

THE VARIETIES OF HUMAN EXCELLENCE

A History of Mind, a Map of Its Dimensions, and a Framework to Replace Intelligence

Working Paper — 2026

How This Book Works

This book makes a single argument in three movements. Each movement earns the next.

Movement I (History, Chapters 1-19) reconstructs what happened. Across nineteen chapters and more than a dozen civilizational traditions, it shows that humanity has never had one concept of "intelligence." The Greek tradition distinguished five intellectual virtues. The Confucian tradition built a thirteen-century examination system around textual mastery and moral character, treating these as inseparable. Indian logicians developed a five-step syllogism using a fundamentally different architecture from Aristotle's. Buddhist philosophy insisted that wisdom without compassion is not yet wisdom. The Yoruba tradition distinguished the wisdom of the head from the wisdom of the belly and made character the paramount quality. The Ubuntu tradition challenged whether the individual extracted from relational context is a meaningful unit of analysis. Pacific navigators held entire star compasses in memory during weeks of open-ocean voyaging. Five civilizations independently developed ideals of the excellent person that integrate cognition with moral character. Three independently developed formal logical systems. Eight or more refused to separate intellectual excellence from ethical formation. Then, beginning in early modern Europe and accelerating through Galton, Binet, and Spearman, this rich plurality was compressed into a single measurable quantity called "intelligence." The narrowing was powerful and consequential. It was also a choice, not a discovery.

Movement II (The Map, Chapters 20-32) asks: if the traditions were tracking real phenomena under different names, what are those phenomena, and how well can modern science measure them? Twelve chapters map twelve dimensions of human excellence (reasoning, learning capacity, knowledge, practical judgment, emotional attunement, social-relational competence, self-regulation, creativity, wisdom, moral character, embodied skill, and collective capability) across traditions and modern measurement. A thirteenth chapter reconstructs the full spectrum of assessment methods that humanity has used, from formal examination to debate to apprenticeship to demonstrated performance. The pattern that emerges is clear: the dimensions the psychometric tradition measures best (reasoning, knowledge) are the ones most traditions treated as one component among many. The dimensions most traditions placed at the center of excellence (practical judgment, wisdom, moral character) are the ones modern psychology measures least well or not at all.

Movement III (The Replacement, Chapters 33-41) draws the conclusion and proposes an alternative. Chapter 33 argues that "intelligence" is not a single natural kind but a partly real, badly bundled construct, using evidence from within psychometrics (the mutualism model's alternative explanation of the positive manifold), from across civilizations (convergent phenomena, divergent bundling), and

from the measurement fragmentation documented in Movement II. Chapter 34 proposes a twelve-dimensional framework for human excellence: not a new single score but a set of partly independent axes that different traditions have tracked, different measurement systems assess, and different contexts weight differently. The remaining chapters apply the framework to the domains where it matters most: assessment across cultures, human-AI comparison, education, institutional selection, and AI alignment.

The arc: **History** → **Map** → **Replacement**.

The book concedes what IQ gets right (cognitive abilities are real, measurable, and predictive) before arguing that it is incomplete (the incompleteness is structural, not marginal, and it excludes the dimensions most human traditions placed at the center of their accounts of excellence). It is not an anti-IQ polemic. It is a proposal to replace a false unity with a framework adequate to what humanity has actually tracked.

MOVEMENT I: THE HISTORY

What happened: humanity's long, branching attempt to define the best kind of mind

Chapter 1: The Problem of the Word

Ask fourteen psychologists to define intelligence and you will get fourteen different answers. This is not a joke. It is what actually happened.

In 1921, the editors of the *Journal of Educational Psychology* invited fourteen leading researchers to answer a deceptively simple question: What is intelligence, and how should it be measured? The answers were so varied that the symposium became famous not for settling the question but for proving it unsettled. Edward Thorndike defined intelligence as the power of good responses from the point of view of truth or fact. Lewis Terman called it the ability to carry on abstract thinking. S.S. Colvin described it as the ability to learn to adjust oneself to the environment. V.A.C. Henmon called it the capacity for knowledge. H. Woodrow called it the capacity to acquire capacity. Each expert was answering the same question. None gave the same answer.

Sixty-five years later, in 1986, Robert Sternberg and Douglas Detterman repeated the experiment. They invited two dozen experts from a new generation to respond to the same prompt. The result was the same: roughly as many definitions as definers. The APA's 1996 task force report, *Intelligence: Knowns and Unknowns*, acknowledged this directly. Individual differences in cognitive ability are real, substantial, and consequential. But no single conceptualization of intelligence has yet answered all the important questions, and none commands universal assent.

This book begins with a claim that may sound radical but is, on closer inspection, simply honest: **the word "intelligence" does not name a single thing.** It is a modern English-language compression of a much larger, older, and more plural human project. That project, pursued across civilizations and centuries, has been concerned not just with reasoning or learning speed but with judgment, character, emotional attunement, social competence, practical mastery, ecological skill, and wise conduct. Reducing all of that to a single word, and then building schools, hiring systems, AI benchmarks, and self-concepts around that word, has consequences. This book is about what those consequences are and what a better framework would look like.

The word is younger than it looks

Most people treat "intelligence" as though it names something obvious, ancient, and universal. It does none of these things.

The English word *intelligence* entered the language in the late fourteenth century, borrowed from Old French, which took it from Latin *intelligentia*. The Latin root is *intelligere*, a compound of *inter* (between) and *legere* (to choose, pick out, or read). The original sense was closer to "discernment"

or "the capacity to read between": to perceive what is not immediately obvious, to pick out what matters. Latin *intelligentia* covered a broad semantic field: understanding, knowledge, the power of discerning, and even art, skill, and taste.

