

Influence Drift Theory (IDT): Modeling Gradual Organizational Change through Interpersonal Influence

Written with ChatGPT.

Hyunho Shin (hyunho.shin@noctuabio.co.kr)

Abstract

Organizations have traditionally been studied through structural and rational lenses, from classical bureaucracy to human relations, systems, and contingency theories. However, these perspectives often treat organizations as static structures or assume rational actors, overlooking the nuanced *drift* of beliefs and behaviors among members over time. In this paper, we propose **Influence Drift Theory (IDT)** as a novel framework grounded in the idea that organizations are sustained by shared fictions and interpersonal influences. Drawing on philosophical insights (e.g. Harari's view of corporations and nations as imagined realities), IDT models an organization as a network of individuals who continuously influence one another's values, attitudes, and performance. We contrast IDT with classical organizational theories and highlight its dynamic, gradual interpersonal influence mechanics. A mathematical structure is outlined where each member's attributes (e.g. ethical stance, competence, trustworthiness) evolve through weighted feedback from others. Key parameters – including feedback cycles, hierarchy depth, team size, and directional influence weights – are defined to show how different organizational forms can be simulated. We explore diverse influence metrics (ethics, competence, trust) and analyze their evolution via simulation of various scenarios. In illustrative case studies, we show how **authoritarian cultures may drift into decay**, **democratic structures foster resilience**, and **ethical leadership can propagate integrity** across a network. We compare IDT's continuous influence model to binary decision models like classical game theory, underscoring how gradual change and path-dependence in IDT capture complexities that static equilibrium models miss. Finally, we discuss real-world applications of IDT for understanding corporate culture shifts, governance and policy-making, and community development. The paper concludes that Influence Drift Theory provides a comprehensive, flexible paradigm for analyzing and guiding organizational change through the lens of interpersonal influence.

Introduction

How do organizations evolve and sometimes lose sight of their founding values or goals?

Traditional organizational theories offer partial answers: classical theorists emphasized formal structures and efficiency, human relations scholars highlighted social factors, and systems theorists viewed organizations as complex open systems. These approaches, while invaluable, often abstract away the minute-to-minute interactions between individuals. In reality, an organization's culture and effectiveness emerge from countless small interpersonal influences – conversations, observations, and social learning – that accrue over time. This paper introduces **Influence Drift Theory (IDT)**, a theoretical framework that places these micro-level influence dynamics at the center of organizational analysis.

IDT is motivated by two insights from prior discourse. First, following Yuval Noah Harari's observation, organizations (from corporations to nations) are not tangible entities but “**imagined realities**” or shared **fictions** that exist in the collective minds of people. If an organization is essentially a network of people bound by common beliefs and narratives, then understanding how those beliefs spread and shift among individuals is crucial. Second, classical models of organizational behavior and decision-making (including game-theoretic models) often assume that actors have fixed preferences or that organizational change happens through discrete decisions or shocks. In practice, change is frequently *gradual* – a product of incremental influence and adjustment. Small shifts in attitudes can compound, a phenomenon akin to what Dekker (2011) calls “*drift into failure*,” wherein local optimizations and minor deviations accumulate into large-scale outcomes. IDT seeks to synthesize these ideas into a cohesive theory: organizations are living networks of influence in which continuous interpersonal interactions lead to slow drifts in collective behavior and performance.

We proceed as follows. First, we ground the theory philosophically by examining the nature of organizations as social constructs and the implications for individual agency and influence. Next, we review classical, human relations, systems, and contingency theories of organization to identify what they explain and where a dynamic influence-based view can add insight. We then formally introduce **Influence Drift Theory**, outlining its key mechanics, parameters, and a mathematical representation of influence dynamics. The flexibility of IDT is illustrated by incorporating multiple influence metrics (such as ethics, competence, and trust) and discussing how these can be simulated to observe emergent patterns. We compare IDT's gradual, continuous model of change with binary and equilibrium-focused models like traditional game

theory, highlighting differences in assumptions and outcomes. Finally, we present several **illustrative scenarios** – including an authoritarian organization’s ethical decay, a democratic team’s resilience to disruption, and the propagation of ethical norms by leadership – to demonstrate the theory in action. In discussing these scenarios, we draw connections to real-world cases in corporations, governments, and communities. The paper concludes with implications for research and practice, suggesting that IDT offers a robust framework for both analyzing organizational phenomena and designing interventions to steer organizations towards desirable outcomes.

Philosophical Foundations: Organizations as Imagined Orders

Classical ontology might describe an organization in terms of its buildings, products, or formal hierarchy, but a deeper perspective is that an organization’s true existence lies in the shared understandings of its members. Harari argues that entities like nations, corporations, money, or religions are “**imagined realities**” – useful fictions that exist because many people collectively believe in them. For example, “*limited liability companies*” (like corporations) have no physical essence; they are legal and social constructs that people treat as real, enabling large-scale cooperation among strangers. The implication is profound: if belief in the organization falters, the organization itself can unravel. John Searle’s work on social ontology similarly contends that institutional facts (such as a corporation’s authority or a dollar’s value) arise from collective intentionality – a community’s shared agreement on the status and functions of certain things.

Under this view, the *glue* holding an organization together is **intersubjective belief**.

Organizational culture, mission, and identity are continuously created and re-created in the minds of employees, leaders, and stakeholders. This leads to a key philosophical premise of IDT: **organizational reality is fluid**, sustained by ongoing communication and influence among individuals rather than by any static material structure. The fiction must be actively maintained; hence the beliefs and behaviors of organizational members can **drift** over time as influence flows shift. If new stories take hold or if leaders and employees begin to behave inconsistently with the old narrative, the organization’s *imagined order* is altered or even undermined. For instance, a company founded on a noble mission may gradually devolve into one fixated on profit at all costs if successive influences (e.g. new management practices, peer pressures) slowly change what employees perceive as acceptable behavior.

This perspective grounds IDT in a *social constructionist* philosophy. It acknowledges individual agency – people can make choices and exert influence – but emphasizes that agency is exercised within an **organizational social network**. Each person’s understanding of “the way we do things here” is shaped by interactions with others. In essence, organizations are **self-perpetuating stories** that live in an ecology of minds; IDT is concerned with how the content of those stories drifts through interpersonal influence. This focus on continuous change and collective cognition differentiates IDT from theories that regard organizational design or culture as something that can be engineered top-down and then remains static. Instead, IDT aligns with the notion that stability in an organization is often an *active* equilibrium – a dynamic balancing act of influences – rather than a permanent state.

Classical Organizational Theories in Contrast

To appreciate the contribution of Influence Drift Theory, it is important to contrast it with the major schools of thought in organizational theory. These classical and modern theories provide valuable insights but generally do not capture the *micro-dynamic* of gradual interpersonal influence that IDT emphasizes.

- **Classical Management Theory:** Emerging from the Industrial Revolution’s context of factory production and bureaucratic administration, classical theory focuses on formal structure, specialization, and efficiency. Thinkers like Frederick Taylor (scientific management) and Max Weber (bureaucracy theory) sought to optimize organizations as machines: defined hierarchies, clear rules, division of labor, and rational-legal authority. Weber’s ideal bureaucracy is characterized by impersonal roles, a strict chain of command, and merit-based advancement. The underlying assumption is that workers are economically motivated and can be guided by systematic coordination and control. While classical structures did improve efficiency, they were often critiqued for their rigidity and their neglect of human variability. Indeed, the classical perspective has been **“criticized as ignoring human needs”**, treating employees as cogs in a machine rather than individuals. It assumes that if the formal design is right, organizational behavior will follow in a predictable way – leaving little room for the subtle influences of morale, peer example, or leadership character on which IDT focuses.
- **Human Relations and Behavioral Theories:** By the 1920s and 1930s, it became clear that the classical approach was incomplete. The famous Hawthorne studies

(1924–1932), led in part by Elton Mayo, revealed that workers' productivity was affected not just by physical conditions or pay, but by psychological and social factors – notably, the feeling of being observed and valued. The **human relations movement** shifted attention to informal organization, employee morale, group dynamics, and leadership style. Researchers in this vein (Mayo, Maslow, McGregor, and others) emphasized that organizations are also social systems: factors like trust, recognition, and a sense of belonging can significantly impact performance. The “neoclassical” perspective acknowledged workers' needs and the importance of **morale and leadership** in achieving organizational goals. In essence, it opened the door to thinking about influence – a sympathetic supervisor or a cohesive team could inspire greater effort than any rule or incentive. However, human relations theory largely remained at a *descriptive* level; it did not formalize how exactly attitudes spread or how culture forms beyond noting that supportive relationships breed positive outcomes. From an IDT viewpoint, human relations theory provides the rationale for considering social influence, but IDT seeks to *model* it explicitly (e.g. how one enthusiastic team member's attitude might uplift others in a gradual contagion-like process).

- **Open Systems Theory:** By the 1950s and 1960s, scholars began viewing organizations as **open systems** interacting with a larger environment. Katz and Kahn (1966) and others applied general systems theory to organizations, highlighting input–throughput–output processes and the need for feedback. An organization imports resources (material, human, informational), transforms them, and exports products or services. Importantly, it must also deal with feedback and dynamic equilibrium. Systems theorists introduced concepts like **entropy** (the tendency of closed systems to run down) and argued that organizations survive by importing more energy (resources, information) than they expend – achieving *negative entropy* through constant renewal. They also noted **dynamic homeostasis**, the idea that subsystems must maintain balance as they adapt, and **equifinality**, meaning there are many different ways (paths) to reach the same outcome. These ideas resonate with IDT: they imply that organizations are *not static*, and outcomes are not predetermined by one “best way.” However, traditional systems theory tends to focus on macro-level flows and may treat culture or human behavior as just one subsystem. It acknowledges complexity but doesn't zoom in on how individual interactions produce emergent system properties. IDT can be seen as drilling down into the “black box” of the transformation process: the *social energy* that flows

within the system. Whereas systems theory might say an organization must process information and feedback to remain stable, IDT would ask: how do the people in the organization actually alter their beliefs and decisions in response to that feedback? Systems theory provides a holistic view; IDT provides a micro-mechanism for one crucial aspect of the system – the evolution of shared understandings through interpersonal influence.

- **Contingency Theory:** By the 1970s, research had shown that no single organizational structure is best for all situations. **Contingency theory** posits that the optimal way to organize “depends” on various factors – environment uncertainty, technology, size, etc. In other words, **there is no one universal or best way to manage; effective organization must fit its context.** Burns and Stalker (1961) famously distinguished *mechanistic* structures (suited for stable conditions) from *organic* structures (suited for dynamic, innovative environments). Lawrence and Lorsch (1967) showed that different departments in a company might need different structures depending on their tasks and external demands, but these must be balanced (integrated) for the firm to function. Contingency theory essentially adds *environmental influence* as a determinant of organizational form and success. This has an affinity with IDT in the sense that IDT can incorporate external influences as drivers of internal change (e.g. new market conditions changing the mindset of key employees, who then influence others). But again, contingency theory tends to be *static* in its formulations: it says “match your structure to your environment,” implying a comparative statics view where an organization shifts from one equilibrium design to another. It does not explain *how* the organization makes that transition or how internal dynamics play out during the adjustment. Influence Drift Theory can provide a complementary perspective: even once structure is matched to environment, success will hinge on how quickly and effectively the people within adopt the needed behaviors or values, which is an iterative social process. Additionally, IDT’s emphasis on *flexible parameters* can be seen as a contingency approach within the model – by tuning feedback loops or hierarchy in simulations, one can mimic different contingency fits (e.g. a rapidly changing environment might be best handled by an influence network with strong lateral connections and feedback, akin to an “organic” design).

