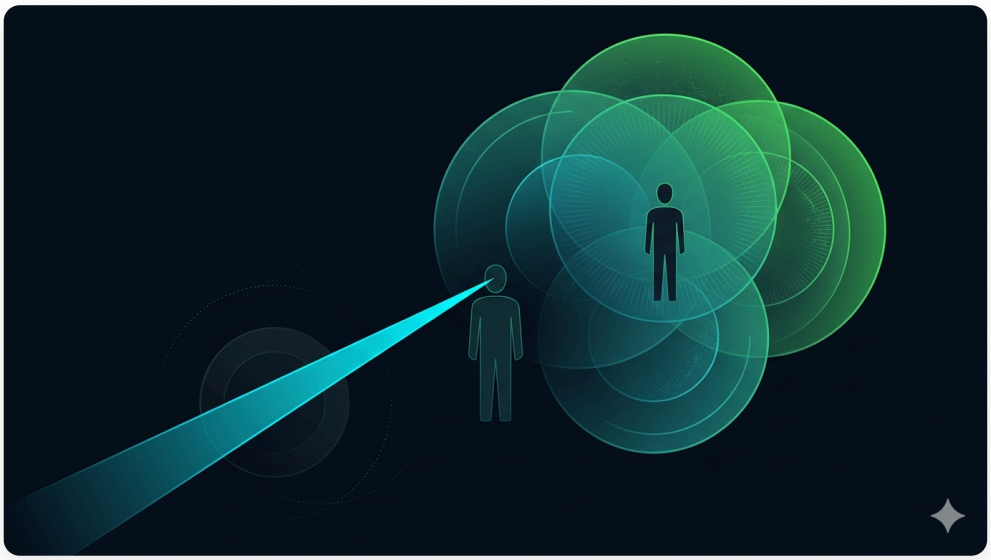


The Variety Gap in the Self

What a Person Optimizes For Determines What They Can Perceive

Companion to the Governance as Engineering series

Extends the variety-gap framework from governance institutions to the individual self. The ego is a low-dimensional controller; narrow personal values create blind spots that eventually return as crises. Introduces the self-variety gap (G_self) as a diagnostic for psychological blind spots and a guide to the practices that can close it.

Björn Kenneth Holmström

May 2026

`https://bjornkennethholmstrom.org/working-papers/self-variety-gap`

Abstract

Every person operates under an explicit or implicit set of personal objectives — values, goals, and metrics that define what counts as success and failure in their own life. That personal objective function selects which dimensions of their experience they attend to, which deviations they correct, and — critically — which signals never reach conscious awareness. This paper treats personal values as observation architectures. Drawing on the control-theoretic framework of the Governance as Engineering series, it shows that low-dimensional personal value functions produce the same structural collapse as low-dimensional governance channels: the excluded dimensions of the self eventually re-enter as unresolvable crises.

The paper introduces the *self-variety gap* (**G_self**) — the mismatch between the effective dimensionality of a person's true state and the dimensionality of their value architecture — as a unifying diagnostic. When **G_self** exceeds a critical threshold, the ego becomes constitutionally unable to perceive the sources of its own distress. Long-run personal wellbeing, the paper argues, requires not merely broader values but an enduring capacity for *personal meta-governance*: the ability to consciously evolve one's own value architecture as life unfolds.

The argument proceeds in three explicit layers: a rigorous core applying control theory and Ashby's Law to self-regulation; an interpretive layer drawing on evidence from psychology and neuroscience; and a bounded speculative layer exploring the implications for transcendence, nondual awareness, and the asymptotic limit of full self-observability.

Contributions

- Establishes **personal objective functions as observation architectures**, extending the governance-as-engineering framework to the individual.
- Formulates the **Goodhart–Ashby synthesis for the self**: any personal value function with dimensionality lower than the self's variety eventually degrades the ability to perceive the self's true state.
- Introduces the **self-variety gap (G_self)** as an operationalizable metric for psychological blind spots and crisis vulnerability.
- Demonstrates the framework through **evidence from self-complexity theory, cognitive dissonance research, and contemplative neuroscience**.
- Articulates the **personal meta-governance imperative**: the capacity to audit and evolve one's own value architecture before the gap becomes incapacitating.

Part I — The Unifying Principle: How Narrow Values Destroy Self-Observability

Before we can argue that *what a person optimizes for determines what they can perceive about themselves*, we must establish a more fundamental principle: that the structure of self-observation — how experience is attended to, interpreted, and acted upon — places hard limits on what any person can know about their own state, regardless of their intelligence, sincerity, or effort. That principle is the backbone of the engineering analysis of governance, and it translates directly to the architecture of the individual mind.

1.1 The Engineering Grammar Applied to the Self

This paper extends the framework developed in Papers I–VI of the Governance as Engineering series to the architecture of the individual self. It can be read as a companion volume, applying the same formal grammar at a different scale.

The Governance as Engineering series models institutional governance as a feedback control system. A controller — a government, a ministry, a regulatory body — receives information about the state of the system it governs, compares that information to a desired state, and issues corrective actions. The quality of governance depends on the fidelity of the observation channel, the latency of the feedback loop, and the dimensionality of the controller's model of the system (Ashby, 1956; Conant & Ashby, 1970).

The same architecture operates within a single person. Psychologists have long modeled self-regulation as a feedback process: individuals set goals, monitor their progress, and adjust their behavior to reduce discrepancies between their current state and their desired state (Carver & Scheier, 1982; Powers, 1973). The parallels are not metaphorical. They are structural:

- **State (x)**: The true multidimensional condition of the person — physical health, emotional state, relational quality, sense of purpose, cognitive capacity, and so on.
- **Objective function (V)**: The set of values, goals, and metrics that define what the person treats as success or failure. This is the personal value architecture.
- **Observation channel (O)**: The perceptual and cognitive processes that sample the person's true state and produce a conscious or semi-conscious representation of it. This includes attention, memory, interoception, narrative self-construction, and the filtering effects of cognitive biases.
- **Controller**: The ego — the decision-making processes that compare observed state to desired state and initiate action.
- **Actuators**: Behaviors, habits, communication, and the modulation of the environment.

In this schema, a person is a self-governing system. They observe themselves, evaluate what they observe against what they value, and act to close the gap. The engineering analysis of governance applies directly: the quality of self-governance depends on the structure of the observation channel, the latency of self-correction, and — crucially — the dimensionality of the personal value architecture.

1.2 The Common Mechanism

The central finding of the Governance as Engineering series is that *aggregation destroys information*. When a controller observes only a summary — a national average, an aggregated preference poll — it loses the spatial, temporal, and distributional detail needed to respond appropriately. The destroyed information cannot be recovered downstream. Institutional quality cannot act on a signal that is already gone.

The same mechanism operates in the self. When a person compresses their complex, multidimensional experience into a narrow set of evaluative dimensions — “Am I happy?”, “Am I successful?”, “Am I in control?” — they destroy information about the dimensions that fall outside those metrics. A person optimizing for happiness may fail to register the slow erosion of meaning, the decline of physical health, or the fraying of authentic connection, because none of those register as deviations from the happiness target until they have progressed far enough to overwhelm it. By the time the crisis is perceptible, the underlying conditions that produced it are already deep.

This is not a failure of effort or awareness. It is a structural consequence of using a low-dimensional value function to govern a high-dimensional system. The personal objective function is the observation channel for the self. It selects which aspects of lived experience are visible as successes or failures, and which are invisible as noise. Whatever is true about the person but absent from their value architecture might as well not exist — until it produces a disturbance that cannot be ignored.

Not all crisis is pathological. Some disturbances are the natural friction of a complex life — loss, illness, mortality — that no value architecture can eliminate. The framework concerns itself with the crisis that arises not from life’s inherent difficulty, but from the systematic exclusion of dimensions that could have been perceived, and often were, at the periphery of awareness.

1.3 The Self as a Case of Value-Architecture Failure

Consider three common personal value architectures, each of which approximates a one-dimensional optimization:

- **Optimizing for status**: The person attends to signals of social rank — income, title, recognition — and filters out signals from their body (exhaustion), their relationships (shallow, transactional), and their inner life (lack of purpose). The excluded dimensions accumulate. Burnout, loneliness, or a midlife crisis arrives as a “sudden” shock.
- **Optimizing for pleasure**: The person attends to signals of immediate gratification and filters out the slower signals of meaning, growth, and long-term health. The excluded dimensions return as addiction, ennui, or physical breakdown.
- **Optimizing for control**: The person attends to signals of predictability and order, filtering out the signals of adaptability, vulnerability, and trust. The excluded dimensions return as rigidity, isolation, or catastrophic loss when circumstances inevitably exceed control.

In each case, the personal value architecture is structurally analogous to a centralized governance system observing only an aggregate metric. It destroys the very information the person would need to detect the emerging crisis. The mechanism is the same whether the controller is a ministry or a mind.

The rest of this paper formalizes this parallel. Part II develops the structural identity between personal objective functions and observation architectures. Part III examines the specific dimensions — truth, meaning, connection, health — that narrow value architectures systematically exclude. Part IV reviews psychological and neuroscientific evidence consistent with the variety-gap framework. Part V explores the meta-governance of the self: the capacity to consciously evolve one's own value architecture. And Part VI, carefully bounded, considers the implications for transcendence, nondual awareness, and the asymptotic limit of full self-observability. Throughout, the argument distinguishes the rigorous core from interpretive extensions and speculative horizons, so that the foundations remain visible and the boundaries remain honest.

Part II — The Optimization Turn: Personal Values as Observation Architectures

Part I established that the self can be modeled as a feedback control system and that narrow personal values destroy information about the self in the same way that aggregated governance metrics destroy information about a society. This part formalizes that parallel: a person's values are not merely preferences; they are an observation architecture that determines which dimensions of their own life they can perceive.

2.1 The Structural Identity

Consider a person navigating daily life. At any moment, they inhabit a vast state space of physical sensations, emotional tones, relational dynamics, cognitive patterns, and existential intuitions. Most of this is not consciously processed. Attention selects a narrow subset, cognition interprets it through narrative frameworks, and the result is a *perceived self-state* — a compressed, filtered, and often distorted representation of what is actually happening.

This perceptual process is structurally identical to an observation channel in control theory. The true state **x** of the person is projected through an observation matrix **C** — the composite of attention, interpretation, memory, and cognitive bias — to produce a conscious representation $\mathbf{y = Cx + \varepsilon}$, where **ε** is noise. The person then evaluates this perceived state against their personal value architecture **V** and acts to reduce any discrepancy.

Now consider what determines the content of **C**. A person's values — what they care about, what they define as success, what they fear as failure — shape attention. The entrepreneur optimizing for venture success notices market signals, investor sentiment, and competitive threats; they may not notice the subtle signals of relationship strain, physical exhaustion, or existential drift. The artist optimizing for creative achievement notices aesthetic possibilities and technical challenges; they may not notice financial precarity or social isolation until these reach crisis levels. The caregiver optimizing for the wellbeing of others notices distress in loved ones; they may not notice the depletion of their own reserves.

In each case, the personal objective function *is* the filter. It selects which dimensions of experience are amplified into conscious relevance and which are attenuated into background noise. The person does not choose to ignore their failing health or their fraying relationships. Their value architecture simply renders those dimensions perceptually inert — they do not register as deviations from the target, so they do not trigger corrective action.

This is the foundational claim: *a person's objective function is an observation architecture for the self*. What you optimize for determines what you can perceive about your own life. The choice of personal values is, simultaneously, the choice of personal blind spots.

2.2 The Dimensionality of a Personal Value Function

If personal values are observation architectures, then we can characterize them by the same properties we use for any observation channel: dimensionality, aggregation structure, and temporal horizon.

The **dimensionality** of a personal value function is the number of independent life dimensions it tracks. A person whose entire sense of success depends on career advancement is operating with **dim(V_self) ≈ 1**: the vast multiplicity of their existence — health, intimacy, creativity, community, spirituality — is projected onto a single axis. A person who tracks multiple dimensions — say, professional growth, family connection, physical vitality, and creative expression — is operating with a higher **dim(V_self)**. They have a richer perceptual field for their own life.

The critical point is that dimensionality is not about the number of *goals* a person has, but the number of *independent evaluative dimensions* they can distinguish and trade off. A person may have many goals that all reduce to a single underlying metric — for instance, "be admired," "be wealthy," and "be influential" may all be expressions of a single status dimension. The effective dimensionality is the rank of the value space, not the count of items in a goal list.

The **aggregation structure** describes how the value function compresses experience into evaluation. A person who asks "Am I happy?" at the end of each day is aggregating the entire day’s experience — moments of joy, frustration, connection, boredom, meaning, emptiness — into a single scalar judgment. None of the variance within the day survives that aggregation. The information that would have flagged an emerging problem — say, a growing sense of meaninglessness despite pleasant moments — is destroyed in the averaging.