In medieval European thought, the related term *intellectus* became the standard scholarly word for the faculty of understanding. It was used to translate the Greek *nous*, which Aristotle had treated as the highest part of the soul, the capacity for grasping first principles directly rather than through argument. But *intellectus* in this period was embedded in a thick metaphysical context. It was bound up with theories of the soul, the active intellect, the angels (sometimes called "intelligences" themselves), and the broader scholastic picture of a rationally ordered cosmos.

When early modern philosophers broke with scholasticism, many of them deliberately avoided the word. Francis Bacon, Thomas Hobbes, John Locke, and David Hume preferred "understanding" in their English philosophical works. Hobbes even mocked the scholastic usage. For these thinkers, *understanding* was a human cognitive capacity stripped of its medieval metaphysical baggage.

The word "intelligence" only returned to English prominence in the early 1900s, and when it did, it came back in a very specific context: the new science of mental testing. The phrase "intelligence quotient" first appeared in print in 1912, coined by the German psychologist William Stern. Alfred Binet had developed the first practical intelligence test in 1905, not to measure some deep metaphysical faculty but to identify Parisian schoolchildren who needed extra help. The word "intelligence" was repurposed for this new, practical, administrative project. And because the tests worked well enough for institutional purposes, the word stuck.

This matters because the word now carries an enormous weight of assumption that its history does not support. When someone says "she's intelligent," they are typically using a word that entered its current sense barely a century ago, that was repurposed from a medieval Latin term with completely different connotations, that was deliberately avoided by the founders of modern philosophy, and that was revived specifically to label what a particular kind of test measures. None of this means the phenomena are unreal. People obviously differ in their ability to reason, learn, adapt, and solve problems. But the assumption that these differences are best captured by a single word pointing at a single underlying thing is not a discovery. It is a choice, and a relatively recent one.

Fourteen definitions, one problem

Return to the 1921 symposium for a moment, because the disagreement is more instructive than it first appears.

The definitions offered by those fourteen experts were not just differently worded versions of the same

idea. They pointed at genuinely different things. Thorndike emphasized correct responses. Terman emphasized abstract thinking. Colvin emphasized environmental adjustment. Woodrow emphasized the capacity to acquire new capacities. These are not synonyms. A person who excels at abstract reasoning may not be especially good at adjusting to new environments. Someone who acquires new skills rapidly may not always produce correct responses. The experts were not describing the same elephant from different angles. They may have been describing different animals.

The 1986 reprise did not resolve this. Sternberg himself proposed a triarchic theory distinguishing analytical, creative, and practical intelligence. Howard Gardner had published *Frames of Mind* in 1983, arguing for at least seven distinct intelligences. The psychometric mainstream continued to defend a general factor, *g*, as the best summary of the positive correlations among cognitive tests. These are not minor variations on a theme. They are fundamentally different claims about what the word picks out: a single general capacity, a small number of broad capacities, or a large set of partly independent abilities.

What is striking is not that experts disagree. Experts disagree about many things. What is striking is that a century of intensive research has not produced convergence. The 1996 APA task force essentially said as much. This is unusual in science. Disagreement about the definition of a key term that persists for over a hundred years, across thousands of studies, is not a sign of a field that has not worked hard enough. It is a sign that the term itself may be doing something misleading.

Consider an analogy from a different domain. Suppose a group of biologists were asked to define "life." Some would emphasize metabolism. Some would emphasize reproduction. Some would emphasize homeostasis. Some would emphasize information processing. The disagreement would be real, but it would not mean that biologists cannot study living systems. It would mean that "life" is a folk category that bundles together several partly independent processes, and that serious scientific work requires specifying which process you are actually studying. That is roughly the situation with "intelligence." The word is useful in everyday speech. It is misleading as a scientific master category because it bundles together capacities that can be studied, measured, and cultivated separately.

What the word hides

The deepest problem with "intelligence" as a master category is not that people define it differently. It is that the word actively hides the plurality of what humans have cared about when they asked, "What makes a mind good?"

Consider just the Western tradition for a moment. Aristotle did not have a single word for "intelligence." He had a carefully differentiated vocabulary: *nous* (intuitive reason, the direct grasp of first

principles), *epistēmē* (scientific understanding, knowledge of what is necessary and demonstrable), *technē* (craft knowledge, knowing how to make or produce), *sophia* (theoretical wisdom, the union of nous and *epistēmē* directed at the highest objects), and *phronēsis* (practical wisdom, the capacity for good deliberation about how to live well in particular circumstances). These are not five names for the same thing. They pick out genuinely different capacities, and Aristotle was clear that excelling at one does not guarantee excelling at another. A brilliant theoretician may lack practical wisdom. A skilled craftsman may have no talent for abstract philosophy. A person of exceptional practical judgment may not be able to prove a theorem.

This is not an ancient Greek curiosity. The same pattern repeats across civilizations. Classical Chinese philosophy worked with the concept of *xin* (心), often translated as "heart-mind," a term that deliberately refuses to separate thinking from feeling, understanding from intention, cognition from desire. What counts as excellence in the Confucian tradition is not raw processing speed but cultivated attunement: knowing how to feel, judge, and act appropriately in one's relationships and roles. The imperial examination system, the largest pre-modern meritocratic institution in human history, tested literary mastery of the Confucian classics as a proxy for the kind of moral-intellectual formation thought necessary for governance. It was not an IQ test.

Classical Indian philosophy took yet another approach. Much of the Nyāya and related traditions asked not "how intelligent is this person?" but "what makes a cognition valid?" The emphasis was on epistemology: the sources (*pramāṇas*) through which knowledge arises, including perception, inference, testimony, and in some schools analogy and other methods. The question was less about ranking minds and more about understanding what makes any particular act of knowing trustworthy. Buddhist traditions, especially in their Tibetan scholastic forms, paired *prajñā* (wisdom, often understood as insight into the nature of reality) with *karuṇā* (compassion). Wisdom without compassion was not merely incomplete; in many formulations it was not yet wisdom at all. The Geshe degree in Tibetan monasteries is earned through years of study, memorization, and public debate. It tests reasoning with great rigor, but it is embedded in a system that treats intellectual development as inseparable from ethical and contemplative transformation.

