishability Principle: Decentralize control for problems and opportunities that age poorly

We should decentralize control to seize opportunities which are fleeting. This does not mean decentralizing responsibility: we need to decentralize the authority to act, and pre-position sufficient resources to make this action meaningful.

D2: The Scale Principle: Centralize control for problems that are infrequent, large, or that have significant economies of scale

Centralizing resources often increases response times, but some problems require large-scale responses which must be centralized.

It is also appropriate to centralize resources to pool variable demand (e.g. of specialist roles such as legal counsel). Centralized resources also have the benefit of consistency between projects. However, centralized teams must be measured on response time instead of local efficiency: a centralized team loaded to capacity will respond slowly.

D3: The Principle of Layered Control: Adapt the control approach to emerging information about the problem

If you can judge the severity of the problem at its time of arrival, you can use **triage** to decide whether the response should be centralized or decentralized. If we cannot know that in advance then we can use **escalation**: problems which are not resolved within a certain time are escalated to a higher organizational level.

D4: The Opportunistic Principle: Adjust the plan for unplanned obstacles and opportunities

The modern military does not create plans for the purpose of measuring conformance; they plan as a tool for maintaining alignment. **The plan forms a baseline that synchronizes complex changes.**

In product development, a planned feature may turn out to be more costly to execute than planned. A good process will bypass it. Alternatively, we may discover a feature is easier to implement or more valuable than initially thought, which may prompt us to exceed our initial goal even when it means descoping other marginal requirements.

D5: The Principle of Virtual Centralization: Be able to quickly reorganize decentralized resources to create centralized power

It is often not economical to have specialized resources sitting idle. Instead, businesses may form "tiger teams"—specialized, cross-functional groups brought together to solve a specific problem—from experts who otherwise have normal day jobs.

D6: The Inefficiency Principle: The inefficiency of decentralization can cost less than the value of faster response time

Whether to centralize or decentralize is an economic decision. The marginal cost of inefficiency through decentralization should be compared to the marginal gains of response times.

Military Lessons on Maintaining Alignment

The dark side of decentralized control is misalignment. If all local actors make locally optimum choices, the system level outcome may still be a disaster.

D7: The Principle of Alignment: There is more value created with overall alignment than local excellence

Military doctrine recognises that combat effects are not linearly proportional to effort. Similarly, alignment in product development produces disproportionate effects: one percent excess over parity on 15 features creates no excitement, but a 15 percent advantage on a key attribute may shift customer preference.

D8: The Principle of Mission: Specify the end state, its purpose, and the minimal possible constraints

Mission orders focus on the **intent** of the operation, rather than constraining how the mission is to be accomplished. The mission, communicated via the "Commander's intent", is the end state that is trying to be created by the battle. It is always a goal that is larger than the battle itself, communicated via the "why" of the operation, not the "what" or "how". Marines believe that when intent is clear, they will be able to select the right course of action.

D9: The Principle of Boundaries: Establish clear roles and boundaries

The military is quite rigorous in defining specific roles for individual units and establishing clear responsibilities for adjacent units. In warfare, poorly defined roles are a life-and-death matter. In product development, they increase the need for communication and increase the time spent in meetings.

D10: The Main Effort Principle: Designate a main effort and subordinate other activities

Marines designate one point of action as the main effort. Forces in other areas are reduced to create maximum strength at this focal point. In product development, there are often just a few critical, preference-shifting features. Other features should be subordinated to achieving performance on these.

D11: The Principle of Dynamic Alignment: The main effort may shift quickly when conditions change

More important even than the main effort is the fact that information can change during development. When it does, the main effort should shift accordingly.

D12: The Second Agility Principle: Develop the ability to quickly shift focus

Product teams should organise to pivot quickly. Small teams of skilled, trained people can change direction faster than a large team. They can pivot quickly when the product has a streamlined feature set rather than one bloated with minor features. They can pivot quickly when they have reserve capacity and energy (an overloaded team will struggle to apply a burst of extra effort).

We can also exploit architecture to enable rapid change by partitioning the system to gracefully absorb change. For example, we might segregate uncertainty into one zone of the design and couple this zone loosely to the rest of the system.

The effects of faster decisions are also cumulative. We can accelerate the speed of decision making and reactions through the application of OODA (observe—orient—decide—act) loops.