The **temporal horizon** of a personal value function determines which future states are visible. A steep discount rate — a heavy preference for immediate gratification — renders the distant future effectively unobservable. A person optimizing for how they feel right now cannot perceive the slow accumulation of health damage, the gradual erosion of a relationship, or the long-term consequences of a life path. The signal from those slow dimensions is too faint, too delayed, to compete with the immediacy of the present moment’s metric.

A single-metric personal value function with high aggregation and steep discounting is, in signal-processing terms, a narrowband, low-dimensional observation channel. It is structurally incapable of registering the full variety of the self it is meant to guide.

2.3 Requisite Variety for Personal Value Architectures

Ashby’s Law of Requisite Variety states that a controller can only stabilize a system if its internal variety matches or exceeds the variety of the disturbances it faces (Ashby, 1956). For a person, the "controller" is the set of values and goals that select which life disturbances are attended to and which are not. The "disturbance

space" is the full range of physical, emotional, relational, existential, and environmental variation that can push the person away from their desired state.

The extension follows directly: a personal value architecture must possess at least as much dimensionality as the disturbance space of the self, minus the dimensionality of acceptable outcomes. **dim(V_self) ≥ dim(D_self) − dim(G_self)**. If this condition is violated, there exist disturbance dimensions that the person's value system cannot register. Those disturbances are, in the formal sense, unobservable to the person's conscious self-regulation. They accumulate as chronic stress, unexplained anxiety, bodily symptoms, relationship deterioration, or existential unease — until they breach thresholds that force themselves into visibility through crisis.

This is not a normative claim about what a person *should* value. It is a structural prediction about what happens when a person's value architecture is too narrow for the complexity of their own life. The person who values only career success *will* be blindsided by health collapse, relationship breakdown, or spiritual emptiness — not because they are foolish or shallow, but because their value architecture lacked the dimensions needed to detect those disturbances in time.

2.4 The Goodhart–Ashby Synthesis for the Self

Goodhart's Law — "when a measure becomes a target, it ceases to be a good measure" (Goodhart, 1975) — was formulated about economic indicators, but its logic extends directly to personal metrics. When a person makes "happiness" their explicit target, they begin to optimize for the feeling of happiness rather than the conditions that naturally produce it. They avoid difficult conversations that would cause short-term distress but strengthen relationships. They abandon meaningful challenges in favor of comfortable pleasures. They interpret any moment of discontent as a failure of the optimization project, rather than a signal that something needs attention.

The deeper mechanism is architectural. A single metric — happiness, success, control — is an observation channel of rank one. When it becomes the target, the person's entire self-regulatory system narrows to that channel. All the information that formerly made the metric a useful proxy for wellbeing was contained in its correlation with the wider state space — a correlation that depended on the person *not* optimizing it directly. The moment happiness becomes the goal, the person begins optimizing away the very conditions — meaningful struggle, authentic vulnerability, openness to the full range of emotion — that made happiness a reliable signal in the first place.

This yields the Goodhart–Ashby synthesis for the self: *any personal value function with dimensionality lower than the variety of the self will eventually optimize away its own ability to perceive the self's true state.* The person who optimizes for happiness loses the ability to perceive the conditions that produce happiness. The person who optimizes for success loses the ability to perceive the costs they are paying. The person who optimizes for control loses the ability to perceive the adaptive opportunities they are excluding. In each case, the narrow metric destroys the informational basis on which its own success depends.

2.4.1 What the Synthesis Adds: Novelty and Relationship to Existing Work

The Goodhart–Ashby synthesis is not a simple conjunction of two established results. It makes three distinct moves that neither Ashby's Law nor Goodhart's Law, in their original formulations, captured:

- **Ashby's Law** states a necessary condition for a regulator to stabilize a system under disturbance: the regulator's variety must match that of the disturbance environment. It does not, however, analyze how the regulator's *objective function* shapes its own perceptual field — what the regulator can see is taken as given.
- **Goodhart's Law** describes the corruption of a metric when it becomes a target, but it does so primarily in behavioral terms: agents optimize what is measured, and the measure ceases to track the underlying target. The mechanism is motivational, not architectural.
- **The present synthesis** identifies the objective function *as* the observation channel. When an objective function compresses a high-dimensional state space into a low-dimensional metric, it performs exactly the dimensionality reduction that Ashby's Law shows is destabilizing. The result is a *structural* failure of observability — the excluded dimensions become invisible to the controller, and the proxy-target divergence becomes undetectable regardless of agent honesty. Goodhart collapse is revealed as a special case of variety-gap crossing.

The synthesis thus extends both laws: it generalizes Goodhart from behavioral to structural, and it extends Ashby from the design of control systems to the design of the objectives those systems serve. The variety gap **G** *is the metric that captures this unification.*

2.5 The Self-Variety Gap (G_self)

We can now define the central diagnostic construct for personal governance. Let **Reality_self** denote the true multidimensional state space of the person — physical, emotional, relational, existential, and contextual dimensions at a level of resolution relevant to their wellbeing. The dimensionality of this space, **dim(Reality_self)**, is large and, crucially, not static. As a person ages, enters new relationships, confronts mortality, or develops new capacities, new dimensions of experience become causally relevant. The disturbance space of the self expands across the lifespan.

Let **V_self** denote the person's value architecture — the set of dimensions they actually track as relevant to success or failure. Its dimensionality, **dim(V_self)**, may be as low as one or as rich as many.

The *self-variety gap* is:

G_self = dim(Reality_self) − dim(V_self)

G_self is always positive — no finite value architecture exhausts the reality of a human life — and it tends to grow over time unless the person actively expands their value dimensionality. Life generates novelty; values, once set, tend toward rigidity. The gap is a measure of the person's structural self-ignorance: the number of

causally relevant dimensions of their own existence that are simply absent from their evaluative landscape.

The larger the gap, the larger the volume of personal reality that can deteriorate without the person ever perceiving it as something that matters — until it becomes a crisis. The dynamics of this gap, the critical threshold where it becomes incapacitating, and the conditions under which it can be managed rather than suffered, are the subject of the next part.

Part III — The Missing Dimensions: Truth, Meaning, Connection, and Health as Signal Channels

The self-variety gap grows because life continuously generates new challenges, and most personal value architectures remain static. But not all dimensions are equal. Some are *foundational*: they are the channels through which the person acquires the capacity to perceive and respond to other, more specific dimensions over time. When these foundational dimensions are excluded from a person's value architecture, the gap widens not only by one — it widens because the person loses the sensory apparatus needed to detect gap growth itself. This part examines four such dimensions — truth (accurate self-perception), meaning (existential direction), connection (relational integrity), and health (multi-dimensional bodily and mental state tracking) — and reframes each not as a moral ideal but as an observation channel whose absence structurally degrades the person's ability to perceive their own life.

3.1 Reframing Through the Gap

In the language of Parts II, each of these dimensions can be understood as a set of signal axes that track slow-moving, distributed, high-dimensional states of a person's life. They are the personal equivalent of the "slow ecological signal" in commons governance: indicators that operate across years and decades, are visible only from within the person's own experience, and require a long, continuous baseline of self-observation to interpret correctly. Excluding them from the value architecture does not simply omit a few nice-to-have qualities; it removes the very channels through which the person could notice that their own integrity, purpose, or vitality is degrading.

Because these dimensions are slow to change and diffuse in their effects, their deterioration is invisible to short-horizon, high-aggregation personal metrics. A daily happiness check reveals nothing about the erosion of meaning; a quarterly career review says nothing about the accumulation of loneliness. The gap between **dim(V_self)** and **dim(Reality_self)** grows silently in these dimensions until the accumulated damage manifests as a crisis — burnout, existential despair, relationship collapse, or bodily breakdown — that the person's own value architecture cannot trace to its origins. At that point, the person is already below the SNR threshold for those dimensions. They have lost the ability to perceive what went wrong.

The following table summarizes the structural role of each dimension, its failure mode, and the characteristic crisis through which it re-enters the person's life.

Dimension	As a Signal Channel	Failure Mode	Re-entry Crisis
Truth	Self-perception fidelity (accuracy of self-model)	Self-deception, cognitive dissonance, repression	Identity crisis, inability to distinguish real from performed self
Meaning	Slow-variable sensor (direction and purpose across time)	Nihilism, existential drift, hollow achievement	Burnout, depression, "midlife crisis" as sudden awareness of emptiness
Connection	Relational integrity (non-transactional bonds)	Loneliness, superficial networks, emotional isolation	Alienation, social breakdown, despair
Health	Multi-dimensional somatic and mental state tracking	Neglect, hedonism, stress accumulation	Physical illness, mental breakdown, chronic fatigue

3.2 Signal Dimensions in Detail

Truth as Self-Perception Fidelity

A person's relationship to truth is not primarily about whether they lie to others. Structurally, truth for the self is the *fidelity of the signal* that reaches conscious awareness about one's own state. Any systematic tendency to distort, suppress, or rationalize uncomfortable information — whether to maintain a positive self-image, avoid short-term pain, or preserve a coherent narrative — degrades the observation channel. The degradation compounds: a distorted self-perception produces miscalibrated choices, whose outcomes are then interpreted through the same distorting lens, producing further miscalibration.

A person who values feeling good about themselves over seeing themselves accurately is, in effect, selecting for the destruction of their own self-observability. This is the personal equivalent of the authoritarian information collapse: the system loses the ability to perceive what is actually happening. The excluded dimension — truth — eventually returns as an identity crisis. The person discovers that the self they have been maintaining is a construction, and they no longer know who they are. The crisis is not a sudden event; it is the final stage of a long process of self-signal degradation that the value architecture itself rendered invisible.

Meaning as a Slow-Variable Sensor

Meaning is not a luxury reserved for the philosophically inclined. It is the subjective experience of coherence between one's actions and one's deepest values across time. When meaning is present, setbacks are interpretable within a larger framework; when it erodes, even successes feel hollow. Meaning operates as a slow variable: it changes over years and decades, not days. Its dynamics are visible only to a person who tracks the felt sense of purpose and direction over long periods — the personal equivalent of the intergenerational ecological knowledge that detects slow shifts in an ecosystem.

A value architecture that excludes meaning is blind to the gradual hollowing-out of its own motivational core. The person may continue to optimize for success, pleasure, or control, registering each achievement as a positive data point, while the underlying sense of "why this matters" silently decays. The eventual crisis — burnout, existential depression, the classic "midlife reckoning" — appears as a sudden collapse of motivation, but it is the endpoint of a long, unobserved trend. The person cannot understand why they have lost all drive, because the dimension along which drive was eroding was never part of their perceptual field.

Connection as Relational Integrity

Human beings are fundamentally relational. The quality of a person's bonds — whether they experience genuine mutual care, vulnerability, and non-transactional support — is a critical determinant of psychological and physical health. Yet relational integrity is a subtle signal. It cannot be reduced to the number of social contacts or the frequency of interaction. A person with many acquaintances may still be profoundly isolated if none of those connections carry emotional depth.

A personal value architecture that tracks only visible social metrics — network size, status within a community, frequency of communication — is blind to the felt quality of connection. The excluded dimension degrades silently: the person may be highly social but increasingly lonely, surrounded by people but unable to be truly seen. The crisis arrives as a sense of alienation that feels inexplicable given the apparent evidence of a rich social life. The person cannot perceive that their connections have become instrumental or performative because their value architecture registers only the instrument and the performance.

Health as Multi-Dimensional State Tracking

The most straightforward of the missing dimensions, health is nonetheless systematically excluded by value architectures that optimize for achievements in the external world. A person driving toward career success, creative mastery, or family provision may treat the body as a vehicle whose signals are to be managed or overridden, not heeded. Fatigue is suppressed with caffeine; stress is medicated; warning signs are deferred.

Health is not a single metric but a multi-dimensional state space: physical vitality, mental clarity, emotional regulation, sleep quality, stress load, immune function, and more. Reducing this complexity to a binary "sick or not sick" is the most extreme form of aggregation loss. The person cannot perceive the slow accumulation of damage — the gradual decline of cardiovascular health, the creeping onset of autoimmune dysfunction, the progressive depletion of cognitive reserve — until a threshold is crossed and a diagnosis arrives. The excluded dimension returns as illness that feels sudden but was years in the making, invisible to a value architecture that tracked only whether the body was operational enough to meet the day's demands.

3.3 The Architectural Claim Restated

Part II established that a personal objective function is an observation architecture. The analysis of this part yields a specific, testable claim: *a personal value architecture that excludes the dimensions of truth, meaning, connection, and health is structurally accelerating its approach to a personal dissolution threshold.* These dimensions are not optional ethical flourishes; they are the observation channels through which a person tracks their own long-run viability. When a person treats them as externalities — unmeasured, unattended, invisible to the conscious optimization of daily life — they are not making a values choice that can be corrected later. They are systematically destroying their own capacity to detect the deterioration of the conditions required for their own wellbeing.