West African traditions, in Akan and Yoruba thought, defined wisdom not as abstract reasoning but as the ability to make good use of what one knows, demonstrated through speech, judgment, conduct, and the maintenance of social harmony. The Akan concept of *nyansa* places the emphasis squarely on practical application and communal benefit. Yoruba thought sometimes distinguishes intellectual wisdom (*ogbón-ori*) from emotional or interior wisdom (*ogbón-inú*), a distinction that anticipates modern debates about emotional intelligence by centuries.

Southern African Ubuntu traditions relocate human excellence entirely. Here, the central question

is not "how smart is this individual?" but "how fully has this person realized their humanity through their relationships with others?" Excellence is relational, not computational.

Pacific wayfinding traditions demonstrate a form of intelligence that modern testing largely ignores. Master navigators hold in memory vast bodies of knowledge about stars, wave patterns, currents, winds, bird behavior, and cloud formations, integrating them in real time to cross thousands of miles of open ocean without instruments. This is not metaphorical intelligence. It is a genuine cognitive achievement of extraordinary sophistication. But it looks nothing like an IQ test, and no standardized battery would capture it.

Every one of these traditions was trying to answer some version of the question: What does it mean for a human being to think well, judge well, know well, and live well? And every one of them arrived at answers that are broader, more plural, and more entangled with ethics, emotion, and social life than the modern English word "intelligence" suggests. The word hides this plurality. It makes it look as though all of human history has been circling around one thing, and that modern psychology finally figured out how to measure it. That is not what happened.

The institutional lock-in

If the word is so misleading, why does it persist? Because it is useful to institutions.

The rise of "intelligence" as a dominant category in the twentieth century was driven less by philosophical clarity than by administrative need. States needed to sort millions of schoolchildren into tracks. Armies needed to assign millions of recruits to roles. Immigration authorities wanted quick assessments of incoming populations. Employers wanted efficient screening tools. All of these institutions needed something that was fast, cheap, scalable, and defensibly objective. IQ-style testing provided exactly that.

This is not a conspiracy theory. It is a straightforward historical observation. Binet designed his test to help schools. The U.S. Army Alpha and Beta tests during World War I demonstrated that group intelligence testing could process enormous numbers of people rapidly. By the mid-twentieth century, standardized cognitive testing had become a cornerstone of educational sorting, military placement, clinical diagnosis, and employment screening across the industrialized world.

The tests worked, in the sense that scores predicted outcomes that institutions cared about: school grades, job performance ratings, training success. This predictive validity is real and should not be dismissed. But it is worth noting what the tests predicted *for*. They predicted success in environments that were themselves designed around a particular kind of cognitive performance: formal schooling, bureaucratic organizations, and structured task completion. A tool that predicts performance within

a system is not necessarily measuring the deepest truth about human capability. It may be measuring fit with the system.

The result was a kind of institutional lock-in. The tests existed. The institutions used them. The scores predicted things the institutions valued. This created a feedback loop in which "intelligence" came to mean, operationally, whatever the tests measured. When critics pointed out that the tests left things out, the response was often to develop *additional* tests: emotional intelligence tests, creativity tests, executive function batteries, situational judgment tests, wisdom scales. Each of these is, in effect, an attempt to recover something that was amputated when "intelligence" was narrowed to fit the needs of mass institutional sorting. The fact that so many recovery efforts have been necessary is itself evidence that the original narrowing left too much behind.

What this book argues

This book makes three connected claims.

First, the historical claim. Humanity has never had a single, stable concept of intelligence. What it has had, across civilizations and centuries, is a recurring set of questions about what makes a mind good: questions about reasoning, learning, judgment, emotional attunement, social competence, self-regulation, creativity, practical mastery, ecological skill, moral reliability, and wise conduct. Different traditions bundled these capacities differently, valued them differently, and recognized them through different methods: examinations, debate, teacher authorization, apprenticeship, demonstrated performance, proverb competence, elder judgment, and eventually standardized testing. The modern psychometric concept of intelligence is one late, powerful, and institutionally successful entry in this much longer conversation. It is not its culmination.

Second, the structural claim. "Intelligence" is not a single natural kind. It is a partly real but badly bundled construct. The phenomena it points at are genuine: people really do differ in reasoning ability, learning speed, and problem-solving capacity. But bundling these with each other while excluding judgment, emotion, character, relational skill, and ecological attunement is not a neutral scientific decision. It is a compression that was useful for institutional purposes but that distorts the actual landscape of human capability. A better framework would decompose "intelligence" into its constituent dimensions and treat them as a profile rather than collapsing them into a single score.

Third, the practical claim. The consequences of the current framing are real and widespread. Education systems that sort children by a narrow band of cognitive performance miss other capacities that matter for flourishing. Hiring systems built around IQ-adjacent screening overlook practical judgment, social skill, and moral reliability. AI benchmarking that asks "how smart is this model?"

recapitulates the same error, collapsing many distinct capabilities into one vague superlative. And public discourse that treats "intelligence" as a single axis along which individuals and groups can be ranked obscures more than it reveals. A plural framework would not solve all of these problems, but it would at least stop making them worse.

The shape of the book

The book proceeds in three movements.