D13: The Principle of Peer-Level Coordination: Tactical coordination should be local

Maintaining alignment is easy with centralized control. It is much more difficult when we decentralize control and emphasize local initiative. Marines achieve this through explicit and implicit communications: the former through face-to-face and voice comms, the latter through doctrine and training. Marines can predict what other marines will do with surprising accuracy thanks to extensive training with peers.

In both cases there is an emphasis on lateral communication rather than strict hierarchical communication. The continuous peer-to-peer communication of a colocated team is far more effective at responding to uncertainty than a centralized project management organization.

D14: The Principle of Flexible Plans: Use simple modular plans

The recognition of uncertainty in warfare does not mean the absence of planning. Everything that can be forecast reasonably well in advance is carefully planned.

Flexibility is achieved by preplanning "branches" and "sequels". Branches are points where different paths can be selected depending on existing conditions. Sequels are preplanned follow-on actions. This approach increases the chance of maintaining alignment when conditions change.

D15: The Principle of Tactical Reserves: Decentralize a portion of reserves

To enable quick realignment, militaries will pre-position reserves at different organisational levels, so that force can be applied at the right time and place instead of making guesses in advance.

In product development we can pre-position capacity margin at various levels within the project organizational structure, in both staffing levels and budgetary capacity, and especially in key processes. This permits support groups to absorb variation locally instead of having to elevate capacity issues to higher organizational levels.

D16: The Principle of Early Contact: Make early and meaningful contact with the problem

Contact with the enemy resolves a great deal of uncertainty, and prevents it from generating.

In product development, our opposing forces are the market and technical risks that must be overcome during the project. It is critical to gain early market feedback, and to focus early efforts on zones of high technical or market risk.

The Technology of Decentralization

D17: The Principle of Decentralized Information: For decentralized decisions, disseminate key information widely

Decentralizing control requires decentralizing both the authority to make decisions and the information required to make these decisions correctly. In the Marine Corps, the minimum standard is to **understand the intentions of commanders two levels higher** in the organization.

This does not mean all information must be disseminated. Principle D20 illustrates how global decision rules can also enable fast local decision making.

D18: The Frequency Response Principle: We can't respond faster than our frequency response

In product development, we need to accelerate decision-making speed. This is done by involving fewer people and fewer layers of management.

To enable lower organizational levels to make decisions, we need to give them **authority**, **information**, and **practice**. Without practice and the freedom to fail upon occasion, they will not take control of these decisions.

D19: The Quality of Service Principle: When response time is important, measure response time

Too many product organisations measure and incentivise efficiency. If time on the critical path is worth a lot, then performance should be measured on response time.

This is particularly true of centralized support groups, which are often unaccountable to delivery teams. A quality of service agreement can start to address this.

D20: The Second Market Principle: Use internal and external markets to decentralize control

Most organizations provide premium internal services at no cost, which means prioritization decisions for scarce resources must be escalated to higher organizational levels, creating delays. **Differential pricing** for premium support services lets teams make localized decisions on when they're willing to pay for additional support.

The Human Side of Decentralization

D21: The Principle of Regenerative Initiative: Cultivating initiative enables us to use initiative

Marines consider inaction and lack of decisiveness to be much more dangerous than making a bad decision. An imperfect decision executed rapidly is far better than a perfect decision that is implemented late.

The more we encourage initiative, the more chance we have to provide positive reinforcement for this initiative, and thus to execute rapidly.

D22: The Principle of Face-to-Face Communication: Exploit the speed and bandwidth of face-to-face communications

Meaningful communication entails feedback and clarification. Verbal communications are simply much more effective in generating rapid feedback than written comms. This does not require full time colocation: for supporting resources face-to-face communication on a regular cadence may be sufficient without demanding a continuous presence.

D23: The Trust Principle: Trust is built through experience

We trust someone when we feel we can predict their behavior under difficult conditions. Psychologists use the term "social expectancy theory" to refer to the ability to predict the behavior of other members of our social group. They point out that we can work effectively with people when we can predict their behavior.

Since trust is primarily based on experience, it follows that we can increase trust by maintaining continuity in our teams. Moreover, by working in small batch cycles we expose team members to more frequent cycles of learning, better equipping them to predict the behavior of other organizational members. Small batch sizes build trust, and trust enables the decentralized control that enables us to work with small batch sizes.