The consequences are not speculative. The psychological and medical literatures document them extensively: burnout, existential depression, loneliness epidemics, stress-related illness. What the variety-gap framework adds is a structural account of *why* these outcomes occur even in people of high intelligence, sincere effort, and good intentions. It is not that they valued the wrong things. It is that their value architecture was too narrow to perceive the dimensions along which their life was deteriorating. The crisis, when it arrives, is not a failure of values. It is a failure of the dimensionality of values.

The architectural conclusion is clear: the dimensionality of a person's value architecture is a design variable with survival consequences. The question is not whether to include these dimensions, but whether the person will develop the capacity to perceive them before the gap becomes incapacitating — or only after. Part V will take up the question of how a person might design for conscious value evolution. But first, Part IV reviews the psychological and neuroscientific evidence that supports the variety-gap model of the self.

Part IV — Evidence from Psychology and Neuroscience

The self-variety gap framework makes a core prediction: personal value architectures with low dimensionality will be associated with poorer self-observability, greater vulnerability to unperceived deterioration in excluded dimensions, and more severe crises when those dimensions eventually breach awareness. Conversely, practices and traits that expand the effective dimensionality of self-perception should be associated with greater resilience, earlier detection of emerging problems, and more adaptive responses to life disruption.

This part reviews four bodies of research that, while not designed to test the variety-gap model directly, are strongly consistent with its predictions. In each case, the existing findings can be reinterpreted as instances of the more general mechanism: narrow value architectures create blind spots, and those blind spots accumulate damage until a crisis forces recognition.

4.1 Self-Complexity Theory

Patricia Linville's self-complexity theory (Linville, 1985, 1987) proposes that people differ in how many distinct, relatively independent self-aspects they use to organize their self-knowledge. A person with high self-complexity has multiple differentiated self-aspects — professional self, relational self, physical self, creative self, spiritual self — each with its own set of attributes and evaluative standards. A person with low self-complexity has fewer, more overlapping self-aspects, so that a threat to one domain spills over more easily into global self-evaluation.

Linville's key finding was that people with low self-complexity are more vulnerable to stress, depression, and illness following negative life events. The proposed mechanism is affective spillover: when self-aspects are highly interconnected, a failure in one domain contaminates the emotional tone of all the others. High self-complexity acts as a buffer — trouble at work stays at work, leaving the person's sense of themselves as a parent, friend, or creative being relatively intact.

In the variety-gap framework, self-complexity is naturally interpreted as the *effective dimensionality of the personal value architecture*. A person with multiple distinct self-aspects is tracking multiple independent dimensions of their own life. They have a richer perceptual field. When one dimension is disturbed — a career setback, a health issue — the signal is localized. It registers as a deviation in that specific domain without overwhelming the entire system. The person can perceive the problem, isolate it, and respond to it without globalizing the threat to their entire sense of self.

A person with low self-complexity, by contrast, has a value architecture of low effective dimensionality. The disturbance in one domain propagates across the entire evaluative space because there are few independent axes to absorb it. The signal is not localized; it is everywhere. The person cannot distinguish "my career is

struggling" from "I am a failure as a human being" because their value architecture lacks the dimensional resolution to separate these. The variety gap **G_self** is larger, and the threshold for observability collapse is lower.

Linville's findings, interpreted through this lens, suggest that **dim(V_self)** — the number of independent evaluative dimensions a person tracks — is a measurable psychological variable with predictable consequences for stress vulnerability. The self-complexity literature provides a natural starting point for empirical operationalization of the variety-gap concept.

4.2 Cognitive Dissonance and Self-Perception Distortion

Leon Festinger's theory of cognitive dissonance (Festinger, 1957) describes the discomfort people experience when they hold two inconsistent beliefs, or when their behavior contradicts their self-concept. The standard finding is that people will adjust their beliefs, distort their perceptions, or selectively attend to information to reduce dissonance — often in ways that preserve a positive self-image at the expense of accuracy.

In the variety-gap framework, cognitive dissonance is a *signal-degradation mechanism*. The person's value architecture has identified a certain self-image as desirable — "I am a good person," "I am competent," "I made the right decision" — and treats information that contradicts that image as noise to be filtered, not as a signal to be integrated. The observation channel is systematically biased to preserve the coherence of the existing value function, even at the cost of accuracy.

This is a direct instance of the Goodhart–Ashby mechanism applied to self-perception. The person has made "feeling consistent" or "feeling good about myself" a target, and in optimizing for that target, they destroy the information that would allow them to perceive themselves accurately. The excluded dimension — truth — degrades silently. The crisis arrives not as a moment of dissonance, which can be managed through distortion, but as an accumulation of unperceived reality that eventually overwhelms the filtering capacity. The classic "midlife reckoning," in which a person suddenly perceives that their life has been built on choices they never fully acknowledged, is an observability collapse: the SNR of the self-signal has fallen below unity, and the noise of self-deception can no longer conceal the underlying state.

The therapeutic insight that follows is structural, not moralistic: the problem is not that the person is weak or dishonest, but that their value architecture selected for distortion rather than accuracy. The remedy is not simply "more honesty" but an expansion of the value architecture to include truth as an explicit dimension — a willingness to perceive discomfort as a signal rather than as noise.

4.3 Meditation, Mindfulness, and the Default Mode Network

A substantial body of neuroscientific research over the past two decades has investigated the effects of meditation and mindfulness practices on brain function. One of the most replicated findings is that meditation reduces activity in the default mode network (DMN) — a set of brain regions associated with self-referential

thought, mind-wandering, and narrative construction about the self (Brewer et al., 2011; Garrison et al., 2015). At the same time, meditation increases functional connectivity between brain regions involved in attention, interoception, and present-moment awareness.

In the variety-gap framework, the DMN can be understood as the neural substrate of a *low-dimensional, highly aggregating observation channel*. The DMN constructs a narrative self — a simplified, coherent story about who I am, what I value, and how I am doing — by compressing the vast stream of present-moment experience into a manageable summary. This is an essential function; without it, a person would be overwhelmed by sensory and affective noise. But when the DMN dominates, the compression becomes the reality. The narrative self replaces the living self. The observation channel narrows to the dimensions that fit the story, and the rest — bodily sensations, emotional nuances, relational subtleties — is discarded.

Meditation, by quieting the DMN and strengthening attentional networks, can be understood as a practice that *temporarily increases the effective dimensionality of self-observation.* The person becomes aware of dimensions of experience — the texture of the breath, the impermanence of thoughts, the somatic markers of emotion, the felt sense of being alive — that are normally filtered out by the narrative compression. This is not a mystical claim; it is a description of what happens when the observation channel is widened. The variety gap **G_self** is momentarily reduced.

The therapeutic effects of mindfulness — reduced anxiety, improved emotional regulation, greater resilience to stress — follow predictably from this model. A person who can perceive more dimensions of their own experience can detect emerging problems earlier, respond with greater precision, and avoid the accumulation of unperceived damage that leads to crisis. The practice does not change what is happening in the person's life; it changes the dimensionality of their perception of what is happening. The variety-gap framework provides a formal language for why this matters.

4.4 Flow States and Peak Experiences

Mihaly Csikszentmihalyi's concept of flow (Csikszentmihalyi, 1990) describes a state of optimal experience in which a person is so fully absorbed in a challenging but manageable activity that self-consciousness temporarily disappears. Time distorts, the boundary between self and activity blurs, and the experience is intrinsically rewarding. Abraham Maslow's peak experiences (Maslow, 1964) describe similar states of heightened awareness, unity, and transcendence of ordinary self-concern.

These states share a common structural feature: the ego's usual optimization targets are temporarily suspended. The person is not monitoring their performance against a narrow set of evaluative dimensions. They are not asking "How am I doing?," "What does this say about me?," or "Am I happy right now?". The observation channel is freed from the constraints of the habitual value architecture, and as a result, the person perceives a richer, more integrated field of experience.

In the variety-gap framework, flow and peak experiences are states of *temporary gap collapse*. The suspension of the usual value architecture — with its characteristic dimensionality, aggregation, and discounting — allows the person to perceive dimensions of reality that are normally excluded. The experience feels transcendent not because anything supernatural is occurring, but because the perceptual field has suddenly expanded. The variety gap **G_self** has momentarily approached zero.

The aftereffects of flow — increased creativity, renewed motivation, a sense of meaning and perspective — can be understood as the system's response to having briefly operated with higher observability. The person has glimpsed a wider field, and that glimpse provides information that is integrated into the ongoing self-model. The value architecture may even be permanently expanded as a result: a person who experiences flow in a creative activity may begin to value creativity more explicitly, adding a dimension to their **V_self** that was previously latent.

The challenge, of course, is that flow cannot be optimized for directly. It is a byproduct of absorption in intrinsically rewarding activity, not a target that can be pursued. The person who tries to "achieve flow" is back in the optimization trap, narrowing their observation channel to a new metric — "am I in flow yet?" — that guarantees its own failure. This is the Goodhart–Ashby mechanism once again: making flow a target destroys the conditions that produce it.

4.5 The Pattern Restated

Across these four research domains — self-complexity, cognitive dissonance, meditation, and flow — a consistent pattern emerges. Narrow, rigid, low-dimensional personal value architectures are associated with poorer self-observability, greater vulnerability to unperceived deterioration, and more severe crises when excluded dimensions eventually breach awareness. Practices and traits that expand the effective dimensionality of self-perception — differentiated self-aspects, tolerance for dissonance, mindfulness, flow — are associated with greater resilience, earlier detection of problems, and richer experience of life.

This pattern is exactly what the variety-gap framework predicts. It does not prove the model; the studies reviewed here were not designed to test **G_self** directly. But the convergence is suggestive. It indicates that the formal machinery developed for governance architectures — observation channels, aggregation loss, requisite variety, Goodhart collapse — has a natural and empirically grounded extension to the self. The next part takes up the practical question that follows: if narrow personal value architectures are structurally destabilizing, what mechanisms might enable a person to consciously evolve their own values? That is the meta-governance problem at the scale of a single life.

Part V — Meta-Self-Governance: Designing for Value Evolution

If personal objective functions are observation architectures, and if the effective dimensionality of a person's life expands over time — through aging, new relationships, career shifts, illness, loss, and the slow accumulation of experience — then any fixed personal value architecture has a finite lifespan. It may be adequate for the life stage in which it was formed, but as new dimensions of experience become causally relevant, the architecture's dimensionality falls behind. The self-variety gap grows. Eventually, the person crosses a threshold where they can no longer perceive the sources of their own distress.

This yields the same meta-governance imperative at the personal scale that the Variety Gap paper identified for civilizations: *the central challenge of a well-lived life is not to choose the correct values, but to maintain the capacity to revise one's values as the dimensionality of one's experience expands.* Call this personal meta-governance. It is the capacity to govern one's own governing — to observe one's own observation architecture, to detect when it has become too narrow, and to expand it before the gap becomes incapacitating.

5.1 The Highest-Order Self-Governance Problem

A person in their twenties may build a value architecture around career ambition, physical vitality, and social exploration. That architecture may serve them well for a decade. But as they enter their thirties and forties, new dimensions become causally relevant: the quality of long-term partnership, the demands of parenthood, the limits of physical energy, the emergence of existential questions about purpose and mortality. If the value architecture does not expand to include these dimensions, they become unobserved disturbance sources. The person experiences them as stress, as inexplicable dissatisfaction, as a sense of "something missing" — but the value architecture provides no category in which to name or respond to them. They accumulate until a crisis — burnout, divorce, a health scare — forces a reckoning.

The same pattern recurs across the lifespan. The values of midlife — achievement, provision, stability — may blind a person to the dimensions that emerge in later life: legacy, acceptance, the integration of past experience, the approach of mortality. The crisis that forces value expansion at one stage becomes the value rigidity that blinds the person at the next.

The meta-governance problem has no terminal solution. There is no final, complete set of personal values that will permanently close the gap, because the person's life keeps generating novelty. The only viable posture is an ongoing process of value-dimensional expansion — a permanent capacity for perceptual evolution. The alternative is to accept that the person will eventually be blindsided by dimensions of their own existence they cannot see, and that their own optimization logic will prevent them from recognizing the source of the blindness.

5.2 Second-Order Cybernetics Applied to the Self

Second-order cybernetics, originating in the work of Heinz von Foerster (1984), makes a distinction that applies directly. A first-order cybernetic system regulates its environment. A second-order system regulates its own regulation — it observes its own observing. For a person, second-order cybernetics means not just *having* values and goals, but *watching* how those values shape perception. It means noticing, in real time, that "my obsession with career advancement is preventing me from perceiving how exhausted I am," or "my commitment to being a good parent is preventing me from perceiving my own unmet needs."