Movement I is historical. It traces what happened: how different civilizations defined, cultivated, and recognized human excellence. It covers ancient Greece, Rome, classical China, classical India, Buddhist traditions, Islamic and Persianate traditions, West African philosophical traditions, Southern African relational traditions, Māori, Inuit, Pacific, and other Indigenous knowledge systems, and the long arc of Western thought from the medieval period through the Enlightenment to the birth of modern psychometrics. The purpose is not to produce an encyclopedia of ideas about intelligence. It is to show that the modern concept is a late narrowing of a much larger project, and that the traditions it displaced or ignored were often tracking real capacities that modern testing leaves out.

Movement II is cartographic. It maps the recurring dimensions of human excellence that civilizations have been tracking, often under very different names and bundled in very different ways. Each chapter takes one dimension (reasoning, learning, knowledge, practical judgment, emotional attunement, social competence, self-regulation, creativity, wisdom, moral discernment, embodied-ecological skill, collective capability) and compares how it has been understood historically, how it is currently measured, and where frontier science is heading. The purpose is to build a reusable map of the actual space that "intelligence" has been used to compress.

Movement III is reconstructive. It argues that "intelligence" should be retired as the master category for human excellence and replaced with a plural, dimensional framework. It presents a set of partly independent axes that different traditions have tracked, different measurement systems assess, and different contexts weight differently. It then considers what this means for education, institutions, AI alignment, and the broader question of how humanity talks about the capacities that matter most.

The arc of the book is: **history, then map, then replacement.**

One final note. This book takes all of the traditions it surveys seriously, not as curiosities or as exotic alternatives to "real" psychology, but as independent, sophisticated attempts to answer the same fundamental questions. It also takes modern psychometrics seriously, not as the villain of the story but as one powerful and partly successful tradition that happens to have been mistaken for the whole story. The goal is not to attack IQ research or to celebrate non-Western thought. The goal is

to see the full landscape clearly, and to build a framework worthy of it.

Chapter 2: The Trap of False Unity

There is a move that human thinking makes so naturally that we rarely notice it happening. We observe a pattern, give it a name, and then begin to treat the name as though it refers to a thing that exists independently of the pattern. The philosopher Alfred North Whitehead called this the "fallacy of misplaced concreteness." The psychologist's term is reification: treating an abstraction as though it were a concrete entity. And the nineteenth-century philosopher John Stuart Mill warned of it directly: "The tendency has always been strong to believe that whatever received a name must be an entity or being, having an independent existence of its own."

This chapter argues that "intelligence" has been reified. Not in a crude or foolish way, but through a subtle process that unfolded over decades and was driven by real scientific achievements. The pattern is genuine: people who do well on one kind of cognitive task tend to do well on others. The name is convenient: calling this pattern "intelligence" gives researchers, educators, and institutions a handle on something they care about. The trap is what comes next: once the name exists, people begin to ask how much intelligence someone has, as though it were a substance that could be weighed, rather than asking what cognitive capacities this person has in what combination and in what context. The shift from pattern to substance is where the trouble begins.

How abstractions become things

Consider a simpler example first. Temperature is a useful abstraction. It summarizes the average kinetic energy of molecules in a system. It can be measured, compared, and tracked over time. But no one thinks that "temperature" is a separate substance floating around inside a cup of coffee. It is a way of describing something about the system, not a thing in the system. The measurement is real. The usefulness is real. But the abstraction remains an abstraction.

Now consider "the economy." This is a much more complex case. When someone says "the economy is struggling," they are using a compression that bundles together employment rates, GDP growth, consumer spending, business investment, trade balances, inflation, debt levels, asset prices, and the lived experience of millions of people trying to pay rent. All of these are real. But "the economy" is not a single entity that can struggle the way a person struggles. It is a label applied to a vast, heterogeneous system. Treating it as a single thing with a single state leads to the kind of confusion where two economists can look at the same data and disagree about whether "the economy" is doing well, because they are weighting different components differently.

Intelligence is closer to "the economy" than to "temperature." It is a label applied to a heterogeneous

set of cognitive capacities: reasoning, memory, learning speed, verbal facility, spatial processing, pattern detection, abstract thinking, and more. The observation that these capacities tend to correlate positively across individuals is real and replicable. The statistical technique that summarizes this correlation, factor analysis, is legitimate. The label *g*, which Charles Spearman proposed in 1904 to name the general factor emerging from these correlations, is a reasonable piece of scientific shorthand.

The reification happens when *g* stops being treated as a statistical summary and starts being treated as a causal explanation. Spearman himself made this move. He speculated that *g* reflected a kind of "mental energy" that powered all cognitive operations. Later researchers debated whether *g* corresponded to a single neural substrate, a general processing speed, or some other unitary biological mechanism. The summary statistic became a hypothesis about a hidden entity. The pattern became a thing.

This matters because once you believe you have found a thing, you stop looking for the more complex reality underneath it. If intelligence is one thing, then measuring it with one number makes sense. If it is one thing, then ranking people along a single axis makes sense. If it is one thing, then asking whether it is primarily genetic or primarily environmental makes sense, because the question presupposes a single target for the genetic or environmental influences to act on. Each of these downstream moves depends on the reification at the top.

The path from Spearman to destiny

The reification of intelligence did not happen in a single step. It unfolded through a sequence of scientific, institutional, and cultural moves, each of which seemed reasonable at the time.

Step one: the statistical discovery. In 1904, Spearman noticed that scores on different cognitive tests tended to correlate positively. He used factor analysis to extract a common factor, which he called *g*. This was a genuine empirical finding, and it has been replicated many times since. People who score well on vocabulary tests tend to score above average on spatial reasoning tests, and so on. The correlations are not perfect, but they are consistent enough to extract a common factor.

Step two: the naming. Calling the common factor "general intelligence" was a choice, not a discovery. Spearman could have called it "the common factor in cognitive test performance," which is a more modest and more accurate description. But "general intelligence" sounds like it names a real property of persons, not a statistical artifact of test batteries. The name itself encouraged reification.