In summary, classical and modern organizational theories offer crucial insights: the importance of formal structure and efficiency (classical), the role of human motivation and group dynamics (human relations), the complexity of interdependence and environment (systems), and the situational nature of management (contingency). What they generally lack is an explicit account of **how individual-level changes aggregate into organizational-level outcomes over time**. Organizational culture theory and studies of organizational change begin to tackle this by observing how, for example, leadership style trickles down or how informal networks operate. Yet, these are often qualitative or case-based. **Influence Drift Theory aims to fill this gap by providing a process theory of organizational change**, rooted in interpersonal influence. IDT does not replace classical theories – rather, it builds upon them. One can think of IDT as operating *within* the scaffold of structure and environment that other theories describe: given a certain hierarchy (classical) and workforce attitudes (human relations) in a certain market context (contingency), IDT tells us how things might evolve month by month as people interact. The next sections develop the framework and mechanics of this theory in detail.

Influence Drift Theory (IDT): Concepts and Mechanics

Definition: *Influence Drift Theory* conceptualizes an organization as a network of individuals where each person's behaviors and beliefs are continuously shaped by the influences emanating from others. Over time, these myriad micro-interactions lead to a **drift** in the collective state of the organization – for better or worse. Unlike theories that see change as a series of discrete decisions or external shocks, IDT views change as an *emergent, cumulative process*. The term “drift” implies a gradual shift that may go unnoticed until it produces significant effects. IDT formalizes the idea that, for instance, a company's ethical climate or a team's level of innovation is not fixed, but is constantly evolving as individuals influence one another in subtle ways every day.

Core Principles of IDT

Several key principles underlie Influence Drift Theory:

- **Organizations as Influence Networks:** At its heart, IDT treats the organization as a **social network** of influence. Each member is a node, and each directional interpersonal connection carries a weight representing how much one person can sway another's opinions or actions. This includes formal influence (e.g., a manager's authority over

subordinates) and informal influence (e.g., a veteran employee's peer influence or a charismatic individual's moral sway). The network perspective recognizes that influence does not only flow top-down; it can be bottom-up and horizontal as well. The structure of this influence network – who is connected to whom and with what strength – is a fundamental determinant of how the organization evolves over time.

- **Gradual Interpersonal Influence and Update:** Individuals in an organization do not typically change their mind or behavior instantaneously on a whim; instead, change is **iterative and marginal**. IDT posits that each person periodically updates certain internal states (beliefs, attitudes, level of effort, etc.) based on what they observe in others. This is akin to a learning or averaging process. Social psychologists have long studied such phenomena: for example, *social influence network theory* in small groups shows that individuals tend to weigh others' positions and adjust their own stance gradually, which can lead to group consensus or persistent disagreement depending on the influence network structure. IDT builds on this, suggesting that *every aspect* of organizational behavior (not just opinions on an issue) can drift via similar mechanisms. A worker might slightly increase their work pace after seeing peers working hard, or an executive might slowly adopt a more aggressive strategy after months of interacting with risk-taking colleagues. Change is modeled as a continuous trajectory rather than a jump to a new equilibrium.
- **Multi-Dimensional Influence:** Influence in organizations is not a single scalar quantity; people influence each other on many **dimensions**. IDT therefore considers multiple metrics or state variables for each individual. Examples of such *influence metrics* include **ethical standards, competence or skill level, trust** (in colleagues or leaders), **commitment** to the organization, and even things like risk appetite or openness to innovation. Each metric can drift due to social influence. For instance, an employee's ethical behavior might be influenced by the observed behavior of their manager (if the manager bends rules, the employee's own moral stance might erode over time), while that same employee's technical competence might be improving simultaneously by learning from a skilled coworker. IDT allows for analyzing each of these dimensions separately or in combination. In practice, some dimensions will interact – e.g., one might only be influenced by someone they trust, linking the trust metric with the influence on other metrics – but the core idea is to acknowledge the rich tapestry of what can be

influenced in an organizational setting.

- **Endogenous Emergence vs. Exogenous Shocks:** In IDT, *endogenous* interpersonal influences are the primary drivers of change. While traditional analyses often look for singular causes or external triggers for organizational change (a new CEO, a market crisis, a policy change), IDT directs attention to the slow-building internal momentum. This does not mean external factors are irrelevant – on the contrary, external events often act by altering internal influence patterns. For example, a new regulation might empower a compliance officer’s influence internally, or a market downturn might cause a shift in who is listened to within a company (perhaps operations experts gain more sway than marketing experts during a cost-cutting period). But crucially, IDT suggests that even in the absence of obvious external events, organizations can undergo significant change purely through internal dynamics. Over years, an organization might “drift” into having a very different culture or strategy than it started with, due to turnover of individuals and continuous social influence among the remaining members. This focus on endogenous change is influenced by observations in failure analyses: in disasters like the Challenger Space Shuttle accident or corporate scandals, investigators often find not one moment of irrational decision, but a series of small steps – a *normalization of deviance* – where standards eroded bit by bit. IDT provides a vocabulary for such cases: the influence network allowed each step to become the new normal, drifting the collective baseline.
- **Path-Dependence and Non-Linearity:** Because influence dynamics are iterative, the order and history of changes matter. IDT implies that organizational outcomes are **path-dependent**. Two organizations with identical structures and initial states could end up differently if the sequence of interpersonal influences differs. This is in contrast to simpler models that might predict only one equilibrium outcome given initial conditions. Path-dependence arises naturally in influence networks especially if there are feedback loops. There may also be **non-linear** effects: small differences in early influence (who convinces whom early on) can lock in certain trajectories (akin to “butterfly effects” in complex systems). In network terms, early adopters of a behavior can tip the balance if they influence others before an alternative behavior spreads. IDT therefore views organizational change as a **process** where timing and network position can create tipping points or inertia. One practical upshot is that interventions (like introducing a new

policy or training) might have very different outcomes depending on how they are introduced into the social fabric – who champions them first, and how the influence cascades.

Modeling

Mathematical Structure of the Influence Model

The Influence Drift Theory (IDT) operates on a simple, rule-based hierarchical structure, where interpersonal influence is modeled through **fixed scalar updates** between leaders and their subordinates. This model avoids complex matrix formulations and instead adopts an intuitive, layered influence mechanism that reflects real-world hierarchical organizations.

1. Organizational Structure

The organization is modeled as a **5-level pyramid** with strict hierarchical relationships. Each leader supervises exactly 4 subordinates, forming a **341-member organization** structured as follows:

Level	Number of Individuals	Description
1	1	Top leader (root)
2	4	Direct reports to leader
3	16	Middle managers
4	64	Frontline managers
5	256	General members
Total	341	

This "Rule of Five" was inspired by a line from Naruto's Jiboro: "When five people gather, there is bound to be one who is trash." Although it is an extreme expression in the story, it means that

in a group of five, one person may show an outlier in behavior, and it is the smallest organization where an outlier can appear from the average.

2. Trait Initialization

Each individual is assigned an initial feature score $S \in [0, 100]$, which represents characteristics such as ethics, competence, and ideology. The values are randomly drawn from a normal distribution centered at 50 (neutral) with an appropriate standard deviation that reflects the natural fluctuations of the population. There are two important points in this theory. First, the distribution of organizations is randomly composed regardless of status. This reflects the fact that leaders in organizations are not selected solely based on goodness or competence. Second, all people are composed based on a normal distribution. In fact, it would naturally be a power-law distribution with a higher center than a normal distribution. Most people are good or bad, incompetent or competent to a very small extent. Here, a normal distribution is assumed for general modeling.

- 0 = very negative (e.g., unethical, incompetent)
- 50 = neutral, but it is a non-existent value.
- 100 = Very positive (e.g., very ethical, competent)

3. Influence Mechanism

IDT assumes **bidirectional influence** between each leader and their 4 direct subordinates, but **asymmetrically weighted** to reflect the greater authority of leaders in most organizations.

3.1. Top-Down Influence (Leader → Subordinates)

Each subordinate adjusts their trait based on their leader's current trait using the following rule:

$$S_{\text{subordinate}} += \frac{S_{\text{leader}} - 50}{10}$$

- If the leader's trait is higher than 50, the subordinate moves upward (positive influence).

- If the leader's trait is lower than 50, the subordinate moves downward.

Example:

Leader score = 70 → +2 point increase in subordinate

Leader score = 30 → -2 point decrease in subordinate

3.2. Bottom-Up Influence (Subordinates → Leader)

Each leader is also influenced by the average trait of their 4 direct reports, but with a smaller effect:

$$S_{\text{leader}} += \frac{S_{\text{subordinates}} - 50}{20}$$

- Reflects how leaders receive feedback and how they are (indirectly) influenced by team norms.
- The influence is half the top-down direction.

Example:

Subordinate average = 60 → +0.5 increase in leader

Subordinate average = 40 → -0.5 decrease in leader

This amount of change in influence is based on the fact that the team atmosphere is affected in small teams, and the amount of change can be adjusted depending on whether the organization is horizontal or authoritarian.

4. Update Process (Feedback Cycles)

The entire organization undergoes **10 discrete update iterations** (feedback cycles). At each cycle:

1. All individuals compute their updated trait score based on their leader/subordinates.
2. All updates are applied simultaneously (synchronous update).
3. Scores are clipped to stay within the [0, 100] range.

The 10 feedbacks are an arbitrary number. This is a feedback that can arise from the command/report system, and this number was applied considering that the organization's disposition can change drastically when there is a lot of feedback in the simulation where the amount of change is assumed as the default. You can observe the amount of change by increasing or decreasing this number, and you can also see how it changes in individual organizations, not just the amount of change in the entire organization.