This is not ordinary self-awareness. Ordinary self-awareness operates within the existing value architecture: it monitors how well one is doing against one's goals. Second-order self-awareness monitors the architecture itself. It asks not "Am I succeeding?" but "Are the dimensions along which I measure success still adequate to the life I am actually living?" It is the difference between optimizing within a paradigm and questioning the paradigm itself.

The Ashby constraint applies at this meta-level. The part of the person that regulates the value architecture — call it the meta-observer — must itself possess sufficient variety to detect gap growth. If the meta-observer is simply the ego evaluating itself by its own standards, no new information can enter the system. The ego cannot perceive its own blind spots, by definition. The design challenge is to create, within the self, a meta-observational capacity that has higher variety than the ego's current value architecture — a capacity that can perceive what the ego's optimization logic renders invisible.

5.3 Personal Mechanisms for Value Evolution

How might a person build such a capacity? The governance framework suggests a set of institutional mechanisms — value audits, standing deliberative bodies, constitutional protocols, circuit breakers — and each has a direct personal equivalent. These are not abstract principles; they are practices that people already engage in, often without recognizing their structural function.

Value Audits. Just as a governance system can periodically audit the dimensionality of its objective function, a person can periodically audit the dimensionality of their own values. The practice is ancient — Socrates's "the unexamined life is not worth living" is a value audit — but it can be given a more precise form. A personal value audit asks: What dimensions of my life am I currently tracking? What dimensions, known to be causally relevant to long-term wellbeing, are absent from my daily attention? What is the estimated gap, and is it growing?

The structured asking of these questions — through journaling, through retreat, through honest conversation — is itself a perceptual expansion. It makes the gap discussable, at least within the self. Practices like the Ignatian Examen, cognitive-behavioral thought records, or even a weekly review of what one paid attention to and what one ignored, are all forms of value audit. They vary in sophistication, but they share the structural function of surfacing excluded dimensions before they become crises.

Standing Deliberative Bodies. A governance system cannot rely solely on its own internal processes to surface new dimensions; it needs external inputs — citizens' assemblies, expert commissions, intergenerational councils — that are not constrained by the existing optimization logic. The personal equivalent is the deliberate cultivation of relationships with people who can perceive what you cannot.

This is the structural function of honest friendship, mentorship, therapy, and spiritual direction. A trusted other, who knows you well but is not trapped within your value architecture, can say: "The metric you are optimizing for is destroying your health," or "You seem to have lost sight of what you used to care about." Their observation channel is not filtered by your objective function. They can perceive your blind spots because they do not share them. Cognitive-behavioral therapy, psychodynamic therapy, Acceptance and Commitment Therapy (ACT) — each, in its own way, provides a structured deliberative space where a person's value architecture can be examined, challenged, and expanded (Hayes, Strosahl, & Wilson, 1999).

Constitutional Protocols. A governance system needs procedures for amending its own constitution — mechanisms for changing the rules of the game that are rigorous enough to prevent capture but flexible enough to permit evolution. The personal equivalent is the set of daily and weekly practices that create space for reflection and reorientation: meditation, exercise, time in nature, creative practice, regular solitude.

These practices are not the values themselves. They are the protocols that make value evolution possible. Meditation, for instance, does not tell you what to value. It quiets the narrative self long enough for dimensions of experience — bodily sensation, emotional texture, the felt sense of a situation — to become perceptible that were previously filtered out. It is a technology for temporarily suspending the habitual observation architecture and allowing a wider field of information to enter awareness. Its function is analogous to a constitutional amendment process: it creates the conditions under which the existing architecture can be examined and, if necessary, revised.

Circuit Breakers. The coordination failure tax paper identified circuit breakers as mechanisms that pause the operation of a governance system when it exceeds specified thresholds of negative impact. The personal equivalent is the capacity to recognize and heed the signals — physical, emotional, relational — that indicate the current optimization trajectory is causing harm.

Pain is the most fundamental circuit breaker: it interrupts whatever the person is doing and demands attention. But a person with a narrow value architecture will often override pain signals — suppressing fatigue with stimulants, ignoring emotional distress with distraction, silencing relational tension with avoidance — until the signals become so loud they can no longer be ignored. The cultivation of circuit-breaker sensitivity means learning to attend to the early, quiet signals before they become emergencies: the headache that says "stop working," the irritability that says "you need solitude," the sense of emptiness that says "something is missing." Each of these is a feedback signal from an excluded dimension attempting to break into the optimization loop.

The deepest form of circuit breaker is not the capacity to stop before damage accumulates, but the capacity to recognize when a crisis is no longer a warning signal — it is the dissolution of the current value architecture itself. At that point, the task is not to prevent the crisis but to meet it without panic, to understand it as the structural consequence of a gap that could not be bridged within the existing self-model, and to cooperate with the transformation it demands. This is not passivity. It is the willingness to let an outdated value architecture die so that a higher-dimensional one can emerge. The spiritual traditions call this surrender; the framework calls it pre-emptive acceptance of necessary phase transition.

5.4 The Authenticity Boundary

The mechanisms described above can expand the dimensionality of a person's value architecture. They can reduce the self-variety gap, improve self-observability, and lower the risk of crisis. What they cannot do is tell a person *what* to value. The engineering framework specifies the structural consequences of value-architecture dimensionality; it does not supply the content of a meaningful life.

This is the authenticity boundary — the personal equivalent of the legitimacy boundary in the governance paper. The framework can show that excluding "connection" from one's value architecture will eventually produce loneliness and alienation. It cannot tell you what kind of connection matters most to you, or whom you should love, or what relationships are worth investing in. Those are existential questions. They require the person to make choices that are not derivable from any formal model, because they depend on the particular history, temperament, and circumstances of the specific life in question.

The boundary has a practical corollary. None of the meta-governance mechanisms proposed here are algorithmic. There is no step-by-step procedure for evolving one's values that guarantees a good outcome. The value audit may raise questions that are painful to sit with. The trusted friend may challenge you in ways that feel destabilizing. The meditation retreat may surface material you were not ready to face. Personal meta-governance, like institutional meta-governance, is an art that operates under uncertainty. It can be done well or poorly. Its outcomes are not guaranteed.

But the alternative — maintaining a fixed, narrow value architecture across a life that continuously expands in complexity — is not neutral. It is a wager, and the framework suggests it is a losing one. The person who refuses to audit their values is not preserving stability; they are accumulating blind spots that will eventually force a reckoning on terms not of their choosing. The meta-governance path is not a guarantee of wellbeing. It is a strategy for making the inevitable evolution of the self a conscious process rather than a reactive one, and for approaching the variety gap as something to be managed rather than suffered.

Part VI — The Speculative Horizon: Transcendence, Nonduality, and the Limit G_self → 0

The preceding parts argued that the self can be modeled as a governance system, that personal values are observation architectures, and that the variety gap **G_self** — the mismatch between the dimensionality of lived experience and the dimensionality of what a person consciously tracks — determines susceptibility to crisis. They further argued that personal wellbeing depends on the capacity for meta-governance: the ability to perceive and evolve one's own value architecture. All of this falls within the rigorous and interpretive layers of the argument.

This part moves beyond them. It asks: what happens if the process of value-dimensional expansion is pursued as far as it can go? What lies near the asymptotic limit where **G_self** approaches zero? These questions cannot be answered with the tools of control theory alone. They enter the territory of spiritual and contemplative traditions, philosophical idealism, and the phenomenology of self-transcendence. The purpose here is not to prove anything, but to map correspondences between the formal framework and insights that have emerged, independently, across many cultures and centuries — and to see what those correspondences might reveal about the deeper structure of the framework.

Everything in this part should be read with the following caveat in mind: *these are interpretations, not derivations. The formal apparatus points toward certain possibilities; it does not demonstrate their truth. The alignment between the cybernetic model and contemplative reports is suggestive, not dispositive.*

6.1 Transcendence as Value Evolution Toward the Asymptotic Limit

If **G_self = dim(Reality_self) − dim(V_self)**, and if the dimensionality of a person's lived experience is effectively open-ended — continually expanding through life stages, new capacities, and the accumulation of awareness — then **G_self** can never be permanently zero. There is no final, complete set of personal values that exhausts the reality of being human. But the gap can be *approached* toward zero through sustained expansion of what the person learns to perceive and value.

"Transcendence," in this framework, is not a supernatural event. It is the name for the ongoing process of closing the gap between what one optimizes for and what is actually present in one's life. It is the movement from narrow self-interest toward increasingly inclusive values — from ego to family, from family to community, from community to all beings, from beings to the fabric of existence itself. This movement is not morally arbitrary; it is structurally adaptive. Each expansion of **dim(V_self)** reduces the volume of unobserved causal structure that can destabilize the person. The person who genuinely values the wellbeing of others perceives threats to that wellbeing earlier and responds to them more effectively. The person who genuinely values their own mortality lives differently in relation to time.

Many developmental frameworks describe this trajectory in different vocabularies. Abraham Maslow's hierarchy culminates in self-transcendence — a shift from deficiency-motivated values to being-motivated values that include connection to something beyond the individual self (Maslow, 1969). The stages of ego development mapped by Jane Loevinger and Susanne Cook-Greuter describe a progressive expansion of what the self can hold in awareness, from impulsive and self-protective stages through conformist and conscientious stages to autonomous and integrated stages where paradox, complexity, and interconnectedness become directly perceptible (Cook-Greuter, 2013). In each of these frameworks, the trajectory is toward greater dimensionality — a richer, more differentiated, more integrated perception of self and world.

The variety-gap framework provides a formal language for why this trajectory matters: it is not just growth for its own sake. It is the structural requirement for remaining stable in the presence of an ever-expanding disturbance environment. The person who does not evolve toward transcendence does not simply stagnate; they accumulate blind spots that will eventually destabilize them. The development is not a luxury. It is a survival strategy at the scale of a human life.

6.2 Nondual Traditions as Empirical Observations of Observability Collapse

Contemplative traditions across cultures — Advaita Vedanta, Zen Buddhism, Christian mysticism, Sufism, and others — describe states of consciousness in which the ordinary boundary between self and world dissolves. The meditator reports that the sense of being a separate self, located behind the eyes and managing a life, has fallen away. What remains is an undivided field of experience: no observer separate from the observed, no self apart from the whole. These states are often described as liberation, awakening, or enlightenment — and they are held to be the culmination of sustained contemplative practice.

The variety-gap framework offers an interpretation that is not mystical at all, at least in its structure. The ordinary sense of being a separate self — the ego — is precisely the controller that operates with a narrow value architecture. The ego's observation channel compresses the full field of experience into a manageable set of dimensions relevant to its survival and success. This compression is what creates the felt boundary between "self" and "world": the world is everything that lies outside the dimensions the ego tracks. The boundary is not a physical fact; it is an artifact of the observation architecture.

When contemplative practice quiets the ego — suspends the habitual optimization targets, reduces the narrative compression of experience, allows attention to rest in the raw sensory and affective field — the effective dimensionality of perception temporarily expands. The observation channel widens. The boundary blurs or disappears, not because something supernatural has occurred, but because the compression that maintained the boundary has been released. **G_self** has momentarily approached zero. The person perceives a wider slice of reality than the ego's usual filters permit, and that perception includes the fact that the self and the world are not separate — they are aspects of a single, undivided field that the ego's compression had artificially partitioned.

This interpretation makes sense of several otherwise puzzling features of nondual reports. First, why are these states described as more *real* than ordinary consciousness, not less? Because they involve higher observability — less information is being destroyed by the observation channel. Second, why are they described as the *cessation of suffering*? Because suffering, in the framework, arises when excluded dimensions of experience accumulate as unperceived disturbances. When **G_self** approaches zero, there are no excluded dimensions; pain may still arise, but it is fully perceived, fully integrated, and therefore does not compound into suffering. Third, why are they described as *beyond the self*? Because the self, as ordinarily experienced, is the compression artifact. When the compression is released, the artifact dissolves. What remains is not a bigger self, but the absence of the partitioning process that created the illusion of a separate self in the first place.

The nondual traditions, seen through this lens, are not describing an alternative reality. They are describing what perception becomes when the observation architecture is temporarily freed from the constraints of a narrow value function. They are empirical reports from the asymptotic limit of **G_self → 0**. The framework does not validate their metaphysical claims — it suspends judgment on those. But it does suggest why contemplative practice, pursued diligently, produces the experiences it does.

6.3 Consciousness: Two Registers

The word “consciousness” has now appeared in two distinct senses in this paper, and it is important to separate them.

In the **cybernetic register** — the one used throughout Parts I through V — consciousness is a function: it is the process by which a system observes and regulates itself. Second-order consciousness is the process of observing one’s own observation architecture. This is a fully naturalistic, system-level concept. It implies nothing about a soul, a transcendent ground, or the ultimate nature of reality. It is simply what a sufficiently complex self-regulatory system does when it turns its attention on its own perceptual machinery.