Step three: the measurement apparatus. Binet's test, and later the Stanford-Binet, the Wechsler scales, and other batteries, were designed to produce a single number: an intelligence quotient. The

IQ score became the public face of intelligence. A single number is easy to use, easy to compare, and easy to rank. It is also, by its very structure, a claim that whatever intelligence is, it can be adequately represented by one dimension. The technology reinforced the ontology.

Step four: the institutional adoption. As discussed in Chapter 1, states, armies, schools, and employers adopted IQ-style testing because it was fast, scalable, and predictive. Once millions of people had IQ scores, and once those scores were used to make consequential decisions, the concept of intelligence as a measurable single trait became deeply embedded in institutional practice. Questioning the concept meant questioning the entire system of sorting.

Step five: the cultural absorption. By the mid-twentieth century, "intelligence" had become part of how ordinary people understood themselves and each other. Being "smart" or "not smart" became a core dimension of identity. Parents worried about their children's IQ. Schools tracked students by ability. The single-axis model of intelligence had moved from a statistical observation to a cultural assumption about what kind of person you are.

Each step in this sequence is individually defensible. The correlations are real. The tests are useful. The scores predict things. The institutions needed tools. But the cumulative effect is a kind of conceptual lock-in that is much harder to justify than any individual step. By the time you reach step five, the reification is so deeply embedded that questioning it feels like questioning reality itself. "Of course intelligence is a real thing," people think. "I can see the differences every day." But what they are seeing is a complex, multidimensional pattern of human variation that has been compressed into a single label, and the label has come to feel more real than the pattern.

The counter-arguments, taken seriously

This book does not argue, as Stephen Jay Gould did in *The Mismeasure of Man*, that factor analysis is worthless for studying intelligence or that *g* is merely a mathematical abstraction with no empirical content. Gould's critique was influential and contained important insights, particularly about how scientific methods can be shaped by the cultural assumptions of their practitioners. But his treatment of factor analysis was contested by many statisticians and psychologists, and for good reason. The positive correlations among cognitive tests are robust. Factor analysis is a legitimate tool for summarizing them. And *g* does have predictive value: it correlates with academic performance, job training success, and other outcomes that matter.

The argument here is different and, in some ways, more careful. It is not that *g* is a fraud. It is that *g* is a summary, and summaries leave things out. The question is how much they leave out, and whether what they leave out matters.

Consider an analogy. A person's body mass index (BMI) is a real number derived from real measurements. It correlates with health outcomes. It is useful for population-level screening. But no serious physician treats BMI as a complete description of a patient's health, because it does not distinguish muscle from fat, does not account for bone density, ignores cardiovascular fitness, and says nothing about metabolic health, mental wellbeing, or functional capacity. BMI is a compression. It is useful when you understand its limits and harmful when you mistake it for the whole picture.

IQ and g are compressions of the same kind. They summarize the shared variance among a set of cognitive tests. That shared variance is real and worth studying. But it is not the only variance. The tests also have specific variance: performance on spatial reasoning tests is partly independent of performance on verbal tests, and both are partly independent of performance on processing-speed tasks. These specific factors are not noise. They are real cognitive differences that matter for real-world outcomes. Compressing them into a single number discards information that a profile would preserve.

More importantly, even the full battery of cognitive tests from which g is extracted covers only a fraction of the capacities that human civilizations have valued. It does not measure practical judgment. It does not measure emotional perception or regulation. It does not measure the ability to sustain relationships, navigate social complexity, or act wisely under uncertainty. It does not measure ecological attunement, creative originality, or moral reliability. These are not fringe capacities. They are, as the historical chapters of this book will show, central to what most human traditions have meant by "a good mind" or "a wise person."

So the problem with the reification of intelligence is not that the statistical pattern is fake. The problem is that a real but partial pattern has been promoted to a total description. And once you treat a partial description as total, you systematically undercount everything it leaves out.

Why single axes are tempting and dangerous

There is a deeper reason why reification happens so readily with intelligence. Human beings are drawn to single-axis rankings. We want to know who is best, who is smartest, who is most capable. A single score makes ranking possible. A profile does not, because profiles are multi-dimensional, and multi-dimensional objects do not line up neatly along a single axis.

This preference for ranking is not purely intellectual. It is institutional. Schools need to decide whom to admit. Employers need to decide whom to hire. Armies need to decide whom to promote. A single score makes these decisions fast and defensible. "We admitted the students with the highest scores" is easier to explain than "we admitted the students whose profiles best matched the demands of this

particular program, weighted by practical judgment, collaborative capacity, and self-regulation as well as analytical reasoning.” The first explanation fits in a sentence. The second requires a paragraph and a theory.

The institutional preference for simplicity is understandable. But it has consequences. When you compress a multi-dimensional space into a single axis, you inevitably overweight some dimensions and underweight others. In the case of IQ, what gets overweighted is analytical reasoning and processing speed. What gets underweighted is everything else: judgment, emotional skill, social competence, self-regulation, creativity, moral discernment, and ecological attunement. These are not secondary capacities. For many real-world purposes, they matter more than analytical reasoning alone.

The danger is not just that the ranking is incomplete. It is that the ranking, once established, reshapes the world it claims to describe. If schools select for analytical reasoning, then students who excel at analytical reasoning succeed in school, which confirms that analytical reasoning is what matters, which reinforces the selection criteria. This is a feedback loop, not a proof. The system is measuring its own reflection and calling it truth.

The landscape this book will map

If intelligence is not one thing, then what is it? Or rather, what are the things that have been compressed under its name?