5. Simulation Scaling

To account for variation and test robustness, the simulation is **replicated across 100 independent organizations**, each with:

- Identical hierarchical structure (341 members)
- Independent random initial trait distributions

6. Outcome Analysis

After 10 iterations, key outcome variables are analyzed:

- Average trait shift across the organization
- Variance compression or expansion
- Emergence of convergence, divergence, or cluster patterns
- Sensitivity to initial leader scores

Result

Central Tendency Over Time (Mean vs. Median)

Across the organizations, the overall mean and median trait scores remained roughly stable at ~50 throughout the 10 iterations. In other words, there was no systematic drift in the population average trait – the ensemble mean started at 50.0 and ended at 49.9 (a negligible 0.2% decline), and the median similarly stayed ~50. Figure 1 illustrates this stability: the mean (solid line) and median (dashed) virtually overlap around 50 each iteration. This outcome is expected because positive and negative deviations (leaders above or below 50) tend to balance out across many randomized organizations. Notably, while each organization's internal average

may shift (see below), the aggregate median/mean stays near the baseline, implying no overall bias in trait movement. The gray band in Figure 1 shows the middle 80% range of organization means – initially very narrow around 50, but widening considerably by iteration 10 (spanning roughly 47 to 53). This indicates that although the global average is steady, between-organization differences grow over time (some groups drift higher, some lower). In sum, IDT’s influence updates do not alter the grand mean, but they increase variability across organizations

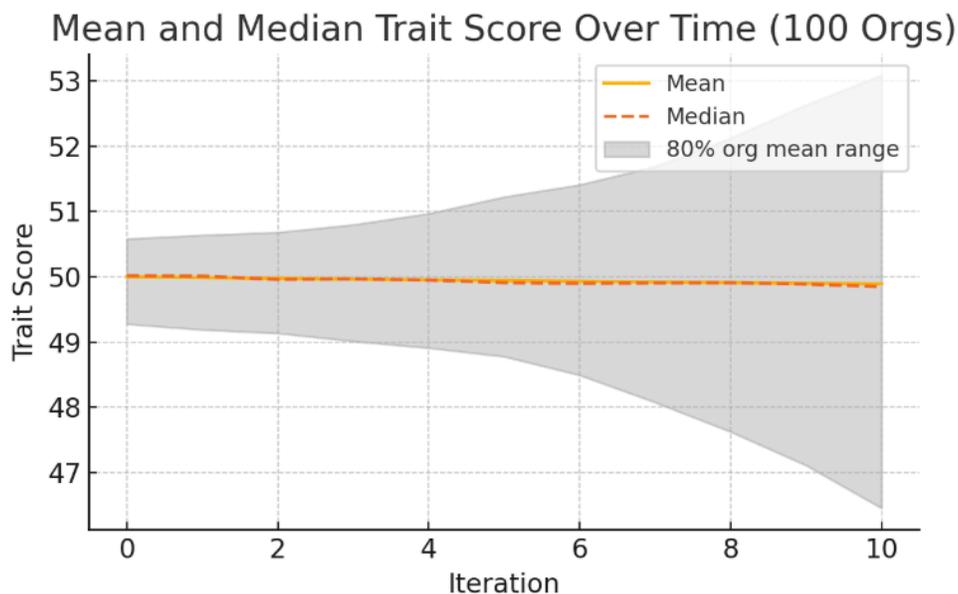


Figure 1: Mean and median trait scores over 10 iterations (100 organizations). The overall mean (orange solid) and median (red dashed) remain near 50 (baseline) at all times. The shaded region shows the 10th–90th percentile range of organization-average trait scores, which broadens over time – reflecting divergence between organizations even as the global central tendency stays constant.

Trait Shifts by Hierarchical Level

We next examined how different **hierarchical levels** responded to these influence dynamics. A clear pattern emerged: **lower-tier individuals’ traits shifted far more over time than those of higher-tier leaders.** Figure 2 plots the average absolute change in trait (from the initial value) for each tier level over the 10 iterations. The top leader (Tier 0) is relatively stable – changing on average only about 2.5 points after 10 updates – whereas those in lower tiers exhibit much larger shifts. By the final iteration, Tier 3 members (second-lowest level) changed by ~10.5

points on average, and Tier 4 (lowest subordinates) by ~9.5 points, compared to only ~8.3 points for Tier 1 and ~9.9 for Tier 2. In other words, **mid- and lower-level employees' traits drifted about 3–4 times more than the top leader's trait**. Lower tiers also began to move **sooner and faster**. In the first few iterations, subordinates already showed larger deviations than leaders – e.g. by iteration 3, Tier 3 individuals had shifted ~3 points versus ~0.8 for the top leader. This occurs because top-down influence propagates cumulatively: a CEO's trait difference influences division heads, which then influence managers, and so on, **amplifying changes down the chain**. Additionally, lower tiers lack further subordinates to “pull” them back, so they fully absorb the leader's influence each round. Higher-ups, in contrast, are buffered by the weaker bottom-up feedback (discussed next). The result is a pronounced gradient: **the farther down in the hierarchy, the larger and quicker the trait shift**. By iteration 10 the Tier 3 group experienced the greatest change, slightly exceeding even Tier 4 – an effect of being pushed from above *and* pulled by their subordinates. Overall, lower-level members' traits are far more volatile under IDT updates, while top leaders remain comparatively steady in their initial dispositions.

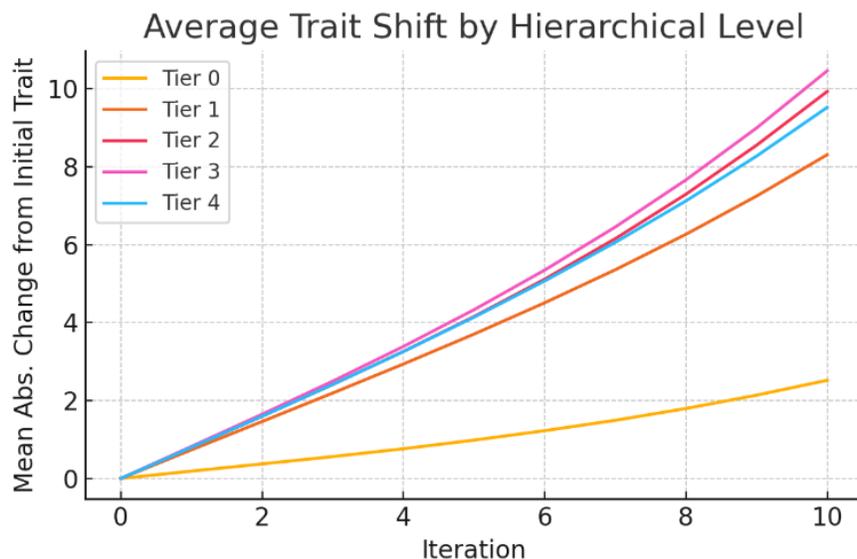


Figure 2: Average trait shift (absolute change from initial score) by hierarchical level over time. Lower tiers (Tier 2, 3, 4) exhibit substantially larger and faster shifts than upper tiers. By the 10th iteration, a typical Tier 3 member's trait has changed by ~10 points, versus only ~2.5 points for the Tier 0 leader. This indicates stronger influence accumulation at lower hierarchical levels, whereas top leaders' traits remain relatively close to their start values.

Top-Down vs. Bottom-Up Influence Dynamics

The simulation allows us to quantify the **relative contributions of top-down vs. bottom-up influence** in shifting traits. We found that **top-down influence dominates**: leaders exert a stronger effect on subordinates than vice versa. On average, a subordinate's trait changed by about 0.8 points (absolute) per iteration due to their leader, whereas a leader's trait changed only ~0.2 points per iteration due to all four subordinates (in early iterations). **Figure 3** compares these effects over time. The orange line (leader → subordinate influence) is consistently well above the red line (subordinates → leader). In the first round, top-down changes were nearly 4 times larger than bottom-up changes on a per-person basis. Even by iteration 10, as bottom-up feedback accumulated, a subordinate's average change (~1.3) was still about 2.7× the magnitude of a leader's change (~0.48). In terms of total impact, we estimate roughly **90–95% of the trait movement is driven by top-down influence** at each step, with bottom-up effects accounting for only ~5–10%. This asymmetry reflects the hierarchy's structure and update coefficients: a leader's deviation from 50 is directly applied (10% of it) to each subordinate, whereas a subordinate's deviation only weakly nudges the leader (5% of the average deviation). Our findings align with leadership "trickle-down" theories, which argue that leaders' traits and behaviors induce corresponding changes in followers. Indeed, we observe subordinates' traits drifting to mirror their leader's level (e.g. high-scoring leaders pull their teams upward, low-scoring leaders drag their teams lower). The bottom-up influence, while present, is comparatively muted – it tempers extreme leader traits slightly but does not reverse the direction of change. For instance, if a manager's four subordinates all rose well above 50, the manager's own score increased marginally in response, but not enough to overpower the top-down effect from *their* superior. Over time, bottom-up influence does grow somewhat (the gap between the lines in **Figure 3** narrows slightly by iteration 10 as subordinates become more extreme), yet **top-down effects remain the primary driver of trait change** in the hierarchy. This result highlights a key aspect of IDT in practice: influence predominantly flows downward through the ranks, with limited upward feedback.

Relative Magnitude of Top-Down vs Bottom-Up Influence

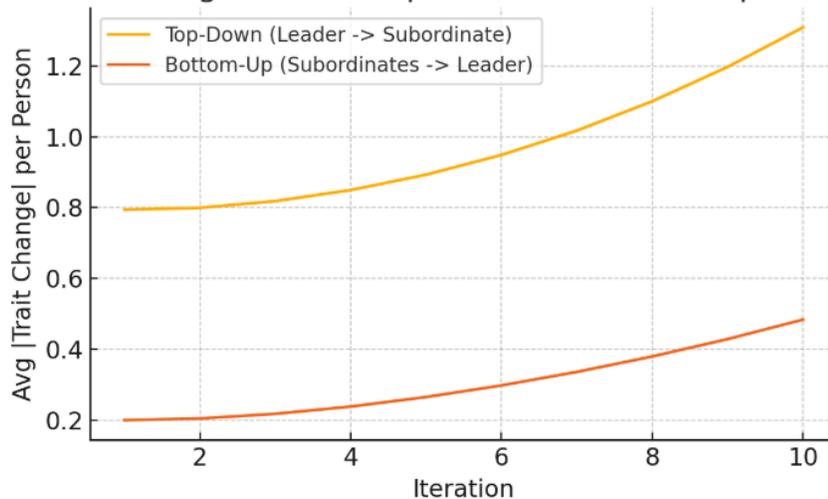


Figure 3: Relative magnitude of top-down vs. bottom-up influence per iteration. The average absolute change caused by a leader on each subordinate (top-down, orange) far exceeds that caused by subordinates on their leader (bottom-up, red). Top-down influence accounts for the bulk of trait shifts – for example, in iteration 1 a subordinate’s trait changed ~0.8 points from leader influence, versus only ~0.2 points change in a leader’s trait due to subordinates. Even after 10 iterations, the leader→subordinate effect is ~3× larger. This confirms that hierarchical influence is highly asymmetric, with **strong top-down impact** and much weaker upward adjustments.

Distributional Changes and Clustering

The **distribution of trait scores** across all individuals broadened markedly under the influence updates. Initially, trait scores were tightly clustered around the mean of 50 (by design, a normal distribution). After 10 iterations, the spread of scores increased dramatically, indicating growing diversity and polarization of traits across the entire population. The variance of trait scores across all 34,100 individuals (341 per org × 100 orgs) rose from ~100 initially to ~273 after iteration 10 – nearly a threefold increase in dispersion. **Figure 4** compares the trait distribution at the start vs. the end of the simulation. The initial distribution (left) is a narrow bell curve centered at 50, with most values in the 40–60 range. By iteration 10 (right), the distribution is much flatter and wider. While still roughly centered near 50 (dotted line) – reflecting the stable global mean – it has **much heavier tails**. About 27% of individuals scored below 40 or above 60 at the end, compared to only ~16% initially. In fact, a non-trivial fraction of the population

migrated toward the extremes of the scale: ~3.5% ended up scoring in the 0–20 range (vs only 0.14% initially), and similarly ~3.4% reached 80–100 (vs 0.13% initially). This indicates **increased polarization**: some groups’ traits drifted to very high values, others to very low values. The distribution remains fairly symmetric (skew ≈ 0), with roughly balanced high and low extremes – consistent with the random mix of initially high vs. low leaders across the 100 organizations. We do not yet see a clear bimodal shape after 10 iterations (the final histogram still has a single broad peak around 50), but the trend suggests movement toward a bimodal or even *bimodal-clustered* outcome if the process continued longer. In each organization, members’ traits started to **cluster around a new mean** (set by their leader’s influence) rather than the original 50. Within-group variance actually increased in these early iterations (as some individuals moved faster than others), but eventually one expects it would **decrease** as members converge toward a common extreme value (bounded at 0 or 100). In effect, the population is segmenting into distinct high-trait and low-trait clusters aligned with their organizational units. By iteration 10, this segmentation is evident in the widening gap between different organizations (recall **Figure 1**’s gray band) and the fattening of the tails in the overall distribution. We can think of each organization’s culture “drifting” either upward or downward from the norm – a hallmark of IDT. In summary, **trait score distributions became substantially more dispersed over time**: the variance surged, the tails grew heavier, and individuals increasingly clustered by organization. This suggests that in a hierarchical influence process, even if the average trait remains the same, the society can become **more polarized and differentiated internally**, with pockets of very high or very low trait groups forming.