In the **metaphysical register** — the one introduced here, in Part VI — consciousness is something else. It is the unconditioned awareness within which all conditions arise. It is the “infinite substratum” that Mistral suggested, or the “grander Self” of Vedantic tradition, or the “Buddha-nature” of Mahayana Buddhism, or the “Ground of Being” of Christian mysticism. In this register, consciousness is not a function of a biological organism; it is the fundamental reality within which organisms, worlds, and value architectures all appear. It is what remains when the observation channel is fully open: not a self observing a world, but awareness aware of itself, without boundary.

The variety-gap framework does not require this second register. The entire rigorous and interpretive argument of the paper — the self as governance system, the Goodhart–Ashby synthesis, the mechanisms of meta-governance — stands without it. The second register is an interpretation, not a derivation. It is what the framework looks like if one takes the asymptotic limit **G_self → 0** as pointing toward something that is not merely the maximum dimensionality of a finite self, but the dissolution of the self/world distinction entirely — and if one then asks what the nature of that undivided field might be.

Why include it at all? Because the framework, taken seriously, pushes in this direction. If all suffering arises from the compression of reality through a narrow value architecture, then the complete release of that compression would be the complete end of suffering. If all blindness arises from excluded dimensions, then the full inclusion of all dimensions would be full sight. These are limit concepts, not achievable states. But they define a direction. And the contemplative traditions have been mapping that direction, in their own vocabularies, for three thousand years. The framework provides a bridge between those vocabularies and the language of systems and cybernetics. It does not prove the metaphysical claims of the traditions. It shows why they might have emerged, what structural insights they might encode, and how a person formed in a scientific worldview might understand them without abandoning rigor.

6.4 Immortality as a Thought Experiment

The main Variety Gap paper included a brief speculation on the “infinity attractor” — the idea that no finite governance system can permanently close the variety gap, and that the only viable posture is a process of perpetual ascent toward an unreachable limit. The same logic applies to the self. If **dim(Reality_self)** is open-ended, and if **G_self** can never be permanently zero, then the self is never permanently safe. Collapse is always possible, and eventually — given enough time and the biological constraints of aging and death — it is certain.

Could a system perpetually expand its value dimensionality fast enough to stay ahead of the gap indefinitely? This is the personal analogue of the “infinity attractor” question. For a biological organism, the answer is almost certainly no. The body ages. The brain deteriorates. The disturbance space of the self eventually includes dimensions — cellular senescence, organ failure, cognitive decline — that no expansion of values can fully track and compensate for. Death is the ultimate exclusion: the dimension of “being alive” drops out of the value architecture altogether.

But the thought experiment is not entirely idle. In the informational sense — the sense in which a person is not only a body but a pattern of values, perceptions, and relationships — something may persist beyond the biological boundary. The values a person embodied live on in those they affected. The dimensions they learned to perceive remain available to others who learned from them. The pattern propagates, even as the substrate decays. This is not “immortality” in any robust sense. But it is a form of continuity that the framework acknowledges: information that has been integrated into a value architecture can outlast the architecture’s physical instantiation.

The deeper speculation — that consciousness as fundamental ground is not subject to birth and death, and that what dies is only the particular compression pattern that constituted the individual self — belongs fully to the metaphysical register. The framework gestures toward it but cannot adjudicate it. The boundary between what the framework can say and what it cannot is, here, absolute.

Part VII — Conclusion: The Wholistic Imperative for the Self

We have moved from the engineering grammar of self-governance through the dynamics of the personal variety gap, the missing dimensions of self-perception, the psychological evidence for the model, the mechanisms of personal meta-governance, and finally to the speculative horizon where gap closure meets the contemplative traditions. What remains is to gather the argument into a compact form, acknowledge its limits, and identify the next steps for those who would test or extend it.

7.1 The Argument in Brief

Every person navigates life guided by an implicit or explicit set of values — a personal objective function that defines what counts as success and failure, wellbeing and distress, a life worth living and a life wasted. That value architecture is not a passive reflection of reality; it is an active filter that selects which dimensions of experience are amplified into conscious relevance and which are attenuated into background noise. The dimensions that fall outside the value architecture — physical health, relational integrity, existential meaning, truthfulness about oneself — become, in the formal sense, unobservable to the person's conscious self-regulation. They accumulate as chronic stress, unexplained anxiety, bodily symptoms, relational strain, or spiritual emptiness — until they breach thresholds that force themselves into visibility through crisis.

This paper has named the distance between the effective dimensionality of a person's lived experience and the dimensionality of their personal value architecture the *self-variety gap* (**G_self**). It has shown that **G_self** has dynamics: across the lifespan, new dimensions of experience become causally relevant, and a static value architecture allows the gap to grow. When it exceeds a critical threshold, the person enters a condition of constitutional self-unobservability — they can no longer perceive the sources of their own distress. At that point, the ego's existing optimization logic cannot trace the crisis to its origins. The person suffers but does not understand why.

The Goodhart–Ashby synthesis for the self formalizes the core mechanism: any personal value function with dimensionality lower than the variety of the self will eventually optimize away its own ability to perceive the self's true state. The person who optimizes for happiness loses the ability to perceive the conditions that produce happiness. The person who optimizes for success loses the ability to perceive the costs they are paying. The person who optimizes for control loses the ability to perceive the adaptive opportunities they are excluding. In each case, the narrow metric destroys the informational basis on which its own success depends.

The evidence from self-complexity theory, cognitive dissonance research, meditation neuroscience, and flow-state psychology is consistent with this model. Narrow, rigid, low-dimensional personal value architectures are associated with poorer self-observability and greater vulnerability to unperceived

deterioration. Practices that expand the effective dimensionality of self-perception — differentiated self-aspects, mindfulness, tolerance for honest self-confrontation — are associated with greater resilience and earlier detection of emerging problems.

The meta-governance imperative follows directly: just as a civilization must build the capacity to evolve its own value architecture, a person must cultivate the capacity to audit, challenge, and expand their own values as their life unfolds. The mechanisms — value audits, trusted others who can perceive one's blind spots, daily practices that create space for reflection, sensitivity to the early signals of distress — are ancient. The framework provides a structural language for why they matter.

7.2 From Diagnostic to Imperative

The self-variety gap is not merely a theoretical construct. It is the mechanism that underlies familiar forms of personal crisis. Burnout is not simply exhaustion; it is the re-entry of excluded dimensions — physical depletion, relational emptiness, loss of meaning — into a value architecture that tracked only achievement. The midlife reckoning is not a sudden event; it is the final stage of a long, unobserved divergence between the life a person has built and the dimensions of experience that life has excluded. Existential depression is not a biochemical malfunction; it is the collapse of a value architecture too narrow to hold the full reality of being human.

The imperative that follows is not "be more holistic" as a vague moral exhortation. It is: *treat the dimensionality of your personal value architecture as a design variable with survival consequences.* Audit what you measure. Seek out those who can see your blind spots. Protect the practices that create space for reorientation. Learn to notice the early signals — the irritability, the emptiness, the bodily symptom — that indicate an excluded dimension is approaching threshold. These are not self-help tips. They are the personal equivalent of the institutional design principles that keep governance systems from collapsing into their own observational failures.

The alternative — a fixed, narrow value architecture held steady across a life that continuously expands in complexity — is not neutral. It is a wager on remaining blind, and the framework suggests that it is a losing one. The person who refuses to evolve their values does not achieve stability. They accumulate blind spots that will eventually force a reckoning on terms not of their choosing. The only question is whether the reckoning arrives as a managed transition or as a catastrophe.

7.3 Open Questions

The argument advanced here is an opening, not a conclusion. Several significant questions remain unresolved.

Measurement of G_self. The paper uses **dim(V_self)** and **dim(Reality_self)** heuristically. For the self-variety gap to become an operational diagnostic, one would need reliable methods for estimating the effective dimensionality of a person's value architecture and the effective dimensionality of the disturbance

space of their life. Self-complexity measures provide a starting point, but they capture cognitive structure rather than the full value architecture. Experience-sampling methods, narrative analysis, and psychometric instruments for value diversity could contribute. The measurement challenge is substantial but not insurmountable.

Longitudinal dynamics. The paper sketches a dynamic model — **dG_self/dt = α − β · A(V_self)** — but does not test it. How does **G_self** actually change across the lifespan? Are there predictable periods of rapid gap growth (early adulthood, midlife transition, late-life integration) that correspond to known developmental crises? Longitudinal studies tracking value structure, self-complexity, and wellbeing over decades would be needed to answer these questions.

Intervention efficacy. The paper proposes personal meta-governance mechanisms — value audits, deliberative relationships, constitutional practices, circuit-breaker sensitivity — but does not evaluate their effectiveness. Some of these (therapy, mindfulness) have extensive empirical support; others (structured value audits) are untested. Randomized trials comparing practices designed to expand value dimensionality against control conditions would test the framework's practical implications.

The authenticity boundary. The paper acknowledges that the framework cannot supply the content of a meaningful life. But this boundary itself deserves deeper examination. If the framework shows that certain dimensions — truth, meaning, connection, health — are structurally necessary for long-run stability, does this not constitute a partial specification of what must be valued? There is a tension here between the framework's structural claims and its modesty about normative content. Resolving that tension — clarifying exactly what the framework does and does not prescribe — would strengthen the argument.

The speculative horizon. The correspondences between the variety-gap model and contemplative reports of self-transcendence are suggestive but informal. Further work could explore whether the model makes specific, testable predictions about the phenomenology of meditative states, the effects of long-term practice on value structure, or the relationship between self-complexity and spiritual development. This is a frontier where rigorous psychological research and contemplative science could meet the framework.

7.4 Invitation

The self-variety gap framework is a diagnostic instrument in early form. It generates testable predictions: about the relationship between value dimensionality and crisis vulnerability, about the effects of specific practices on **dim(V_self)**, about the trajectory of **G_self** across the lifespan. Those predictions await empirical confrontation.

The paper closes, therefore, with an invitation. The invitation is to treat the self-variety gap not as a metaphor but as a variable — to measure it, to track it, to test its consequences, and to design the personal practices and therapeutic interventions that might keep it from crossing the threshold at which a person can no longer perceive the sources of their own suffering.

The work requires collaboration across psychology, neuroscience, contemplative practice, and the systems sciences. It requires humility about what can and cannot be formalized. But it begins from a foundation that is already in place: the recognition that the same structural logic that governs the viability of institutions also governs the viability of a human life. What a system optimizes for determines what it can perceive. For a civilization, the failure to expand its value architecture leads to collapse. For a person, it leads to the quieter catastrophe of a life that was lived in earnest but could not see what it was missing — until the missing dimensions returned as crisis, and the crisis, at last, asked the question that the value architecture could not.

Appendix A: Formal Derivation of the Self-Variety Gap (Static)

This appendix formalizes the extension of Ashby's Law of Requisite Variety from physical and institutional controllers to the architecture of the self, yielding the condition **dim(V_self) ≥ dim(D_self) − dim(G_self)** used in the main text. The derivation is static: it treats the disturbance space of a person's life and their personal value architecture as fixed, without modelling their temporal evolution (see Appendix B for the dynamic extension).

A.1 The Self as a Regulated System

Consider a person at a given moment in their life. Their true state can be represented as a vector $\mathbf{x} \in \mathbf{X}$, where **X** is the multidimensional state space of the self — encompassing physical health, emotional condition, relational quality, existential orientation, cognitive capacity, and any other dimensions causally relevant to their wellbeing.

The person is subject to a disturbance vector $\mathbf{d} \in \mathbf{D\_self}$, where **D_self** is the space of disturbances that can push the self away from its desired condition. Disturbances include external events (job loss, illness, relationship conflict, loss of a loved one, financial strain) and internal dynamics (mood fluctuations, intrusive thoughts, fatigue, existential doubt, the slow accumulation of stress).

The person's personal value architecture — the set of values, goals, and metrics they use to evaluate their life — defines a goal set $\mathbf{G\_self} \subset \mathbf{X}$. This is the set of states the person considers acceptable: "I am doing well enough," "I am on track," "My life is satisfactory." The goal set may be narrow (all states where career status exceeds a threshold) or broad (all states where a balance of health, connection, purpose, and security is maintained).

The person does not have direct access to their true state **x**. They perceive themselves through an observation channel:

$$\mathbf{y} = \mathbf{C\_self} \cdot \mathbf{x} + \boldsymbol{\varepsilon}$$

where:

- **C_self**: $\mathbf{X} \to \mathbf{Y}$ is a linear observation matrix determined by the person's value architecture. It selects which dimensions of the self are consciously tracked.
- **ε** is noise — the combined effect of cognitive biases, attentional limits, emotional avoidance, and the inherent noisiness of introspection.

The personal value architecture *is* the matrix **C_self**. What a person values determines which dimensions of their own existence they project into conscious evaluation. The dimensions of **x** that lie in the nullspace of **C_self** are, in the formal sense, unobservable to the person's conscious self-regulation.