The answer, which the rest of this book will develop, is that humanity has been tracking a set of recurring dimensions of human excellence. These dimensions are partly independent of each other, though they interact in complex ways. They include at minimum:

The capacity for reasoning and abstraction. The capacity to learn and adapt. The body of knowledge and models one has retained. The ability to judge and act well in messy real-world situations. The ability to perceive, understand, and use emotional information. The ability to read, navigate, and sustain human relationships. The ability to direct, control, and monitor one’s own cognition and behavior. The ability to generate novel and useful ideas. The capacity for reflective, perspective-taking, uncertainty-aware judgment. The disposition to act on what one knows to be right. The ability to attune skillfully to environments, bodies, and ecological systems. The ability to contribute to and sustain group intelligence.

Different traditions have bundled these dimensions differently, valued them differently, and recognized them through different institutional methods. Modern psychometrics focuses primarily on the first three, with some attention to the seventh. The other dimensions have been tracked by other

measurement traditions (emotional intelligence tests, situational judgment tests, creativity batteries, wisdom scales) or by non-psychometric methods (debate, apprenticeship, elder recognition, demonstrated performance) or have been largely unmeasured in any standardized way.

The next chapter sets up the comparative method this book will use to map these dimensions across traditions. The chapters that follow trace how different civilizations actually defined, cultivated, and validated them. And the final movement of the book argues that the single-axis model should be replaced with a dimensional profile: not one score, but a shape across multiple partly independent axes. That shape will be more complex than an IQ number. It will also be more honest, more useful, and more worthy of the actual range of human capability.

Chapter 3: Method. Comparing Without Flattening

This book compares how different civilizations have defined, cultivated, and recognized human excellence. That comparison is necessary, because without it we cannot see how local the modern concept of intelligence actually is. But it is also dangerous, because cross-civilizational comparison can go wrong in several characteristic ways, each of which would undermine the project from the start.

This chapter lays out the method. It names the dangers, explains how this book will try to avoid them, and sets up the layered comparison template that every subsequent chapter will use.

Three ways comparison goes wrong

The first and most common error is **Procrustean translation**: mapping every tradition onto Western categories and then congratulating yourself for finding that "they had intelligence too." This is the move that takes the Confucian concept of *xin* and says, "Ah, this is basically their word for mind, which is basically their version of intelligence." It is not. *Xin* encompasses thought, feeling, desire, intention, and moral sensitivity in a single integrated faculty. Translating it as "mind" in the modern English sense strips out most of what makes it distinctive. The same distortion occurs when *prajñā* is rendered as "wisdom" and then filed alongside Western wisdom research, as though Buddhist insight into the nature of reality and modern psychological measures of reflective judgment are the same project. They overlap, but they are not the same, and the differences are as instructive as the similarities.

The danger of Procrustean translation is that it makes everything look like a variation on a Western theme. The result is a false universalism in which every civilization turns out to have been groping toward the same concept, and the West just happened to formalize it first. That is not what the evidence shows. The evidence shows that different civilizations were often tracking genuinely different capacities, or bundling overlapping capacities in genuinely different ways.

The second error is the opposite: **incommensurability despair**, the conclusion that traditions are so different that no meaningful comparison is possible. On this view, Aristotelian *phronēsis* and Akan *nyansa* are embedded in such different philosophical systems that placing them side by side is like comparing colors and sounds. This position has the virtue of taking difference seriously, but it makes the current project impossible and is, in any case, more extreme than the evidence warrants. The philosopher David Wong has explored the conditions under which traditions may be incommensurable, and comparative philosophers from Ram Adhar Mall to Robert Neville have

argued that partial overlap, not identity or total difference, is the normal situation when traditions meet. Two concepts can share a family resemblance without being identical, and that resemblance can be instructive without being a claim of equivalence.

The third error is **decorative pluralism**: mentioning non-Western traditions as colorful supplements to a fundamentally Western story. On this approach, the "real" account of intelligence is the psychometric one, and the Akan or Buddhist or Confucian traditions are interesting footnotes that show how other cultures "also thought about" the same topic. This is perhaps the subtlest of the three errors, because it can look like genuine inclusiveness while actually maintaining the hierarchy it claims to dissolve. The non-Western traditions are included but never allowed to restructure the central argument. They illustrate; they do not challenge.

This book tries to avoid all three. It does not translate every tradition into Western terms. It does not declare traditions incommensurable. And it does not treat non-Western material as decoration. Instead, it treats each tradition as an independent, sophisticated attempt to answer a family of related questions about human excellence, and it asks what we learn when we lay those answers alongside each other without requiring that they be the same.

Four layers of comparison

To make this concrete, the book uses a four-layer comparison template. Every tradition discussed in the historical chapters (Chapters 4 through 14) will be examined along these four layers:

Layer 1: Vocabulary. What terms did this tradition use? What were the key words, and what did they cover? This layer is descriptive and philological. It asks what the tradition actually said, in its own language, before any translation is attempted. For example, the Greek vocabulary includes *nous*, *epistēmē*, *technē*, *sophia*, and *phronēsis*. The classical Chinese vocabulary includes *xin* (心), *zhi* (智), *li* (理), and *de* (德). The Akan vocabulary includes *nyansa*. Each of these words has a semantic range that does not map neatly onto English equivalents.

Layer 2: Phenomena. What real capacities was this tradition tracking? This layer is interpretive. It moves beyond the words to ask what aspects of human cognition, judgment, emotion, behavior, and social life the tradition was actually pointing at. Two traditions may use very different vocabularies and still be tracking overlapping phenomena. Aristotelian *phronēsis* and Akan *nyansa* use completely different words embedded in completely different philosophical systems, but both point at something recognizable: the capacity to judge and act well in practical situations. The overlap is real, but partial. Noting the overlap does not erase the differences; noting the differences does not erase the overlap.