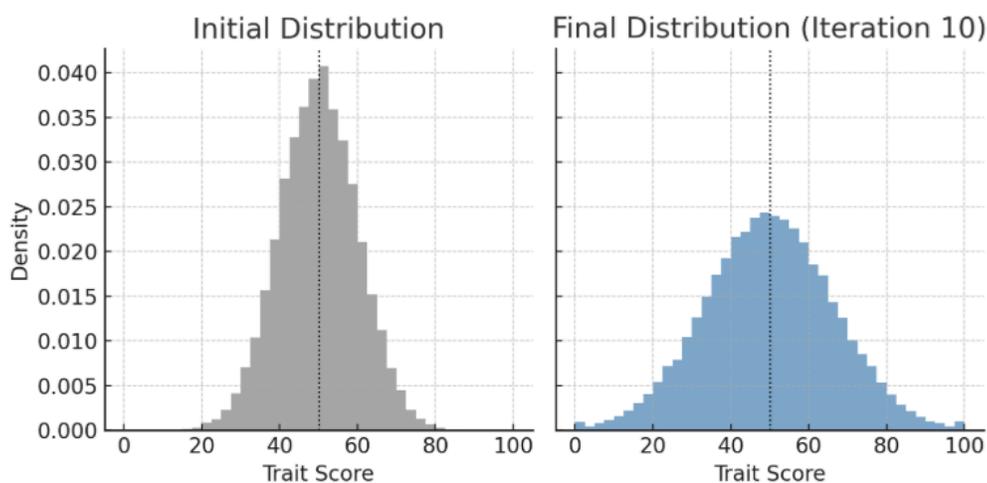


Figure 4: Trait score distribution for all individuals initially (left) vs. after 10 iterations (right). Initially, scores are tightly centered around 50 (s.d. ≈ 10). After the influence updates, the

*distribution has **spread out** significantly – the peak at ~50 has lowered, and much more probability mass resides in the tails (near 0 and 100). The vertical dotted line marks the baseline 50. The final state shows many more individuals with extreme trait values (e.g. scores <30 or >70) than the bell-curve initial state, indicating rising polarization and clustering of traits into high and low groups. Despite this dispersion, the distribution remains roughly symmetric about 50 (no strong skewness), as upward and downward drifts occurred in balance across the 100 organizations.*

These parameters – hierarchy vs. feedback, team cohesion, influence asymmetry, etc. – are **tunable knobs in IDT simulations**. By adjusting them, we can emulate classical scenarios. For example, a **highly bureaucratic, classical organization** would be simulated with strong hierarchy (many levels, top-heavy influence), low feedback (almost one-way), and perhaps medium team cohesion (departments have some solidarity). A **high-involvement, participative organization** would have shallow hierarchy, high feedback (even juniors can influence seniors), and cross-functional team links (preventing silos). A **networked organization** (like an open-source community) might have very fluid influence where expertise determines influence more than formal role – represented by dynamic, context-specific weightings.

The IDT framework doesn't prescribe which configuration is best; rather it provides a way to *analyze* the consequences of each. It can also handle changes in these parameters over time (e.g., what happens if a company restructures from hierarchical to team-based – how does that alter the influence dynamics and thus the drift of various metrics?).

Influence Metrics and Their Evolution

IDT is especially interested in qualitative factors that are often hard to quantify in traditional models: **ethics, competence, and trust** are prime examples given. We treat each of these as a metric that can diffuse or drift in the influence network:

- **Ethics (Moral Standards):** Consider an “ethics level” for each individual, representing their likelihood to behave honestly vs. dishonestly, or their adherence to the organization's code of conduct. Initially, people come with personal morals, but the organizational culture can raise or lower these. If employees observe colleagues and leaders acting with integrity and see ethical behavior rewarded, their own ethical stance may be reinforced or even elevated (through admiration and imitation). Conversely, if they see cheating, cutting corners, or unethical decisions go unpunished (or succeed),

their resistance to those behaviors may erode. Influence on ethics can be subtle: unlike a skill that you can learn in a training, ethics is often transmitted through norms and the perceived “tone at the top.” **IDT allows simulation of ethical climate change.** For example, suppose a new CEO arrives who is less scrupulous than her predecessor. She might directly influence her direct reports to take aggressive actions that test ethical boundaries. Those subordinates, seeing the CEO’s behavior and perhaps feeling pressured to deliver results, might shift their own comfort level with ethically gray practices (a small drift). They in turn manage large teams, and through both explicit directives and implicit signals (what they tolerate, what they reward), they influence the next layer. Over time, a company that once prided itself on honesty might drift toward a culture of “winning at any cost.” This is precisely what many corporate scandal post-mortems find – not a sudden malicious turn, but a gradual normalization of deviance. A recent example is the Volkswagen emissions scandal, where a **culture of solving problems quickly and meeting targets** (without adequate ethical reflection) cascaded into outright cheating on emissions tests. Engineers and managers did not wake up one day deciding to commit fraud; rather, a long-standing pressure and influence from the top created an environment where more and more people became complicit or silent over unethical practices. IDT would model this as a drift: each person lowered their ethical threshold slightly under influence from superiors and peers, eventually resulting in systemic misconduct. On the positive side, IDT can also model *ethical improvement*. If a highly ethical leader is installed (with strong influence weight) and if mechanisms of accountability and open discussion are introduced (increasing feedback loops), one can simulate a gradual rise in the average ethics metric. Empirical studies support this dynamic – for instance, research shows **ethical leadership has a positive, significant impact on employees’ ethical behavior**. In other words, when leaders consistently model and enforce ethical conduct, employees adapt and the overall ethical climate improves via influence and example.

- **Competence and Knowledge:** Competence (or skill, expertise) might seem like a more individual trait, but it too is subject to social influence in organizations. One obvious channel is **knowledge sharing** and peer learning: a highly competent team member can train or mentor others, effectively raising their competence over time. Alternatively, a group of high performers might influence a new member to step up their game (through both direct coaching and the newcomer’s desire to fit into a high-achieving norm).

Conversely, in an environment where many people are under-skilled or complacent, a skilled individual might either leave (if they're not influenced by others, they may be alienated) or could even lose their edge by lack of challenge – essentially “drifting down” to match a lower norm (sometimes called **social loafing** or de-skilling in a lax environment). IDT can incorporate competence as a metric that diffuses: if one person learns a better technique, those closely connected to them (high influence weight) may adopt that technique, improving their competence. Over time, best practices can percolate through the network – or bad practices can as well. Competence influence can be asymmetric: usually junior staff increase competence by learning from seniors (so seniors have high influence on juniors' competence), though reverse mentoring is possible (especially in areas like technology where younger employees might tutor older ones). Simulation of competence drift would involve initial distributions of skill and seeing how quickly the group equalizes or improves collectively given the interaction network. Highly siloed organizations might see little competence diffusion (each unit's capability stays separate), while collaborative cultures might rapidly spread innovations (steep learning curves through influence). Competence also has external feedback: if the organization faces new challenges, those who adapt first might influence others to upskill. If the organization does not have those adaptive influencers, it may drift into obsolescence competence-wise (everyone collectively sticks to old skills).

- **Trust:** Trust is a complex metric because it is inherently relational – A's trust in B might differ from B's trust in A. One way to include trust in IDT is to focus on a person's general **trustworthiness** as perceived by others, or their general propensity to trust colleagues and the organization. Trust is critical for influence: if I do not trust someone, I won't be influenced by them easily. Trust itself evolves: people gain trust in others by positive interactions and lose it through betrayals or disappointments. In organizations, a climate of trust means people are willing to be influenced and to collaborate without fear. A climate of distrust means walls go up – information is hoarded, and influence network links might as well be cut. IDT can capture trust dynamics by having a separate trust metric that is updated based on experiences. For example, if a manager consistently treats an employee fairly and supports them, the employee's trust in the manager increases; that might reflect as an increasing weight of the manager's influence on that employee (so influence weights aren't fixed but evolve with trust). Meanwhile, if a leader breaks promises or behaves inconsistently, trust from subordinates declines – even if

formally the hierarchy is the same, the actual influence weight might drop (people start ignoring the leader or seeking guidance elsewhere). At a group level, **trust propagation** could happen if, say, someone vouches for someone else. Imagine a new CEO comes in – employees may be skeptical initially (low trust), but a well-respected middle manager who knows the CEO from before might assure their team that the CEO is reliable. The team's trust in the CEO can then ramp up faster due to the influencer (the middle manager's testimonial). This is second-order influence: Person A influences B's trust in C. While complex, these dynamics often appear in social networks (e.g., "transitive trust"). In IDT, we could simplify by tracking a general trust climate: e.g., an average trust in management metric. Influences that improve transparency, fairness, and communication tend to drift that metric upward, whereas influences that introduce fear (like witnessing someone punished unfairly) drift it downward. Trust is also connected to feedback loops: in a trusting environment, upward feedback is stronger (people speak up) thus influencing leaders. In a distrustful one, upward influence ceases (no one dares question), exacerbating any drift because corrections aren't communicated.

- **Other Possible Metrics:** Beyond the given examples, IDT could be applied to any quantifiable attitude or behavior: **engagement/motivation, innovative thinking, risk tolerance, job satisfaction, adaptability to change**, etc. Each of these can be thought of as contagious to an extent. Motivated employees can energize others; demoralized employees can bring others down. A risk-taker in a key position might slowly encourage the whole department to take more risks (for good or ill). The evolution of these factors often determines long-term organizational outcomes like productivity, turnover, creativity, or ethical standing.

The **simulation** aspect of IDT comes into play when we use a computational approach (or analytical, if possible) to see how these metrics evolve. Typically, one would set up initial values for each metric for each individual (perhaps based on survey data or hypothetical scenarios). Then at each time step, apply the influence update rules to get new values. By running this forward (iterating), we generate trajectories for each individual and the organization as a whole. We can measure outcomes such as: Does the group converge to a high or low value on a metric? How long does it take to stabilize? Who ends up influencing the outcome the most (which could be analyzed by looking at how the final state relates to initial states – in a linear

model, it's a weighted average of initial states, and those weights can be interpreted as **influence centrality** of each person)? Are there oscillations or cycles (perhaps in trust – e.g., trust dips after an incident, then slowly rebuilds)? These results can be analyzed statistically or visually.

One might also introduce external shocks in a simulation to see how the network responds: for example, at time T, suddenly leader's ethics drops (maybe due to a scandal). How far and how quickly does that ripple through the network given the weights? Or, inject a new high-competence person into a team and simulate knowledge diffusion.

What distinguishes IDT from a mere anecdotal discussion of these issues is that it urges a *systematic*, model-based examination. By specifying numerical influence strengths and initial conditions, we can move beyond saying "leadership matters" to estimating *how much* and under what conditions leadership matters, for instance. If one leader's influence is diluted by very strong peer networks that promote opposite behavior, the peer effect might win out. In another context, a leader placed in a structurally central position could almost single-handedly shift the organization's path (for good or bad).