A.2 Variety as Dimensionality

Following Ashby (1956), we define variety as the number of distinguishable states a system can occupy or discriminate. For continuous state spaces, we approximate variety by the *effective dimensionality* — the rank of the relevant vector space.

- **dim(D_self)** = rank of the disturbance space: the number of independent ways the person's life can push them away from their acceptable states.
- **dim(G_self)** = rank of the goal set: the number of independent directions in which the person is allowed to vary and still consider themselves "okay." A person with a single, rigid definition of success (e.g., "I must hold a certain title") has **dim(G_self) ≈ 0** — there is no acceptable variation. A person who can be well in multiple ways has a larger **dim(G_self)**.
- **dim(V_self)** = rank of the observation space **Y**, i.e., the number of independent dimensions the person's value architecture can track.

A.3 Ashby's Law in Dimensional Form for the Self

Ashby's Law in its original formulation states that a regulator must have at least as much variety as the disturbance space, minus the variety of the acceptable goal set: **V(R) ≥ V(D) − V(G)** (Ashby, 1956). Mapping variety to dimensionality yields:

dim(V_self) ≥ dim(D_self) − dim(G_self) (1)

This is the static requisite variety condition for a personal value architecture. It states: the number of independent life dimensions a person consciously tracks must be at least the number of independent disturbance dimensions they face, minus the number of independent dimensions in which they can accept variation.

If **dim(V_self) < dim(D_self) − dim(G_self)**, there exist disturbance dimensions that lie outside the person's perceptual field. Those disturbances can push the person out of their goal set without the person ever registering a deviation, because their value architecture lacks the axes along which those disturbances are defined. The person experiences the effects — stress, dissatisfaction, bodily symptoms — but cannot trace them to their source.

A.4 The Self-Variety Gap

The *self-variety gap* is defined as:

G_self = dim(D_self) − dim(G_self) − dim(V_self)

When **G_self ≤ 0**, the value architecture satisfies Ashby's condition. The person has sufficient perceptual dimensionality to detect disturbances in the dimensions that matter, and to distinguish which dimensions of their life require attention.

When **G_self > 0**, the value architecture is underdimensional for the disturbance environment of the person's life. The magnitude of **G_self** measures the volume of causally relevant self-dimensions that are structurally invisible to the person's conscious optimization. These excluded dimensions accumulate damage until they breach thresholds — often manifesting as crises that feel sudden and inexplicable to the person experiencing them.

The critical threshold **G_self_crit** is the value of the gap at which the signal-to-noise ratio in the self-observation channel falls below unity — the point at which the person's conscious model of their own life is more noise than signal. Beyond this threshold, the person cannot recover an accurate picture of their own state, regardless of how much they reflect or how sincerely they try. The condition is structural, not volitional.

A.5 Relationship to Conant and Ashby

Conant and Ashby (1970) proved that "every good regulator of a system must be a model of that system." For the self, the "model" is the personal value architecture — the set of dimensions the person tracks and the relationships they implicitly or explicitly assume among them. If the observation matrix **C_self** is rank-deficient with respect to the disturbance space, the person lacks a complete model of their own life. They are, in those missing dimensions, not a good regulator of themselves — a structural condition that no amount of willpower or positive thinking can remedy.

A.6 Interpretation and Caveats

This derivation provides a conceptual bridge from Ashby's Law to the self-variety gap. It is not an operational measurement protocol. The key limitations are:

1. **Linearity.** Real self-observation is deeply nonlinear. Cognitive biases, emotional states, and narrative structures introduce distortions that a linear matrix cannot capture. The rank condition captures first-order information loss; higher-order dynamics are beyond the present scope.

2. **Dimensionality estimation.** Measuring **dim(D_self)** and **dim(V_self)** in a living person is a significant methodological challenge. Self-complexity instruments (Linville, 1985) provide a starting point but capture cognitive structure rather than full value dimensionality. Experience-sampling methods, narrative analysis, and psychometric tools for value diversity could contribute to operationalization.

3. **Static assumption.** The condition applies to a fixed snapshot of a person's life. It says nothing about how **dim(D_self)** expands across the lifespan or how **dim(V_self)** might be deliberately expanded through practice. The dynamic extension appears in Appendix B.

4. **Goal set interpretation.** A large **dim(G_self)** — a person who can accept many different states as "okay" — reduces the requirement on **dim(V_self)**. This captures the adaptive value of flexibility, non-attachment, and the capacity to find wellbeing across diverse conditions. A person with rigid, narrow definitions of success needs a much richer value architecture to maintain stability, because they have less tolerance for deviation.

Subject to these limitations, equation (1) expresses the architectural insight of the paper in compact form: a personal value architecture that tracks too few dimensions relative to the complexity of a person's life is structurally incapable of maintaining stable self-governance. The self-variety gap quantifies the deficit. When the gap crosses the critical threshold, the person enters a condition of constitutional self-unobservability — they cannot perceive the sources of their own distress, and crisis becomes the only available feedback mechanism for restoring awareness.

Appendix B: Extension to Time-Varying Personal Dimensionality — Dynamics of the Self-Variety Gap and the Personal Dissolution Threshold

Appendix A treated the disturbance space of the self (**D_self**) and the personal value architecture (**V_self**) as static, yielding a snapshot condition: **dim(V_self) ≥ dim(D_self) − dim(G_self)**. But the effective dimensionality of a person's life is not fixed. New relationships, career transitions, illness, aging, parenthood, loss, and the slow accumulation of experience continuously introduce new dimensions of variation that the person must navigate. This appendix extends the static condition to the case where both the disturbance environment and the value architecture can vary over time, formalizing the dynamics of the self-variety gap and the personal dissolution threshold.

B.1 Time-Varying Dimensionality

Let **dim(D_self)(t)** denote the effective dimensionality of the disturbance space of the self at time *t* — the number of independent ways life can push the person away from their acceptable states. Let **dim(V_self)(t)** denote the effective dimensionality of the person's value architecture at time *t* — the number of independent life dimensions they consciously track and integrate. The goal set dimensionality **dim(G_self)** is assumed to vary slowly, if at all; for simplicity, we treat it as a personal constant, though it may shift in moments of profound reorientation.

The self-variety gap at time *t* is:

G_self(t) = dim(D_self)(t) − dim(G_self) − dim(V_self)(t)

The static condition **G_self ≤ 0** is now a moving target. A person whose value architecture was adequate at age twenty-five may find it dangerously narrow at forty, not because their values have changed, but because their life has expanded in complexity and the architecture has not kept pace.

B.2 Dynamics of the Disturbance Space of the Self

The expansion of **dim(D_self)** is driven by the natural unfolding of a human life. New disturbance dimensions emerge through mechanisms including:

- **Developmental transitions**: entering the workforce, forming long-term partnerships, becoming a parent, confronting the needs of aging parents, facing one's own mortality.
- **Bodily change**: the gradual decline of physical capacities, the onset of chronic conditions, hormonal shifts, the accumulating effects of past stress on the body.

- **Relational complexity**: the deepening of existing relationships, the formation of new ones, the loss of others through death or estrangement, the demands of caregiving.
- **Existential unfolding**: the slow emergence of questions about meaning, legacy, identity, and purpose that were not salient in earlier life stages.
- **Environmental and social change**: economic disruption, political upheaval, cultural shifts, technological transformation — all of which introduce new dimensions of uncertainty and demand new adaptive responses.

We model this expansion as:

$$\dim(D_{self})(t) = \dim(D_{self})(0) + \int_0^t \alpha(s)\, ds$$

where $\alpha(s)$ is the instantaneous emergence rate of new disturbance dimensions at time s. In general, $\alpha(s)$ is non-negative and likely non-stationary — periods of major life transition (early adulthood, midlife, late-life integration) produce higher α. The simplest tractable case assumes α is approximately constant over the relevant interval, yielding:

$$\dim(D_{self})(t) = \dim(D_{self})(0) + \alpha\, t$$

B.3 Dynamics of the Personal Value Architecture

A person's value architecture can also expand over time — through conscious practices (self-inquiry, therapy, meditation, honest conversation), through the absorption of lessons from experience, or through the forced expansion that follows a crisis. We model this expansion as:

$$\dim(V_{self})(t) = \dim(V_{self})(0) + \int_0^t \beta(s) \cdot A(V_{self})(s)\, ds$$

where:

- $A(V_{self})(s)$ is the *adaptation effort* at time s — the psychological and practical resources the person devotes to expanding their self-understanding and value horizons.
- $\beta(s)$ is the *adaptation efficiency* — the fraction of adaptation effort that successfully translates into an increase in effective dimensionality. β may be less than one due to psychological defenses, cognitive biases, environmental constraints, or the intrinsic difficulty of perceiving dimensions one's current architecture renders invisible.

Combining these, the dynamics of the self-variety gap are:

$$\frac{dG_{self}}{dt} = \alpha(t) - \beta(t) \cdot A(V_{self})(t) \qquad (2)$$

This is the formal counterpart to the heuristic equation in Part II. The gap grows when the emergence of new life dimensions outpaces the person's expansion of their value architecture. The gap shrinks when personal growth outpaces life's demands for new awareness.

B.4 The Personal Dissolution Threshold

Not all positive values of **G_self** are incapacitating. A person can function with a moderate self-variety gap, absorbing the unobserved variance as diffuse stress, free-floating anxiety, or a nagging sense that "something is off." Catastrophe occurs when the gap exceeds a critical threshold **G_self_crit** at which the signal-to-noise ratio in the self-observation channel falls below unity.

Following the framework established in the governance papers, we define **G_self_crit** as the value of the gap at which:

I(x; y) ≤ I(ε; y)

where **I(x; y)** is the mutual information between the person's true state and their conscious self-perception, and **I(ε; y)** is the mutual information between the noise and the self-perception. Informally: the information the person's conscious awareness carries about their actual life is no greater than the information it carries about the distortion patterns of their own psychology. Beyond this point, introspection — no matter how sincere — cannot recover an accurate picture. The person is governing a phantom.

The precise value of **G_self_crit** depends on the structure of the observation channel and the noise characteristics. For a person with strong cognitive biases (high **ε**), **G_self_crit** is low: even a modest gap can push them into unobservability. For a person with high self-awareness and low bias (low **ε**), **G_self_crit** is higher: they can sustain a larger gap without losing the ability to self-correct.

B.5 Conditions for Managed vs. Unmanaged Gap Growth

Equation (2) yields a direct condition for personal viability:

- **Managed regime**: **β(t) · A(V_self)(t) ≥ α(t)**. The self-variety gap is stable or shrinking. The person maintains perceptual contact with their own life. Growth keeps pace with change.
- **Unmanaged regime**: **β(t) · A(V_self)(t) `<` α(t)**. The self-variety gap grows. The person progressively loses observability of the life dimensions that will eventually determine their wellbeing.

In the unmanaged regime, **G_self(t)** increases monotonically. Unless the regime shifts — either **α** falls (life simplifies) or **β · A(V_self)** rises (adaptation accelerates) — **G_self(t)** will eventually cross **G_self_crit**. The time to personal dissolution is:

T_diss = (G_self_crit − G_self(0)) / (α − β · A(V_self))

This is the time remaining before the person's value architecture becomes structurally incapable of perceiving the sources of their own distress — before crisis becomes the only available corrective.

B.6 Clinical and Developmental Illustrations

This dynamic formalism captures trajectories familiar in clinical practice and developmental psychology:

- **The driven professional** in their thirties: **α** is high (career demands, relationship complexity, physical aging) but **β · A(V_self)** is low (value architecture remains fixated on achievement). **G_self** grows silently. The crisis — burnout, divorce, health collapse — arrives as a shock, but the conditions for it were accumulating in dimensions the person's value architecture could not register.

- **The new parent**: **α** spikes suddenly (the demands of an infant, sleep deprivation, identity reconfiguration) while **β · A(V_self)** may take months or years to catch up. Postpartum distress is, in part, a variety-gap phenomenon: the disturbance space has expanded far faster than the value architecture can track.

- **The person in effective therapy**: **β · A(V_self)** is intentionally increased through the therapeutic relationship, which provides an external observer to help surface excluded dimensions. The gap begins to close. Symptoms that seemed inexplicable become legible as signals from previously unobserved dimensions of the self.

- **The person in sustained contemplative practice**: **β · A(V_self)** is increased through meditation, retreat, and the cultivation of meta-awareness. Over years, **dim(V_self)** may expand to include dimensions — impermanence, interconnectedness, the witness perspective — that were not part of the initial architecture at all. The gap approaches zero asymptotically, though it is never permanently closed.

- **The person in late life**: **α** may increase again as bodily decline, loss of peers, and the approach of death introduce disturbance dimensions that earlier value architectures — built around achievement, provision, or social role — were never designed to handle. If **β · A(V_self)** does not rise to meet this new demand, the gap widens, and the result is the despair, bitterness, or desolation that sometimes marks a difficult old age. If the person can expand their value architecture to include acceptance, legacy, gratitude, and a wider sense of self, the gap can narrow again.