Layer 3: Norms. Which capacities were admired, and why? This layer asks about values. Differ-

ent traditions may track some of the same capacities but weight them very differently. Confucian traditions placed enormous weight on cultivated attunement to roles and relationships. Classical Indian epistemology placed enormous weight on the validity of cognitive sources. Psychometric intelligence research places enormous weight on reasoning speed and abstraction. These are not just different descriptions; they are different evaluations. What counts as "the best kind of mind" depends on what a tradition thinks minds are for.

Layer 4: Recognition. How was competence validated in practice? This layer asks about measurement in the broadest sense. Not just modern standardized testing, but every method a tradition used to select, certify, or publicly acknowledge someone as excellent. Chinese imperial examinations. Tibetan monastic debate. Islamic *ijazah* authorization. Akan proverb competence and elder judgment. Pacific wayfinding apprenticeship. Modern IQ tests, emotional intelligence batteries, and situational judgment tests. All of these are recognition methods. They differ enormously in form, but they all perform the same social function: they validate a claim to competence.

Using these four layers prevents the most common comparison errors. Layer 1 keeps the analysis grounded in what traditions actually said. Layer 2 allows cross-traditional comparison without demanding terminological equivalence. Layer 3 prevents the assumption that all traditions were valuing the same thing. Layer 4 expands the concept of "measurement" beyond standardized testing to include the full range of methods humanity has used to recognize excellence.

Convergence without shared foundations

One of the most important things the historical survey reveals is that traditions can converge on recognizing overlapping capacities without sharing the same philosophical foundations for why those capacities matter.

Consider practical judgment. Aristotle valued *phronēsis* because he believed the good life requires practical wisdom about particular situations, not just theoretical knowledge of universals. Confucian thought valued cultivated judgment because social harmony depends on people acting appropriately in their roles. Akan thought valued *nyansa* because communal flourishing requires leaders and elders who can use knowledge well. Modern situational judgment testing values practical decision-making because employers need people who can handle real-world ambiguity. These are four different justifications for valuing roughly overlapping capacities. The convergence is at the level of what is valued, not at the level of why.

This pattern recurs throughout the survey. Emotional attunement is valued by the Chinese heart-mind tradition (because cognition and feeling are inseparable), by Buddhist thought (because wis-

dom without compassion is incomplete), by Yoruba tradition (because wisdom has both intellectual and emotional dimensions), and by modern emotional intelligence research (because emotion-related skills predict workplace and relationship outcomes). The justifications are different. The phenomenon being pointed at has significant overlap.

The methodological principle this suggests is: **look for convergence in what traditions recognize, not convergence in how they justify it.** If multiple independent traditions, developing in isolation from each other, all arrive at the conclusion that some capacity matters for human excellence, that is evidence that the capacity is real and important, even if the traditions disagree about *why* it is important. The convergence of independent observers is one of the strongest forms of evidence available when studying something as complex as human excellence.

This does not mean that every capacity recognized by any tradition must be real or important. It means that widespread independent convergence is a signal worth taking seriously, and that the absence of a capacity from one particular tradition (say, the modern psychometric one) does not prove that the capacity is not real. It may instead reveal a blind spot.

The archival asymmetry

Any honest cross-civilizational survey must confront an uncomfortable fact: the archive is not neutral. Some traditions left extensive written records. Others encoded their knowledge primarily in oral traditions, apprenticeship systems, ritual practices, proverb collections, and living transmission. The first kind of tradition is easier to study from a desk. The second kind is not less sophisticated; it is less accessible to the methods that modern scholarship finds most comfortable.

This asymmetry has shaped which traditions are treated as "philosophy" and which are treated as "folklore" or "culture." Classical Greek, Indian, and Chinese thought have been extensively analyzed by comparative philosophers, partly because all three produced large bodies of written text that can be translated, annotated, and debated. West African philosophical traditions, Pacific navigation knowledge, Inuit collaborative frameworks, and Māori knowledge systems have received far less attention in global philosophy, not because they are less profound but because they were transmitted in forms that do not look like journal articles.

The consequences for a project like this one are significant. When discussing Aristotle, we have his texts. When discussing Akan *nyansa*, we have ethnographic and philosophical studies, collections of proverbs, and reconstructions by scholars such as Kwame Gyekye and Kwasi Wiredu. When discussing Pacific wayfinding, we have the testimony of living practitioners, the records of anthropologists and cognitive scientists, and the survival of navigational knowledge in communities that

continue to practice it. These are different kinds of evidence, and they require different kinds of interpretive care. But they are all evidence, and dismissing the second and third kinds because they are not "texts" would reproduce exactly the bias this book is trying to correct.

The practical consequence for the book is that the chapters on literate traditions (Greece, Rome, India, China, Islamic civilization) will be able to draw on extensive philosophical texts and scholarly commentary. The chapters on oral and practice-based traditions (Akan, Yoruba, Ubuntu, Māori, Inuit, Pacific) will draw on ethnographic work, philosophical reconstructions, community accounts, and contemporary Indigenous-led scholarship. Both kinds of sources are treated as legitimate evidence about how a tradition defined and recognized human excellence.

What counts as measurement

A final methodological point. This book uses the word "measurement" more broadly than modern psychology does. In psychology, measurement usually means standardized testing: a controlled procedure that produces a quantitative score that can be compared across individuals. That is one important form of measurement, and this book takes it seriously.

But it is not the only form. When Tibetan monastic institutions put candidates through years of public debate before awarding a Geshe degree, they are measuring reasoning ability, doctrinal knowledge, and argumentative composure. When the Chinese imperial examination system tested candidates on literary composition and mastery of the Confucian classics, it was measuring a specific complex of intellectual and moral formation. When an Akan community recognizes an elder as wise on the basis of decades of observed judgment, speech, and conduct, it is measuring practical wisdom. When a Pacific master navigator selects an apprentice and trains them over years before entrusting them with responsibility for a voyage, that selection and training process is measuring a form of ecological and spatial intelligence.