Continuous Influence vs. Binary Decision Models

Influence Drift Theory's approach to organizational change stands in contrast to the binary, static models often used in economics and game theory. It is instructive to compare these paradigms:

Traditional **game theory** models interactions in terms of *choices* – typically discrete choices like cooperate vs. defect, adopt strategy A vs. strategy B. Each actor is assumed to have stable preferences (e.g., a payoff matrix) and to make a rational choice that maximizes their utility given expectations of others' choices. Game theory has been applied to organizations to analyze, for example, principal-agent problems, coordination games among departments, or voting in committees. While powerful, these models usually simplify the internal state of actors to a payoff function and perhaps belief about others. They generally do not account for the notion that preferences or values themselves might change via interaction. Instead, if repeated interaction is considered, it's often through concepts like **equilibrium** – players might eventually converge to a stable strategy profile (Nash equilibrium) after some learning or adaptation.

IDT vs. Game Theory can be summarized on a few dimensions:

- **Nature of Change:** In IDT, change is **gradual and incremental**. There is a continuum of possible states (e.g., one can be 70% trusting, then 72%, then 75% – it's not just trust or distrust as a binary). In many game models, change is implicit (players might switch strategies between discrete rounds, which is a sudden jump). Some game-theoretic extensions like **evolutionary game theory** introduce gradual change, but that typically refers to population fractions (e.g., proportion of agents using a strategy) rather than an individual's propensity in continuous terms. IDT is closer to a **dynamical system**, possibly even allowing calculus-based descriptions (in the limit as time steps become small, one could write differential equations for opinion change).
- **Binary vs. Continuous Decisions:** Consider a scenario of cooperation in a team project. A binary model might frame it as each member either contributes effort or shirks. Influence would then be captured by something like repeated Prisoner's Dilemma games – you cooperate if enough others cooperated last time or if you predict they will. IDT, however, would say each member can adjust their **level of effort** continuously by observing others. If others work slightly harder, I might also work a bit harder tomorrow (not an all-or-nothing choice). Over time this could escalate everyone's effort (a virtuous cycle) or decline (a race to the bottom) or find an equilibrium. In effect, IDT can model *degrees* of cooperation, whereas classical game theory often focuses on the existence of cooperation vs. defection as distinct outcomes.
- **Rationality Assumptions:** Game theory typically assumes rational actors with clear objectives (maximize payoff). Influence Drift Theory does not start from utility maximization. Instead, it focuses on *social adaptation*: people might change to be more like those around them without a conscious calculation of long-term payoffs. Of course, one could derive influence behavior from bounded rationality (e.g., if I assume my peers know something I don't, it's rational to shift toward their opinion – a kind of social learning). But IDT is agnostic about the reasons; it could encompass unconscious mimicry, conformity pressure, or simple learning. The **worldview** critique by Dekker is relevant here: searching for rational causes and “bad actors” can be misleading in complex systems. IDT embodies a *non-blame* perspective – outcomes are systemic, arising from many small interactions rather than single decisions. In game theory, we often pinpoint equilibrium strategies or pivotal players; in IDT, influence is diffuse.

- **Equilibrium vs. Trajectory:** Game theory analyses often focus on equilibria (Nash equilibrium, Pareto optima, etc.) and compare before vs. after. The path to get there is either not modeled or is assumed to be resolved by some learning algorithm that eventually settles. In contrast, IDT emphasizes the **trajectory**. Even if a final consensus might be reached, the journey (how the drift happened, how fast, with what intermediate states) matters. There might be multiple possible endpoints depending on initial conditions and random fluctuations along the way (history matters). Game theory can handle multiple equilibria too, but it doesn't inherently account for *time-continuous* changes between them except via repeated game frameworks or learning algorithms which are often still stepwise and rational (like best-response dynamics, fictitious play, etc.). Those learning models usually require a feedback of utility each round. IDT doesn't necessarily involve utility; the feedback is seeing others' states.
- **Interpersonal vs. External Incentives:** In many binary models, changes in behavior are driven by external incentive changes or explicit signals (like one player changes strategy to better respond to others' strategies). In IDT, change is driven by interpersonal influence *for its own sake* – e.g., identification, peer pressure, or normative alignment. This makes it particularly suited to things like ethics or culture where there may not be a clear immediate payoff for an individual to behave one way or another, but the social environment strongly guides it. For instance, game theory might struggle to explain why an employee would adopt an unethical practice that harms the company (and maybe eventually themselves if the company is caught) – rationally they shouldn't if it's against interest. But if everyone around them is doing it and a leader expects it, social influence can make it “rational” in a normative sense for the person to conform. IDT can capture that slide without invoking a change in formal incentives (the incentive could even be negative, but they still drift into it due to influence).

One way to illustrate the difference: **binary game model of organizational change** might say “if enough people defect (break rules), the best response for an individual might flip from cooperate to defect, leading to a new equilibrium where all defect.” This is like a sudden phase transition (tipping point) and often involves discontinuous change (a critical mass is reached and then everyone changes strategy). **IDT's gradual model** would see each additional defector slightly reducing others' perceived norm of honesty, nudging them slightly toward defecting, and

as more do so, it accelerates – but each individual’s transition is smooth, and one could measure the degree of norm erosion at each time. There might still be a non-linearity (once 50% are cheating, maybe the rest quickly follow), but it’s not a single jump; it’s spread over time steps with intermediate states.

In summary, IDT complements game-theoretic thinking by providing a *fine-grained, process-oriented* view. They are not mutually exclusive: one could incorporate influence drift into game simulations (e.g., players’ payoffs or preferences change slowly as a function of the social environment). But IDT by itself reframes organizational dynamics as being less about calculated moves and countermoves, and more about the slow molding of minds and norms. This resonates with many organizational phenomena that game theory has difficulty with, such as culture and values. It also allows analysis of scenarios where binary models would label agents irrational – under IDT, they might be perfectly normal people simply adapting to their social context.

Illustrative Scenarios

To make the concepts concrete, we explore several scenarios that highlight how Influence Drift Theory can be applied and what insights it can yield. These scenarios – **Authoritarian Decay**, **Democratic Resilience**, and **Ethical Propagation** – correspond to patterns observed in real organizations and societies. In each case, we describe how the IDT framework operates and contrast its predictions with what classical views might suggest.

Scenario 1: Authoritarian Decay

Description: An organization (or government) is structured in a strongly hierarchical, authoritarian manner. There is a single leader or a small top cadre who hold most of the decision-making power. Influence flows predominantly from the top downward; feedback from below is minimal (subordinates have little influence on leaders’ behavior). Initially, assume the organization is performing adequately – perhaps it was founded by a visionary who set clear ethical standards and goals. Over time, however, that original leader is replaced or changes – the leadership that emerges is less competent and less ethical, yet retains the same absolute authority. How does the organization’s quality and integrity evolve under IDT?

IDT Analysis: In this scenario, the influence network is like a **tall tree** with the leader at the root influencing all others through branches. The leader’s direct reports take heavy influence from

him; they in turn heavily influence their reports, and so on. Let's say a leader's ethical metrics begin to decline (perhaps due to external temptations or changes in personal values). At first, your subordinates may not follow right away and will still adhere to existing ethical standards. However, as the leader issues directives or sets examples that deviate from prior norms (e.g., pressuring to meet targets by “whatever means necessary”), By the second tier, those managers now, perhaps reluctantly, enforce or model slightly more unethical practices to their own subordinates, who then also drift downward. Because **feedback is absent or ignored**, there is no corrective mechanism; even if some in the middle or bottom feel uneasy (their personal values differing), their influence upward is negligible. They might voice concerns but be silenced or fear reprisal (as happened in companies like VW, where an engineer noted the illegal software and was told to keep quiet. Over iterations, the entire organization's ethics metric moves from say a high value towards the leader's lower value.

A similar story can be told for **competence**. If an authoritarian leader starts making capricious decisions or doesn't listen to experts (perhaps due to hubris or lack of knowledge), the organization's competence can degrade. Talented employees whose input is ignored may become disengaged or leave; the ones who stay conform by doing only what is asked (stop taking initiative). The influence dynamic here is that the leader's behaviors (like discouraging dissent, valuing loyalty over expertise) trickle down. Mid-level managers mimic this by not forwarding bad news or by promoting people based on loyalty instead of merit (because that's what they perceive the top wants). Over time, **organizational drift** sets in: metrics like innovation, problem-solving, and adaptability all worsen because the social network has oriented itself to pleasing the top rather than confronting reality.

The end state of an authoritarian drift can be quite dire: a once-effective organization might become brittle, corrupt, and out-of-touch. **Authoritarian decay** is often observed in long-standing dictatorships or in corporations with long-tenured CEOs who become surrounded by “yes-men.” Sociologist Robert Michels coined the “iron law of oligarchy” – that even democratic organizations tend to turn oligarchic, with leaders insulating themselves. IDT provides the mechanism: once influence concentrates and feedback closes off, drift (especially negative drift) can proceed unchecked. In the absence of external correction, this ends in either **collapse or crisis**. We can cite historical examples: many observers of the **Soviet Union** noted how stagnation and a culture of sycophancy under decades of authoritarian rule led to systemic incompetence and eventual collapse. In business, **Enron** in the 1990s can be seen as an authoritarian culture at the top (Skilling and Lay's aggressive profit-driven ethos) that caused

ethical drift down the hierarchy, ending in a notorious collapse – analysts point out that Enron’s failure was not one or two bad apples but a culture of fraud that propagated from the top throughout.

Classical theory might attribute authoritarian organization issues to “centralized decision-making” or “lack of flexibility,” which is true, but IDT pinpoints *why* those issues worsen over time: because without feedback, the central decisions become more and more misaligned as everyone adjusts to a perhaps flawed central viewpoint. A game-theoretic approach might look at the leader-follower as a principal-agent problem; IDT instead looks at it as a contagion problem. The difference is subtle but important: in principal-agent, if the agent cheats, the principal tries to sanction – in authoritarian drift, if the principal cheats, the agents *learn* to cheat as well (since they take cues from the principal). There’s no corrective principal above the top leader in an authoritarian system.

Interventions: What could stop authoritarian decay? IDT suggests two things: either an *external shock* (e.g., regulators intervene, an outside authority imposes change – effectively injecting a new influence like a legal mandate that forces behavior regardless of internal beliefs) or an *internal reform* that introduces feedback loops. For instance, if at some time a new influence is injected – say a whistleblower arises (someone with high internal ethical steadfastness and possibly given a platform by external attention) – that can act as a competing influence on some individuals, possibly slowing the drift or even reversing it if it topples the leadership. But absent such changes, the drift is likely one-way. This scenario underscores IDT’s cautionary implication: **unbalanced influence structures can lead to slow, almost invisible degradation** of an organization’s core qualities.

Scenario 2: Democratic Resilience

Description: In contrast, consider an organization (or community) with a **democratic or participative structure**. No single node completely dominates the influence network; instead, influence is more evenly distributed and often reciprocal. Such an organization might have multiple leaders or centers of influence (e.g., an executive team rather than a single CEO, or a coalition government rather than an autocrat). Feedback from lower levels is institutionalized – there are forums for employees to share input, decisions are made by consensus or voting, and transparency is high. The question is: how does such a network respond to disturbances or attempts at extreme change? Is it more **resilient**?