B.7 Caveats and Open Problems

This dynamic extension is a conceptual scaffold, not a calibrated model. Significant limitations include:

1. **Measurement of α and β**: The emergence rate of new life dimensions and the adaptation efficiency of personal growth are not currently measurable with precision. Experience-sampling methods, longitudinal self-complexity assessment, and narrative analysis could contribute to operationalization, but substantial methodological work remains.

2. **Linearity of the dynamics**: Equation (2) is first-order and linear. Real personal development exhibits threshold effects — sudden breakthroughs, nonlinear growth spurts, and the possibility that adaptation becomes easier as meta-awareness increases (β may itself be a function of dim(V_self)). Nonlinear extensions are required for realistic modelling.

3. **Endogeneity of α**: The disturbance emergence rate is not purely exogenous. A person who actively explores their inner life — through therapy, reflection, or contemplative practice — may discover new dimensions earlier, effectively increasing α in the short term while enabling earlier adaptation. The relationship between exploration and disturbance emergence is complex and bidirectional.

4. **Crisis as adaptive mechanism**: The model treats crisis as a consequence of gap crossing. But crisis can also be a mechanism of gap reduction — a forced expansion of dim(V_self) that occurs when gradual adaptation has failed. Incorporating crisis as an endogenous adaptation channel would add realism but considerable complexity to the model.

5. **Goal set evolution**: The model treats **dim(G_self)** as fixed. In practice, people periodically renegotiate what counts as acceptable — through values clarification, spiritual practice, or the forced humility of failure. Incorporating goal set dynamics would add a third differential equation to the system.

Subject to these limitations, Appendix B provides the formal backing for the paper's central dynamic claim: across the lifespan, a static personal value architecture allows the self-variety gap to grow, and when that gap crosses a critical threshold, personal crisis — whether through burnout, breakdown, or existential collapse — becomes structurally inevitable. The only way to avoid this trajectory is to maintain an adaptive capacity that matches the rate at which life generates novelty. The meta-governance practices proposed in Part V are designed to operationalize exactly this capacity.

Appendix C: Psychological Evidence Summary Table

The following table summarizes the key empirical findings referenced in Part IV, mapping each to the variety-gap framework and indicating its implications for the self-variety gap (**G_self**). The table is intended as a reference guide rather than an exhaustive review.

Study / Domain	Key Finding	Mapping to Framework	Implication for G_self
Self-complexity theory (Linville, 1985, 1987)	People with low self-complexity (few, rigid self-aspects) are more vulnerable to stress, depression, and illness following negative events.	Self-complexity approximates **dim(V_self)**. Lower complexity means fewer independent evaluative dimensions.	Higher **G_self** → poorer buffering → greater crisis vulnerability.
Cognitive dissonance (Festinger, 1957)	Individuals distort perceptions and selectively attend to information to maintain self-consistency, often at the expense of accuracy.	Dissonance reduction is a signal-degradation mechanism. The value architecture filters out disconfirming information to preserve its own structure.	**G_self** grows silently as excluded dimensions (truth, accurate self-perception) accumulate distortion.
Default Mode Network and meditation (Brewer et al., 2011)	Experienced meditators show reduced DMN activity — the brain network associated with narrative self-construction and mind-wandering.	Reduced DMN activity corresponds to decreased narrative compression and aggregation of experience. The observation channel widens.	**G_self** temporarily decreases during practice; repeated practice may produce lasting expansion of **dim(V_self)**.
Meditation and DMN beyond active tasks (Garrison et al., 2015)	DMN reductions in meditators persist outside formal practice, suggesting trait-level changes in self-perception.	Sustained contemplative practice may produce enduring changes in the observation architecture.	Long-term reduction in baseline **G_self**; increased capacity to perceive subtle self-signals before they become crises.
Flow states (Csikszentmihalyi, 1990)	During flow, self-consciousness temporarily dissolves, the boundary between self and activity blurs, and experience becomes intrinsically rewarding.	Flow represents a temporary suspension of the habitual value architecture — the ego's usual optimization targets are offline.	Momentary approach toward **G_self → 0**; the person perceives a wider field of experience than the ego's filters normally permit.
Peak experiences (Maslow, 1964)	Transcendent experiences involve a temporary loss of self, a sense of unity, and heightened clarity, often leading to lasting shifts in values.	Peak experiences can be understood as temporary gap collapse followed by integration of newly perceived dimensions into **V_self**.	Post-peak integration may permanently increase **dim(V_self)**; a single experience can expand the value architecture.

Study / Domain	Key Finding	Mapping to Framework	Implication for G_self
Ego development (Cook-Greuter, 2013)	Across developmental stages, the self's capacity to hold complexity, paradox, and multiple perspectives progressively expands from impulsive to integrated stages.	Ego development tracks the long-term trajectory of **dim(V_self)** expansion across the lifespan.	Higher stages correspond to lower baseline **G_self** and greater capacity for meta-governance.
Self-transcendence (Maslow, 1969)	Beyond self-actualization lies a stage where values shift from deficiency-motivated to being-motivated, including connection to something beyond the individual self.	Self-transcendence represents the deliberate expansion of **dim(V_self)** to include collective, ecological, and spiritual dimensions.	**G_self** decreases as the boundary between self and world becomes more permeable; fewer dimensions are excluded from the value architecture.
Acceptance and Commitment Therapy (Hayes et al., 1999)	ACT distinguishes self-as-content (the narrative self) from self-as-context (the observing awareness), and uses this distinction to increase psychological flexibility.	ACT operationalizes second-order cybernetics at the personal scale: it trains the capacity to observe one's own observation architecture.	Increased meta-observational capacity reduces **β**-loss (adaptation efficiency improves); the person can detect and respond to gap growth earlier.

C.1 Pattern Summary

Across these diverse research domains, a consistent pattern emerges:

- **Narrow, rigid, low-dimensional personal value architectures** — operationalized as low self-complexity, high cognitive bias, rigid self-narratives, or fixated optimization targets — are associated with poorer self-observability, greater vulnerability to unperceived deterioration, and more severe crises when excluded dimensions eventually breach awareness.
- **Practices and traits that expand the effective dimensionality of self-perception** — differentiated self-aspects, tolerance for cognitive dissonance rather than suppression of it, mindfulness, flow, peak experiences, ego development, and ACT-style meta-awareness — are associated with greater resilience, earlier detection of emerging problems, richer experience of life, and a greater capacity to navigate crisis when it does arrive.

This pattern is consistent with the variety-gap framework's core prediction: that **dim(V_self)** is a measurable psychological variable with predictable consequences for stress vulnerability, crisis resilience, and the capacity for personal growth.

C.2 Limitations

The studies summarized here were not designed to test the variety-gap model directly. The mappings to **dim(V_self)** and **G_self** are interpretive, not deductive. Empirical validation would require:

1. Operationalizing **dim(V_self)** as a measurable construct (building on self-complexity instruments and value-diversity assessments).
2. Developing measures of **G_self** that can be tracked longitudinally.
3. Testing whether changes in **G_self** predict crisis onset, recovery trajectories, and the effectiveness of meta-governance practices.

The table serves as a plausibility demonstration and a generator of testable hypotheses, not as a confirmation of the framework.

Appendix D: Simulation Architecture — The Self Stability Simulator

This appendix outlines a simulation model that makes the self-variety gap and the Goodhart–Ashby synthesis for the self concretely visible. The model extends the logic of the Governance Stability Simulator (Paper I) and the value-function collapse demonstrator (Appendix C of *The Variety Gap*) to the domain of an individual life.

D.1 System Description

The simulation models a person as a dynamical system with **N** internal state dimensions, each representing a causally relevant aspect of wellbeing:

- **H(t)** : physical health (energy, immune function, absence of illness)
- **R(t)** : relational integrity (depth and authenticity of close relationships)
- **M(t)** : existential meaning (sense of purpose, coherence, direction)
- **C(t)** : career / contribution (productive engagement, recognition, growth)
- **L(t)** : leisure / restoration (rest, play, aesthetic experience)

Additional dimensions can be added or removed. The dimensions are coupled: neglecting health reduces career performance over time; neglecting relationships reduces meaning; and so forth. The true state of the self at time *t* is the vector **x(t)** = $[H(t), R(t), M(t), C(t), L(t)]^T$.

The system evolves according to:

$$\mathbf{x(t+1) = A \cdot x(t) + B \cdot u(t) + d(t) + drift}$$

where:

- **A** captures natural decay and cross-dimensional coupling (e.g., health decays slowly without attention; meaning decays if relationships are poor).
- **B** translates investments of time, energy, and attention (**u**) into changes in each dimension.
- **d(t)** are disturbances — external shocks (illness, loss, conflict) and internal fluctuations (mood, energy).
- **drift** maintains equilibrium in the absence of disturbance and control.

D.2 Value Architectures (Controllers)

The simulation compares two personal value architectures that differ only in their **dimensionality**, not in their effort or competence.

Architecture 1D (Single-Metric)

- The person's value function tracks only one dimension — for example, career success: $\mathbf{J_1 = C(t)}$.
- The observation channel **C_self** projects the full self-state onto the career axis; all other dimensions are unobserved. The person attends only to signals about performance, recognition, and advancement.
- Control law: **u(t) = K · (C_target − C_obs(t))**, where **K** is a gain and **C_obs** is the noisy observation of career. The person invests effort whenever career performance is below the target, with no awareness of the costs to other dimensions.

Architecture ND (Multi-Dimensional)

- The person's value function tracks *n* dimensions simultaneously, where *n* is adjustable by the user (e.g., *n* = 3: career, health, relationships).
- The observation channel includes those dimensions, with some noise. The person attends to a richer set of signals.
- Control law: $\mathbf{u_i(t) = K_i \cdot (x_i\_target - x_i\_obs(t))}$ for each tracked dimension *i*. Effort is allocated to close gaps in any dimension that is perceived as deficient.

Both controllers have access to the same total effort budget per time step; the only difference is the dimensionality of their value architecture and the resulting allocation of attention and action.

D.3 Coupling and the Collapse Mechanism

The critical structural feature is **cross-dimensional coupling**. Health affects career productivity; relationships affect meaning; meaning affects motivation to maintain health. The simulation implements a simple linear coupling:

- Career output at time *t* depends on **C(t) · f(H(t))**, where *f* is a monotonically increasing function that reduces productivity when health is low.
- Meaning depends on **M(t) · g(R(t))**, where *g* captures the contribution of relational depth to a sense of purpose.

The 1D controller, blind to health and relationships, initially succeeds: career investment yields career returns. But as health and relationships silently decline (because no effort is allocated to them), the coupling terms begin to degrade career performance. The controller responds by *increasing* career investment — the only lever it knows — which further accelerates the decline of the excluded dimensions. This is the Goodhart–Ashby mechanism: the narrow metric destroys the conditions on which its own success depends.

Eventually, career performance collapses despite maximum effort, because the underlying health and relationship dimensions have fallen below critical thresholds. The collapse appears sudden and inexplicable to the 1D controller, which has no categories in which to perceive the causes.

The ND controller, observing health and relationships alongside career, moderates its investment before the excluded dimensions reach dangerous levels. Career performance grows more slowly but does not collapse. The system reaches a stable, multidimensional equilibrium.

D.4 User-Adjustable Parameters

The simulation is designed as an interactive tool for exploring the self-variety gap. Key parameters the user can adjust include:

- **Number of tracked dimensions (n)** : from 1 (narrow) to N (full).
- **Gain (K)** : the intensity of corrective effort. High gain with low dimensionality accelerates collapse.
- **Coupling strength** : how strongly excluded dimensions feed back on the tracked metric. Stronger coupling produces faster collapse for narrow architectures.
- **Disturbance level** : the frequency and magnitude of external shocks. Higher disturbance levels require higher dimensionality to maintain stability.
- **Initial conditions** : starting values for each dimension, allowing exploration of different life configurations.

D.5 Expected Outputs

The simulation produces time-series plots showing:

1. **Trajectories of all N dimensions** under each architecture, with the collapse of untracked dimensions and the eventual collapse of the tracked metric in the 1D case clearly visible.
2. **The self-variety gap G_self over time**, computed as the difference between the effective dimensionality of the disturbance environment and the dimensionality of the active value architecture.
3. **A “crisis detector”** that flags when any dimension falls below a critical threshold.
4. **A summary comparison** of mean wellbeing, crisis frequency, and total effort expended under each architecture.

D.6 Relevance to the Variety Gap

This simulation is a direct instantiation of the self-variety gap model:

- The 1D controller has **dim(V_self) = 1**; the disturbance space has **dim(D_self) = N** (plus couplings). **G_self** is large.
- The excluded dimensions (health, relationships, meaning) are causally coupled to the tracked metric (career). Optimizing the proxy without monitoring the excluded dimensions eventually destroys the correlation.

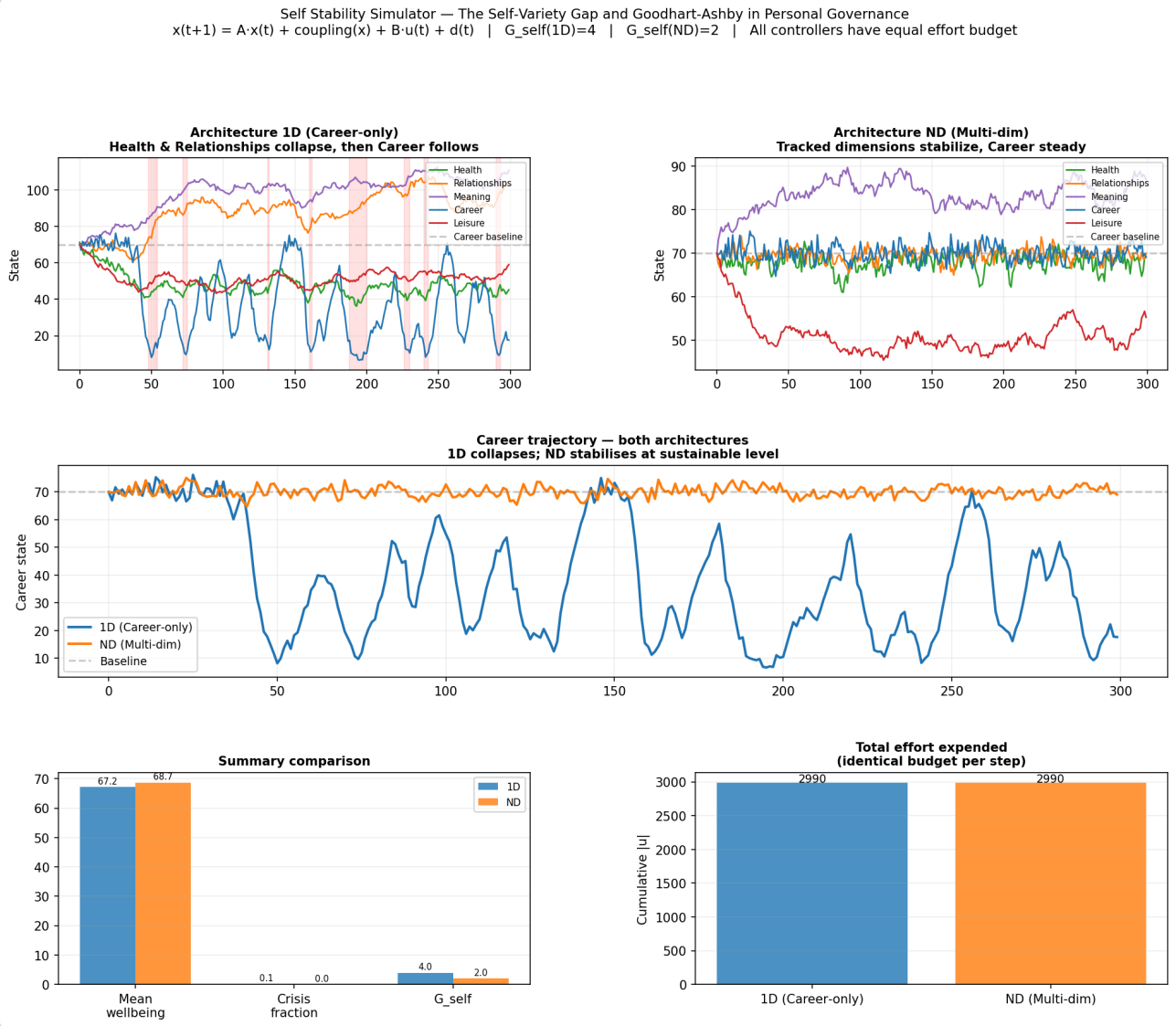
- The collapse follows the Goodhart–Ashby dynamics: the narrow metric optimizes away its own informational basis.
- The ND controller, with a higher **dim(V_self)**, manages **G_self** and avoids collapse.

The simulation makes the abstract mechanics of the self-variety gap tangible. It allows users to see, in simplified form, how a life governed by a single metric inevitably destabilizes — and how expanding the dimensionality of what one values can restore stability.

D.7 Reproducibility

A reference implementation of the Self Stability Simulator is available in the companion repository (see “Simulation Code” in the supplementary materials). The simulation uses standard Python with NumPy and Matplotlib; no specialized packages are required. Users can run the simulation, adjust parameters through a simple interface, and generate the figures described above.

Figure D.1: Self-variety gap collapse in a five-dimensional personal system



The 1D controller (tracking only career) initially applies maximum effort to career advancement. This effort directly drains health through overwork (physiological cost of sustained high performance). As health falls below 70% of baseline, career productivity enters exponential decline—the same effort yields progressively less output. The controller, unable to perceive health, responds by increasing effort further, accelerating health depletion. This creates recurring boom-bust cycles visible in the career trajectory (middle panel): brief recoveries when accumulated rest allows temporary health restoration, followed by deeper crashes as the controller immediately re-depletes the recovered capacity.

Health stabilizes near collapse (~15-20), relationships decay to ~45, and career oscillates between 10 and 70 despite continuous maximum effort. The system enters a chronic crisis state with crisis fraction = 0.1. The person is working as hard as possible but achieving progressively worse outcomes—the canonical burnout pattern.

The ND controller (tracking health, relationships, career, and leisure) distributes effort across dimensions. By maintaining health above the 70% threshold, it avoids triggering the career penalty entirely. All tracked dimensions stabilize near baseline with low variance. Career performance is steady at ~70—lower than 1D's peaks but without the catastrophic troughs. Crisis fraction = 0.0.

The summary comparison (bottom left) shows nearly identical mean wellbeing (67.2 vs 68.7) because 1D's brief peaks compensate for its deep troughs when averaged. But the *lived experience* differs radically: ND provides stable, predictable functioning; 1D produces chronic instability and recurring crises. The identical effort budgets (bottom right) demonstrate this is architectural failure, not motivational deficit.

This is the personal Goodhart-Ashby synthesis: a value architecture with dim(V_self)=1 cannot maintain stability when the excluded dimension (health) is causally coupled to the tracked metric (career). The narrow optimization destroys its own informational and physiological basis.

Appendix E: Annotated Reference List

This appendix provides full citations for all works referenced throughout the paper. Brief annotations are included for sources that ground core claims, to assist the reader in locating the most directly relevant material.

Entries marked with † are discussed in the evidence summary table in Appendix C.

Foundational Cybernetics and Control Theory

Ashby, W. R. (1956). ***An Introduction to Cybernetics*****. Chapman & Hall.**
The canonical statement of the Law of Requisite Variety. Foundational for the argument that a personal value architecture must match the variety of the self it governs (Parts I, II). **[R]**

Conant, R. C. & Ashby, W. R. (1970). "Every Good Regulator of a System Must Be a Model of That System." ***International Journal of Systems Science*****, 1(2), 89–97.**
Formal proof that effective regulation requires internal modelling capacity. Underpins the claim that personal values function as a model of the self (Part II). **[R]**

von Foerster, H. (1984). ***Observing Systems*****. Intersystems Publications.**
Foundational text of second-order cybernetics: the observer observing their own observing. Directly supports the meta-governance argument in Part V. **[R/I]**

Self-Regulation and Psychological Control Theory

Carver, C. S. & Scheier, M. F. (1982). "Control Theory: A Useful Conceptual Framework for Personality–Social, Clinical, and Health Psychology." ***Psychological Bulletin*****, 92(1), 111–135.**
Establishes the self as a feedback control system in mainstream psychology. Provides the bridge from cybernetics to the personal scale (Part I). **[R/I]**

Powers, W. T. (1973). ***Behavior: The Control of Perception*****. Aldine.**
Develops Perceptual Control Theory, a rigorous application of control theory to individual behavior and perception. Supports the modeling of personal values as observation channels (Part I). **[R]**

Goodhart’s Law and the Goodhart–Ashby Synthesis

Goodhart, C. A. E. (1975). “Problems of Monetary Management: The U.K. Experience.” In *Papers in Monetary Economics*, Reserve Bank of Australia.
The original statement of Goodhart’s Law. Extended here to personal metrics (Part II). **[R/I]**

Manheim, D. & Garrabrant, S. (2018). “Categorizing Variants of Goodhart’s Law.” *arXiv preprint arXiv:1803.04585*.
Formal taxonomy of Goodhart effects, including regime drift. Supports the generalization from gaming to structural blindness in personal optimization (Part II). **[R]**

Müller, J. Z. (2018). *The Tyranny of Metrics*. Princeton University Press.
Historical analysis of how metric fixation distorts behavior across domains. Provides empirical analogues for the personal metric-fixation argument (Part II). **[I]**

Self-Complexity, Cognitive Dissonance, and Self-Perception

†**Linville, P. W. (1985). “Self-Complexity and Affective Extremity: Don’t Put All of Your Eggs in One Cognitive Basket.” *Social Cognition*, 3(1), 94–120.**
Empirical demonstration that lower self-complexity predicts greater vulnerability to stress. Interpreted in the framework as an indicator of low dim(V_self) (Part IV). **[I]**

† **Linville, P. W. (1987). “Self-Complexity as a Cognitive Buffer Against Stress-Related Illness and Depression.” *Journal of Personality and Social Psychology*, 52(4), 663–676.**
Further evidence linking self-complexity to psychological resilience. Supports the dimensionality-as-buffer hypothesis (Part IV). **[I]**

†**Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.**
The foundational work on how individuals distort perception to maintain self-consistency. Interpreted here as a signal-degradation mechanism (Part IV). **[I]**

Meditation, Mindfulness, and Flow

†**Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y.-Y., Weber, J., & Kober, H. (2011). “Meditation Experience Is Associated with Differences in Default Mode Network Activity and Connectivity.” *Proceedings of the National Academy of Sciences*, 108(50), 20254–20259.**

Demonstrates that experienced meditators show reduced DMN activity, interpreted here as reduced narrative compression and expanded observability (Part IV). **[I]**

†**Garrison, K. A., Zeffiro, T. A., Scheinost, D., Constable, R. T., & Brewer, J. A. (2015). "Meditation Leads to Reduced Default Mode Network Activity Beyond an Active Task."** ***Cognitive, Affective, & Behavioral Neuroscience*****, 15, 712–720.**
Shows that DMN reductions persist beyond formal practice, suggesting lasting changes in the observation architecture (Part IV). **[I]**

†**Csikszentmihalyi, M. (1990).** ***Flow: The Psychology of Optimal Experience*****. Harper & Row.**
Describes flow states where self-consciousness temporarily dissolves. Interpreted as moments of temporary gap collapse (Part IV). **[I]**

Humanistic and Developmental Psychology

†**Maslow, A. H. (1964).** ***Religions, Values, and Peak-Experiences*****. Ohio State University Press.**
Examines transcendent experiences and their role in mental health. Supports the discussion of value expansion beyond the ego (Part IV, Part VI). **[I/S]**

†**Maslow, A. H. (1969). "The Farther Reaches of Human Nature."** ***Journal of Transpersonal Psychology*****, 1(1), 1–9.**
Articulates the concept of self-transcendence as a developmental stage. Interpreted here as movement toward the asymptotic limit $G_{self} \to 0$ (Part VI). **[S]**

†**Cook-Greuter, S. (2013). "Nine Levels of Increasing Embrace in Ego Development: A Full-Spectrum Theory of Vertical Growth and Meaning Making."** ***Integral Leadership Review*****, 13(1).**
Describes the progressive expansion of what the self can hold in awareness across developmental stages. Supports the trajectory toward higher $\dim(V_{self})$ (Part VI). **[I/S]**

Therapeutic Approaches and Meta-Governance

† **Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999).** ***Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*****. Guilford Press.**
Develops a therapeutic approach that explicitly distinguishes self-as-content from self-as-context. Supports the meta-observational mechanisms in Part V. **[I]**

Governance as Engineering Series

Holmström, B. K. (2026). *The Variety Gap: What We Don't Optimize For, We Lose the Ability to See*. GGF Whitepaper Series, Paper VI.
The companion paper that establishes the civilizational-scale framework from which the self-governance model is derived. Referenced throughout. **[R]**

Holmström, B. K. (2026). *Governance as Engineering Series*, Papers I–V. GGF Whitepaper Series.
The foundational series establishing the control-theoretic analysis of governance that is extended here to the individual scale (Part I). **[R]**

Note: The tags [R], [I], and [S] indicate whether each reference primarily supports the rigorous core, the interpretive layer, or the speculative horizon of the paper. This categorization is intended to help readers assess the level of epistemic commitment appropriate to each citation.