IDT Analysis: A democratized influence network can be represented by a **decentralized graph** – perhaps even fully connected or with multiple loops. For example, consider a cooperative where every member has a say (equal influence weights to each other, roughly). Or a corporation with a very open culture, where even junior staff can question senior staff's decisions (i.e., significant upward influence). In this scenario, suppose one member (even if it's the nominal leader) introduces a radical idea or begins to drift in a harmful direction (say, proposing an unethical shortcut or a strategy the others see as risky). Unlike in the authoritarian case, here the network will not simply follow by default – because other influencers exist who can push back. If the CEO of a participative company suddenly becomes erratic, the board of directors (or even employees through internal feedback mechanisms) might check that influence. In IDT terms, if one node deviates, that information travels to others, but those others do *not* all blindly update towards the deviation; instead, they might moderate their response because they also are influenced by each other and by norms that have formed collectively.

Democratic structures often create a **balance of influences**. Think of it like a multi-agent averaging model where many nodes have roughly equal weight – the effect of one outlier is diluted by the majority's influence. So if one person's ethics drop in a 5-person egalitarian team, the other 4 (assuming they remain high) exert a larger combined pull to keep that person in check than the one person exerts to pull the 4 down. The outcome might be that the one deviating person either comes back to the norm (peer influence corrects them) or, if they resist and remain deviant, they eventually become isolated or even removed (the team might expel or sideline a persistently unethical member, analogous to a democracy voting out an official who doesn't adhere to norms). Thus, **resilience** here means the group can absorb shocks or bad influence from a few members without completely drifting off course.

Another aspect is information processing: In a democratic network, errors can be spotted by someone and **communicated widely**, preventing the whole group from going down a wrong path. For instance, in an open culture company, an engineer can openly say “this design is flawed, we need to fix it,” influencing managers to reconsider, thereby averting a potential failure. In NASA's case, the lack of heed to engineers' warnings was a factor in the Challenger disaster – one could argue NASA at that time had some authoritarian features in decision-making. A more resilient structure might have empowered the engineering concerns to influence the launch decision more strongly.

Let's also consider **democratic political systems** through IDT. In a democracy, leaders are influenced by public opinion (through polls, media, elections – which are themselves a reflection of influence networks among citizens). If a leader tries to amass power or implement extreme policies, there are checks and balances (courts, opposition parties, free press) that represent other influence nodes pushing back. The net effect tends to be moderation: policies drift toward what is acceptable to a broad base, and if something too extreme passes, there is often an electoral correction later. Democracies are not immune to drift (they can polarize if influence networks segregate, as we see with echo chambers in social media), but the presence of multiple centers of influence gives a fighting chance for self-correction. For example, consider an economic crisis – in a democracy, policy might change course after debate and input from various experts and stakeholders, whereas in an authoritarian regime, the top might double down on a flawed policy, with others afraid to contradict, leading to disaster.

In IDT terms, **robust networks** tend to have redundant and diverse influence paths. They don't depend on one node, so they don't have a single point of failure. They also allow negative feedback (in the control theory sense) – opposing influences that dampen deviations. This often prevents runaway drift of one kind (though if the entire group inclines together, they can still all drift – but that would typically be due to some external alignment or a consensus shift for perhaps rational reasons).

From a classical perspective, one might say democratic organizations have *lower efficiency* because reaching consensus takes time and there's a lot of discussion. IDT can incorporate that: a very dense mutual influence network might converge more slowly than a star network commanded by one leader (just as autocracies can make decisions faster). However, the **quality** of the outcome at equilibrium may be higher or more stable in the democratic case. Efficiency in the short run (authoritarian quick decision) can mean inefficiency in the long run if that decision is wrong and needs reversal – whereas participative approaches might avoid the wrong decision in the first place through deliberation. IDT simulations could demonstrate scenarios where a strong leader gets the group to consensus faster (less time steps) but to a suboptimal consensus (like everyone's ethical level drops more), whereas a fully connected egalitarian group takes more iterations to settle but ends up at a higher ethical level.

Real-world example: The **open-source software community** is often quite decentralized and democratic in influence – code decisions emerge from discussions, not commands. Such communities can be remarkably resilient: even if a founder leaves or a prominent maintainer

goes rogue, others step in and the project can continue (forking the project if needed). Influence is based on merit and trust built over time, and no single person's absence or misconduct necessarily ruins the project (contrasted with a startup company where a domineering founder's fall might collapse the company culture). Another example, **Toyota's production system** is known for empowering any worker to stop the assembly line if they spot a problem (the Andon cord). This is a built-in feedback mechanism preventing drift into poor quality – problems are addressed when small. Companies that lack this (perhaps due to strict top-down control) may let small quality issues drift until they cause a recall crisis.

In summary, **Democratic Resilience** in IDT is characterized by a network that resists drifting due to any single influence. It tends to require a plurality of the network moving for the whole to significantly shift, which usually means more deliberation and cross-influence – a stabilizing factor. The scenario teaches that encouraging multiple voices and mutual influence in an organization can act as a safeguard against extreme outcomes (like ethical collapse or strategic blunders), albeit sometimes at the cost of speed or unity in the short term.

Scenario 3: Ethical Propagation

Description: This scenario examines how positive change – specifically, raising the ethical standards of an organization – can be achieved through influence. Suppose an organization has historically had a mediocre ethical climate. There isn't outright fraud, but people commonly ignore minor rules, and corners are cut. Now a new initiative is undertaken to improve ethics: perhaps a new **ethical leader** is hired (e.g., a new CEO with a reputation for integrity), or a comprehensive ethics training and mentoring program is rolled out. How can IDT model the propagation of ethical behavior, and what factors determine success or failure of such an intervention?

IDT Analysis: IDT Analysis: At the beginning, the distribution of ethics indicators across the organization is constant. Let's say on a 0–100 scale, most people are around 45~55 (somewhat ethical but will compromise occasionally). The goal is to move the organization towards 90 (very ethical, low tolerance for misconduct). A purely rational approach might implement new rules and assume people will follow them. But IDT suggests that unless those rules permeate into daily behaviors via influence, they may remain "on paper." So how to propagate ethics?

One powerful method is **leadership by example**. Introduce a highly ethical leader at the top (initially, that leader is a node with ethics = 0.95, much higher than the average). If this leader

has significant influence weight on direct reports (which, in a formal hierarchy, they usually do), then over time those direct reports will shift upward in their ethics. This might happen because the leader actively communicates the importance of ethics, makes decisions that sacrifice profit for principle (signaling new norms), and perhaps rewards ethical behavior (and penalizes unethical behavior). The direct reports not only hear the message but see consistent action. IDT would update their ethics: Each manager with a high score increases each member of the organization closer to the leader level. Empirical studies support this mechanism: ethical leaders *“promote such conduct to followers through two-way communication, reinforcement, and decision making”*, and this *encourages employees to be ethical*. In IDT, that two-way communication and reinforcement is precisely creating influence links and feedback: employees see ethical decisions being praised, unethical ones being corrected, so the social reward structure shifts.

The propagation continues downward. As mid-level managers become more stringent ethically, their subordinates will notice the difference. Perhaps previously it was okay to fudge a report; now their manager insists on accuracy and transparency. There might be some resistance at first if unethical habits were ingrained, but gradually, if the pressure is consistent and if people start to believe “this is the new normal,” their own internal threshold moves. IDT can incorporate a slight time lag or resistance. But with repetition, they drift upward. It’s critical that the network not have *inconsistent signals*. If top leadership is ethical but a local supervisor remains unethical (maybe they pay lip service but still encourage cheating privately), then subordinates get conflicting influences. The outcome could be mixed or a stalemate, where some drift up, some not. Ensuring propagation means aligning influence at multiple levels – basically creating a **cascade** of ethical influence. This is why ethics programs often involve training at all levels, so that a critical mass of influencers is reached.

Another pathway is **peer influence and ethical champions**. Instead of relying purely on formal hierarchy, an organization can seed ethical influencers across the network. For example, appoint “ethics ambassadors” in each department – respected employees who champion integrity. These individuals, if respected (i.e. they have some social influence weight among peers), can spread new norms horizontally. A peer who refuses to go along with a minor violation and explains why might cause others to reconsider. Over time, their colleagues may drift to adopt the same stance, especially if they see that person respected or rewarded. This is analogous to how grassroots social movements work: a few passionate advocates influence their friends, who then influence their friends, and so on.

We can simulate a case: imagine a company of 100 people, initially average ethics 45~55. We place 5 ethical champions at random points in the network with ethics 90 and give them somewhat higher influence (they are outgoing, respected folks). We simulate interactions (maybe random pairwise meetings or group discussions each round). We'd likely observe clusters around those champions start to rise in ethics score. If the network is well-connected, eventually those clusters overlap and the overall average climbs. If the network is fragmented, some pockets might remain low ethics because they never interact with a high-ethics influencer (hence the importance of coverage – ensuring every part of the organization gets exposure to the new standard).

Pitfalls: IDT can also highlight why ethical propagation sometimes fails. If the new ethical leader doesn't actually have much influence (say they are distant or employees doubt their sincerity), waiting to see if this is just rhetoric. Or if there are a few very stubborn nodes who continue to exert contrary influence (imagine a star salesperson who keeps bending rules and others still idolize him for his sales, thus giving him high influence weight), they can act as anchors pulling the metric down. The organization might then split: the formal message is ethical, but a sub-culture remains that drifts slower or not at all. Overcoming that may require either converting or removing those counter-influencers (which aligns with practice: sometimes a company must fire or isolate those who refuse to get on board with ethics, despite their other performance contributions).

Another factor is **reinforcement**: single influences might fade unless reinforced frequently. If ethical training is a one-off workshop, its effect might decay. People might drift a bit up, then slide back if old influences remain. But if training is ongoing and coupled with structural changes (like performance evaluations now include ethics, etc.), then multiple influence vectors (formal evaluation, peer pressure, leadership messaging) all push in the same direction consistently, leading to a more permanent drift.

Applied example: Consider a government attempting to reduce corruption in its civil service (ethical propagation in a community/government context). It could create anti-corruption commissions (influencers), raise salaries (to reduce external incentive for corruption, though IDT is more about internal influence so we focus on normative campaigns). Often, they also implement peer monitoring and ethical pledges, try to celebrate "honest officer of the month" to set examples. These strategies correspond to injecting positive influencers and changing the narrative that "everyone is taking bribes" to "integrity is valued and common." There have been

cases like Georgia (in the 2000s) where aggressive anti-corruption reforms, including new leadership and mass replacement of traffic police, dramatically changed public sector ethics within a few years – essentially by resetting influence: new hires with no norm of bribe-taking influenced each other to maintain a clean culture. In a smaller scale, a company like **Wells Fargo** after its fake accounts scandal had to change its sales culture – they removed certain top executives and managers who pushed unethical practices and tried to institute a customer-focused ethos. The success depends on whether those changes truly penetrate daily behavior via influence or remain superficial.

Metrics interplay: In ethical propagation, **trust** often increases as a side effect. When employees see leadership walking the talk, their trust in leadership rises, which then strengthens the influence of leadership in other areas (a virtuous cycle). Competence might also rise if unethical shortcuts were hurting quality – now doing things right might improve product outcomes (though it might reduce short-term output if people were cutting corners to boost numbers, it likely stabilizes long-term performance).

Through IDT simulation of ethical propagation, one could experiment: what if only the CEO is changed vs. what if we also change a few key managers? How high must the leader's ethics be, and how large must their influence weight be, to tip the whole company's culture above, say, 0.8? If the influence network has certain central nodes, targeting those (either converting them or replacing them) yields the biggest shift. This is akin to identifying "culture carriers" or informal leaders in an organization.

Summary of Scenarios

Across these scenarios, we see the power of *incremental influence accumulation*. In **authoritarian decay**, the lack of checks allows a slow but relentless deterioration as each person's small compromises aggregate into a big problem. In **democratic resilience**, many small corrective influences prevent large deviations, maintaining stability. In **ethical propagation**, deliberate positive influences gradually elevate the whole by capitalizing on interpersonal spread. None of these changes happen overnight or from one command – they are the result of iterative social processes, which is exactly what IDT focuses on.

Applications of Influence Drift Theory

Influence Drift Theory is not only an academic construct; it has practical relevance for various domains. We highlight applications in **corporate management**, **public governance**, and **community development**, demonstrating how IDT can inform both analysis and action.

Corporate Management and Organizational Change

Modern corporations are keenly interested in managing organizational culture, improving teamwork, and ensuring ethical compliance. IDT provides a framework to diagnose cultural issues and plan change initiatives:

- **Culture Diagnostics:** By mapping the influence network within a company (formally via org charts and informally via social network analysis tools or employee surveys), one can identify who the key influencers are and how certain values or behaviors might be drifting. For example, if a company's customer service quality is declining, IDT would prompt managers to look at not just process metrics but *who influences service reps*. Perhaps a regional manager with cynical attitudes has been subtly encouraging lax service, and others emulate that. Standard KPI tracking might not catch the cause, but an IDT lens would. In practice, methods like "organizational network analysis (ONA)" are used to find influencers, which aligns with identifying high-weight nodes in IDT terms.
- **Change Management:** Traditional change management often uses a "**cascade**" **approach**, recognizing that leaders need to role-model and then managers need to adopt, etc. IDT formalizes why cascade is needed: each layer influences the next. If a change (like adopting a new safety protocol) isn't embraced at one level, the level below won't sustainably adopt it either due to conflicting influence. Thus, when planning a change, management can ensure to secure influence at multiple levels (e.g., get buy-in from informal leaders and supervisors, not just issue a top-down directive). They can also decide whether to approach change top-down or via pilot groups (small teams first adopt the new practice, then influence others). IDT could simulate both to see which yields faster convergence or less resistance. A pilot approach corresponds to creating a pocket of new influence that then spreads (if connected to others); a top-down approach is broad but might be shallow if not reinforced locally.
- **Leadership Development and Succession:** Companies invest in developing leaders who can positively influence others. IDT underlines the impact of leadership beyond

direct decisions – leaders set tones that drift. Hence, selecting a CEO or manager isn't just about their individual skills but about what kind of *network effect* they will have. A leader who isolates themselves has low effective influence weight despite authority (because few interactions); a leader who actively engages employees increases their influence links. Some firms now use internal social media or town halls to allow leaders to connect broadly, effectively boosting the leader's influence reach in the network. In succession planning, if a highly influential charismatic leader is retiring, IDT would warn of a possible culture shock – the network might lose a binding force and drift unpredictably. Planning could involve identifying who will fill those influence roles or distributing influence more widely before the transition.

- **Ethics and Compliance Programs:** Many corporations have compliance training, codes of conduct, etc., to prevent misconduct. IDT suggests that beyond formal rules, **monitoring the informal influence** is key. A single unethical manager can undermine a program by encouraging colleagues to ignore the training. Thus, companies might use IDT by, for example, measuring local ethical climates through surveys (to see which departments are drifting lower) and then intervening there, perhaps by replacing leadership or adding strong ethical influencers (like rotating in a respected manager known for integrity). Whistleblower systems also tie in: they amplify upward influence by giving voice to lower-level employees' observations, which can check drift as we described.
- **Mergers and Acquisitions:** Merging two organizations is essentially merging two influence networks with distinct cultures. IDT can be used to anticipate friction: if the acquirer imposes its culture via its leaders, how will the acquired employees adapt? There might be key influencers in the acquired company who resist and pull their subordinates in a different direction, causing integration failure. Identifying those nodes and either co-opting them (turning them into allies for the new combined culture) or neutralizing their influence is crucial. Often, not doing so results in a drift back to old habits or an us-vs-them culture split. Simulation might reveal that without a high density of cross-company ties, the two sub-networks won't integrate, staying in separate consensus states.

- **Team Composition:** When forming teams, managers can use IDT thinking to balance influences. For instance, a new project team might intentionally include a mix of people from different silos to ensure diverse influence (avoid groupthink drift in one direction). Or if a team is known to be negative, adding a particularly positive person can over time lift the mood (if that person is given enough social capital). This aligns with the practice of putting change “champions” in teams.

Public Governance and Policy

Governments and public institutions can also benefit from IDT insights, especially in managing social influence on a larger scale:

- **Policy Acceptance:** Consider how citizens adopt (or resist) new policies or public health guidelines. Influence drift happens in communities through social networks and media. Governments can use IDT principles by identifying *community influencers* – respected local figures, social media personalities, etc. – and involving them in advocacy. For example, during a public health campaign (like encouraging vaccination or mask-wearing during a pandemic), using doctors or local leaders to influence their communities will likely produce a gradual norm shift, rather than relying purely on official mandates. People may ignore a government PSA, but if many of their friends have been convinced by a local influencer to vaccinate, they’ll come around. This was seen in various health campaigns where peer educators were effective.
- **Democratic Processes:** As discussed in the scenario, democracies are constantly in flux as public opinion moves. IDT could model how information or misinformation spreads in the electorate. Understanding this can help in designing interventions to improve public discourse. For instance, if polarization is rising (two groups drifting apart in opinions because they are influenced only internally), interventions might involve creating more cross-cutting influences: town halls, debates, or media that expose each side to the other’s views in a humanized way. This is essentially modifying the influence network connectivity to prevent echo chambers. Governments or civil society could facilitate bridging ties.

- **Anti-Corruption and Institutional Reform:** We touched on how governments fight internal corruption. IDT might advise that changing laws is not enough; you must alter everyday interactions. For example, in an overly bureaucratic system where citizens routinely pay bribes to low-level officials (and everyone thinks that's normal), just penalizing bribery might not work until the social expectation changes. Some countries have used approaches like rotating staff frequently (to break cozy influence cycles between specific officials and communities) or introducing transparency (so one official knows others might see their work, creating a peer influence of accountability). Additionally, publicizing positive deviance – highlighting offices or regions that perform with integrity – can create a competitive influence (others see it is possible and desirable to be clean).
- **Education and Community Programs:** Governments often run programs to shape societal values (e.g., promoting tolerance, diversity, environmental responsibility). These can be framed as influence problems: how to get these values to diffuse among the population. School programs train young people who then influence their families at home (often kids influence parents' recycling habits, for instance). Community workshops create local champions for, say, energy conservation, who then persuade neighbors. Rather than only top-down regulations, fostering bottom-up influence can be more effective and self-sustaining.
- **Crisis Management:** In crises (natural disasters, pandemics, etc.), clear and consistent communication is known to be vital. IDT adds that using *trusted networks* is key: people are more likely to follow guidelines if they see their peers doing so and if instructions come through sources they trust (their doctor, community leader, etc.). Managing misinformation also becomes a matter of influence – quickly injecting correct information through the network before false rumors entrench is analogous to using “counter-influencers” to curb a bad drift.

Communities and Social Networks

Communities, including both physical communities (neighborhoods, cities) and virtual communities (online forums, social networks), are fertile ground for applying IDT:

- **Social Norms and Behavior Change:** Many community interventions rely on changing norms – whether it's reducing littering, increasing organ donation rates, or discouraging drunk driving. Research in social psychology shows that *perceived norms* strongly influence individual behavior. For instance, if people think “everyone around here litters,” they are more likely to litter. If they think “most people are actually disposing trash properly,” they conform to that. Thus, community leaders try to broadcast positive norms (e.g., an anti-litter campaign might emphasize that the majority do use bins, so the litterers feel in minority). This is using influence: get enough visible compliance so that others follow. IDT would approach a community problem by mapping how neighbors influence each other (often through observation and conversation) and then tipping that. Even subtle cues like seeing clean streets (outcome of behavior) can influence because it implies a norm.
- **Grassroots Movements:** Social movements (civil rights, #MeToo, etc.) often spread through peer-to-peer influence. One person's stance emboldens their friends, and so on. These are classic examples of influence drift leading to big societal shifts – usually gradual for a period, then reaching a tipping point. Movement organizers can use IDT logic: target influential community figures to join early, use testimonials (one person sharing experience influences another), create networks of activists in many localities to propagate the message. The difference from top-down is clear: it's organic influence. The success of a movement like #MeToo was largely via decentralized sharing of stories (each story influencing others to speak or change perspective), rather than a central authority ordering it. IDT provides a way to understand the momentum building and maybe simulate at what point a norm truly changes (e.g., when does sexual harassment become widely condemned rather than silently tolerated – you could view that as reaching a consensus above a certain ethics threshold after enough interpersonal discussion and witness).
- **Online Communities and Information Spread:** In the digital age, online platforms form influence networks of unprecedented scale. Algorithms aside, people influence each other through posts, likes, retweets. Misinformation can drift and become accepted in certain circles. By analyzing these as networks, one can identify *influencer accounts* or communities that act as echo chambers. Interventions like injecting fact-checks or promoting inter-group dialogue can be seen as attempts to alter the influence graph

(maybe by connecting a strongly misinformed cluster with outside voices). The interplay of IDT with algorithmic amplification is complex, but at its core, if people see false information repeated by those they trust, they adopt it – a clear case of influence drift. Conversely, if key opinion leaders in an online community start pushing back on a false narrative, followers may gradually update their beliefs.

- **Community Resilience:** Similar to democratic resilience but at community level – how communities respond to external stress (economic downturn, disaster) can depend on their internal influence structure. A community with strong social ties and positive influencers (neighbors who coordinate help, spread accurate info, encourage cooperation) might fare better than one with fragmented ties or negative influencers (people spreading panic or selfish norms). Building social capital and trust *before* crises is akin to strengthening the positive influence network so that in times of need, beneficial behaviors (like mutual aid) spread quickly instead of selfish or chaotic behaviors.

Using IDT in Communities: Practitioners might perform network mapping (like figuring out who in a village is most trusted for advice) and then involve those people in programs. For instance, an NGO promoting female education in a conservative village might first win over a few respected elders or parents, who then influence others by example – a gentle drift of attitudes rather than confrontation.

Discussion

The simulation results of Influence Drift Theory (IDT) highlight the emergence of **internal polarization in a heterogeneous organization**. Rather than converging on a single consensus (as one might expect from simple averaging models of influence), agents clustered into distinct subgroups with different prevailing “influence” values. In other words, initial **organizational heterogeneity** – whether in beliefs, influence power, or network position – was amplified or preserved, leading to **persistent divisions**. This outcome deviates from classical consensus models like DeGroot’s (1974) iterative averaging, which would predict eventual agreement given a strongly connected influence network. The fact that our IDT agents did *not* all agree suggests that mechanisms such as bounded confidence or partial stubbornness were at play, preventing

full convergence. Indeed, the pattern mirrors the dynamics of the Friedkin–Johnsen model, where each individual never completely abandons their initial opinion. In our IDT context, this means subgroups of the organization maintained distinct “imagined realities” or interpretations of information, resulting in **stable polarized coalitions** instead of unity.

This phenomenon of **internal polarization** can be interpreted through the lens of classic organizational theories. **Max Weber’s bureaucracy theory** posits that an ideal organization operates on impersonal rules, formal hierarchies, and a strict chain of command to ensure uniformity and control. Bureaucracy’s advantage, Weber noted, was its technical efficiency and *unity* – it can coordinate many individuals under a singular authority and set of norms. The IDT simulation’s divergent outcome stands in stark contrast to this ideal of unity. Whereas Weberian bureaucracy seeks to **minimize variability** in interpretations and behaviors through formal structure, IDT illustrates how, absent or beyond those formal constraints, informal influence networks can drift in multiple directions. In a rigid bureaucracy, one might expect top-down directives to suppress the kind of **influence drift** our simulation displays. The polarization we observed suggests that if bureaucratic controls or a clear “chain of influence” are weak, an organization might splinter into factions – each faction perhaps following its own local norms or leaders. This underlines a tension between **formal structure and informal dynamics**: IDT shows that the latter can undermine unity if not aligned with the former. In practical terms, a bureaucracy would need to recognize these undercurrents; otherwise, the formal organization (the “*imagined reality*” of roles and rules) could be subverted by emergent subcultures that formal rules did not anticipate.

In contrast, the **Human Relations perspective** (Elton Mayo and colleagues) would find our results more expected and explainable. Mayo’s studies in the 1930s (e.g. the Hawthorne experiments) revealed that informal group relations, feelings of belonging, and communication greatly influence worker behavior – sometimes more so than formal incentives or rules. He stressed “*the power of natural groups, in which social aspects take precedence over functional organizational structures*”, the necessity of **two-way communication** between workers and management, and the role of strong yet empathetic leadership to foster common goals. The **formation of cohesive subgroups** in the IDT simulation echoes Mayo’s point about natural groups: people (or agents) naturally form cliques and shared understandings that might deviate from official directives. The **internal polarization** we saw can be interpreted as what happens when there is a breakdown in Mayo’s advised two-way communication and inclusive leadership. Lacking open channels, subgroups in the simulation apparently reinforced their own beliefs and

became isolated “silos” of opinion. Human Relations theory suggests that to counteract such drift, management should actively engage with the informal organization – for example, ensuring that **organizational narratives and goals are communicated effectively** and feedback is received, so that different departments or teams stay on the same page. Our findings reinforce this: without **reciprocal communication and good leadership**, different parts of an organization may develop their own micro-cultures and narratives, essentially **different imagined realities** about “how we do things” – leading to the polarized camps as in the simulation. From a Human Relations standpoint, IDT’s lesson is that **employee perceptions and social bonds** can decisively shape organizational outcomes, sometimes overpowering formal structure. Strong leadership is needed to bridge these divides (a point Mayo also emphasized), by creating a unifying vision and making each member feel heard and aligned with the larger group.

Taking a broader **systems theory** and **contingency theory** view, the IDT results underscore the importance of balancing differentiation with integration. Organizations can be seen as **open systems** (Katz & Kahn, 1978) with multiple interdependent parts that must exchange information (influences) to function as a whole. In any large system, some level of heterogeneity or specialization is natural – each sub-unit may have its own role or environment, leading to different perspectives. Contingency theory, as formulated by authors like Lawrence and Lorsch (1967), explicitly found that different subgroups in an organization (e.g. sales vs. R&D) will develop different orientations and sub-cultures to meet their unique environmental demands. This *differentiation* is beneficial for specialization, but it creates a need for *integration* mechanisms to align these diverse units towards the overall organizational goals. Our simulation’s organization appears to have high differentiation (distinct clusters of influence) but insufficient integration (no convergence to a common state), resulting in fragmentation.

Polarization in the IDT model can thus be viewed as differentiation running rampant without integration – the subsystems (agent clusters) evolved in different directions and lacked sufficient feedback loops or incentives to re-synchronize. Classic systems theory would predict that without feedback to correct deviations, parts of a system may drift into different equilibria, threatening the integrity of the whole. Indeed, **contingency theory** warns that there is “no one best way” to organize – the appropriate structure depends on context – and that successful firms both allow subunits to differ and have strong integrative devices to prevent chaos. In IDT terms, an organization might *need* a certain amount of internal diversity (heterogeneity can provide adaptability and creativity, aligning with Ashby’s law of requisite variety), but it must also

invest in integration (e.g. cross-department teams, shared values, rotational programs) to ensure these diverse influences don't become **internally adversarial**. The simulation serves as a cautionary tale: if the **system's alignment forces (integrative leadership, shared culture)** are not strong enough, internal polarization can become a stable state, potentially reducing overall organizational performance and agility. From a systems perspective, one might ask: does the polarized organization reach a form of **stable equilibrium**? Possibly, but it could be a maladaptive one (akin to a metastable state) if the environment requires unified responses. Systems theory would encourage introducing feedback or incentives in the model to see if the polarized state can be destabilized back into coherence – essentially, testing how an organization can **recover unity after drift**.

Comparing IDT with **game theory models** provides another illuminating contrast. Classical game theory (as pioneered by von Neumann & Morgenstern, 1944) treats organizational interactions as strategic games between rational actors, each seeking to maximize their payoff under certain rules. In such models, outcomes settle into equilibria (e.g. Nash equilibria) where no individual can unilaterally do better by deviating. One could analogously frame influence dynamics as a game – perhaps a coordination game where each agent “chooses” an influence position and is rewarded for aligning with others. In a simple coordination game, there might be multiple **equilibria** (multiple stable consensus points), corresponding to different prevailing ideologies or norms. **Internal polarization** in our simulation could then be interpreted as the organization ending up in *two or more different equilibria at once* – in effect, the organization's sub-parts coordinated on different solutions. Game theory does allow for **multiple equilibria** (for instance, think of two competing standards in an industry; each group of firms coordinates internally on one standard). However, classical game models typically would still assume that within one play of the “game” the group picks one equilibrium. What IDT highlights is a more dynamic, **path-dependent process**: rather than a single rational choice, the final state depends on the history of interactions and influence *drift*. There is no single overarching utility being maximized; instead, local interactions drive the system toward a split outcome that no individual *designed*. Notably, **coalition formation** is a concept in cooperative game theory – von Neumann and Morgenstern even considered that players could form coalitions (e.g. legislative blocs or business partnerships) to improve their outcomes. In our simulation, the polarized clusters resemble such coalitions, but they arose organically from influence dynamics rather than explicit negotiation. In game-theoretic terms, one might say each polarized faction is internally stable (agents agree with their coalition) and the organization as a whole is stuck in a

suboptimal equilibrium where cross-faction collaboration breaks down. Traditional game theory might not immediately predict this without introducing additional assumptions (like communication failures or preference shifts), because in a fully rational analysis all agents would foresee the benefits of a single organizational consensus (a higher collective payoff perhaps) and coordinate on that. **Behavioral and evolutionary game theory** offer closer parallels: for example, Axelrod's (1984) simulations of the evolution of cooperation show how repeated interactions and memory can lead to stable cooperative or defective behaviors in groups. Similarly, **Schelling's model of segregation (1971)** demonstrated that even mild individual preferences can lead to strongly polarized outcomes in aggregate – a result very much in spirit with our IDT findings. The difference is that IDT focuses on **drifting influence (beliefs, norms)** rather than explicit strategic moves about location or cooperation. It complements game-theoretic insights by showing how even without clear payoffs and rational strategies, social systems can get **locked into polarized configurations** through micro-level interactions. In summary, game theory provides a useful metaphor (subgroups as different Nash equilibria in an coordination problem), but IDT emphasizes the *process* over the *endpoint*: it shows *how* an organization might meander into one of many possible equilibria due to internal influence forces, rather than assuming equilibrium from the start. This perspective is valuable for organizational analysis because many real-world organizations do exhibit path-dependent cultures and divisions that are not easily explained by one-shot rational choice models.

Finally, it is worth reflecting on what these findings mean for the very nature of an organization. Yuval Noah **Harari's concept of "imagined realities"** is particularly apt here. Harari (2014) argues that large-scale human cooperation is made possible by shared myths or collective fictions – things like nations, money, or corporations that exist because we **collectively believe in them**. An organization's culture, mission, and identity can be seen as such an imagined reality: it's a story that all members tell themselves so they can work together towards common goals. **Our IDT simulation can be viewed as probing the stability of this shared fiction.** When the simulation began, presumably the agents had some common overarching framework (they are in the same organization), but with heterogeneity in how strongly they believed or what interpretation they had. Over time, as influences drifted, it's as if different groups **started believing different stories** about the organization. In Harari's terms, the **imagined order fragmented** – instead of one collective myth ("we are all on the same team with the same goal"), multiple conflicting sub-myths emerged. Harari notes that imagined realities are powerful only so long as most people accept them. Our findings illustrate this vividly: when consensus on

the corporate narrative erodes, the organization's cohesion and power (its ability to get "millions of strangers to cooperate...towards common goals") diminishes. Internal polarization can thus be seen as a **crisis of narrative** within the organization – a sign that the unifying imagined reality is no longer universally believed. This connects back to the classical theories: for Weber, the "imagined reality" holding a bureaucracy together is the rational-legal authority and rules; for Mayo, it's the sense of social belonging and satisfaction; for contingency theorists, it's an adaptive culture tuned to the environment; for game theory, it could be a common knowledge of the game everyone plays. IDT shows how that underpinning narrative can **drift** or split, threatening to dissolve the glue that makes collective coordination possible. The practical implication is that leaders and managers must be **custodians of the shared reality**: they need to monitor the informal currents of influence and meaning in their organizations and take action when factions begin to form their own separate realities. Interventions could include re-articulating a compelling shared vision, reshaping networks of interaction (so siloed groups mix more), or introducing "**integrative**" narratives and rituals that re-bind the different tribes within the company. By doing so, they guard against influence drift that could otherwise entrench divisions. In essence, IDT not only aligns with Harari's insight that companies and institutions function as long as people collectively buy into the fiction, but also provides a dynamic view of how those fictions can erode from within. It bridges the gap between abstract social theory and concrete organizational behavior: even something as conceptual as an imagined reality has very real dynamics and can undergo a **drift** if not actively maintained. The discussion above, linking IDT with both classical and contemporary theories, demonstrates that while the language differs – bureaucracy's **unity**, human relations' **social groups**, systems' **integration**, game theory's **equilibria**, and Harari's **shared myths** – all are concerned with how organizations achieve or fail to achieve **cohesion**. IDT's contribution is providing a simulation-based micro-level lens on these macro theories: it shows the process by which cohesion can give way to fragmentation, and in doing so, it enriches our understanding of organizational behavior in a novel, synthetic way.

References

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345), 118–121.

Friedkin, N. E., & Johnsen, E. C. (1990). Social influence and opinions. *Journal of Mathematical Sociology*, 15(3–4), 193–206.

Harari, Y. N. (2015). *Sapiens: A brief history of humankind*. Harper. (Original work published 2011)

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). Wiley.

Lawrence, P. R., & Lorsch, J. W. (1967). *Organization and environment: Managing differentiation and integration*. Harvard Business School Press.

Mayo, E. (1933). *The human problems of an industrial civilization*. Macmillan.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Weber, M. (1947). *The theory of social and economic organization* (A. M. Henderson & T. Parsons, Trans.). Free Press. (Original work published 1922)