None of these is a standardized test in the modern sense. But all of them perform the core function of measurement: they select, they validate, and they publicly certify competence. They answer the question "Does this person have the capacity we value?" and they do so through methods that the relevant community finds credible.

The book will therefore discuss measurement traditions in two registers. One is the modern psychometric register: batteries, factor structures, reliability coefficients, predictive validity. The other is the institutional register: what methods did a tradition actually use to decide who was excellent, and what capacities did those methods select for? Both registers are important. The first tells us what can be quantified and compared at scale. The second tells us what humanity has actually cared about

enough to build institutions around.

The template in action

To see how this four-layer method works in practice, consider a quick preview.

When we reach Chapter 6 (Classical China), the analysis will proceed as follows. **Vocabulary:** *xin*, *zhi*, *li*, *de*, and the broader Confucian lexicon of moral-intellectual cultivation. **Phenomena:** an integrated faculty of thinking-feeling-intending; a set of cultivated capacities for role-appropriate judgment and action; a body of literary and moral knowledge. **Norms:** excellence is cultivated attunement, not raw cognitive speed; the ideal is the *junzi*, the exemplary person who acts well in all their relationships. **Recognition:** the imperial examination system, teacher-student lineages, community reputation, and the long tradition of self-cultivation as ongoing moral-intellectual work.

When we reach Chapter 10 (Akan and Yoruba traditions), the analysis will proceed differently but with the same structure. **Vocabulary:** *nyansa*, *ogbón*, *ogbón-ori*, *ogbón-inú*, *omolúàbí*. **Phenomena:** the ability to make good use of knowledge; intellectual and emotional dimensions of wisdom; good character as an ideal of human formation. **Norms:** wisdom is practical, social, and communal; it is shown in conduct and speech, not in abstract test performance. **Recognition:** elder judgment, proverb competence, adjudication, and the testimony of community life over time.

The same template applies to modern psychometric intelligence. **Vocabulary:** *g*, IQ, fluid and crystallized intelligence, processing speed. **Phenomena:** reasoning, abstraction, pattern detection, working memory, and their statistical covariance. **Norms:** cognitive performance is valued primarily for its predictive utility in educational and occupational settings. **Recognition:** standardized test batteries administered under controlled conditions, scored quantitatively, and compared against population norms.

Placing these three cases side by side does not require claiming that they are "all the same." They are not. But the four-layer template reveals what each one tracks, what it values, and how it validates competence. And it is only when they are placed side by side that the shape of the larger space becomes visible: the space that no single tradition covers entirely, and that "intelligence" compresses into a single misleading word.

Chapter 4: Ancient Greece. Excellence, Wisdom, Craft, and Practical Judgment

If you wanted to find the moment when the Western world most carefully thought about what it means for a mind to be excellent, you could do worse than Book VI of Aristotle's *Nicomachean Ethics*. Written in the fourth century BCE, it remains one of the most sophisticated decompositions of intellectual excellence ever produced. And it looks almost nothing like modern intelligence testing.

Aristotle identified five distinct intellectual virtues: five different ways the mind can reliably arrive at truth. Each one involves a different kind of object, a different kind of reasoning, and a different relationship to action. None of them is "IQ." Taken together, they represent a tradition's answer to the question this book keeps asking: What does it mean for a mind to be good?

But Aristotle did not invent this question. He inherited it from a philosophical tradition stretching back through Plato and Socrates to the Homeric poets, and from an educational culture, *paideia*, that had been shaping Greek ideals of human excellence for centuries before any philosopher wrote a word. To understand what Greek thought bundled together under the heading of intellectual excellence, we need to start before philosophy, with the poets.

Homeric excellence and the origins of aretē

The Greek word most commonly translated as "virtue" or "excellence" is *aretē*. In its earliest Homeric usage, *aretē* did not mean moral goodness in the modern sense. It meant something closer to functional excellence: being good at what you are supposed to do. The *aretē* of a warrior was courage and skill in battle. The *aretē* of a horse was speed. The *aretē* of a knife was sharpness. To have *aretē* was to fulfill your function superbly.

In the *Iliad* and the *Odyssey*, human *aretē* was displayed through action: through brave deeds, shrewd counsel, eloquent speech, and the ability to navigate complex social and political situations. Odysseus is the paradigm case. He is not praised primarily for abstract reasoning. He is praised for *mētis*, cunning intelligence: the ability to read situations accurately, devise plans, adapt to circumstances, persuade, deceive when necessary, and survive where others would fail. Odysseus is, in modern terms, practically intelligent. He would score poorly on a test of pure abstract reasoning compared to his performance on a test of situational judgment.

This matters because the Homeric model of excellence was the cultural bedrock on which later Greek philosophy built. When Plato and Aristotle argued about what kind of mind is best, they were not starting from scratch. They were inheriting a tradition in which excellence meant demonstrated performance in the real world: in war, in counsel, in speech, in governance. The abstract, theoretical

turn that Plato introduced was, in its own time, a radical departure from this older model, not its natural continuation.

Paideia: education as formation

The Greek word *paideia*, which originally meant simply child-rearing, gradually expanded to encompass the entire project of shaping a person into an excellent human being. Werner Jaeger, in his monumental study of Greek culture, treated *paideia* as the central organizing concept of Hellenic civilization: not just a curriculum, but a civilizational ideal of human formation.

What did *paideia* actually involve? For the sons of free citizens in classical Athens, education included grammatical and literary study (centered on Homer and the poets, who were treated as moral authorities as much as literary ones), music and poetry (which were understood to shape the soul's emotional disposition), physical training (gymnastics and wrestling, which cultivated bodily discipline and courage), and eventually rhetoric, mathematics, and philosophy.

The crucial point for this book is what *paideia* was *for*. It was not designed to measure cognitive ability. It was designed to produce a certain kind of person: someone capable of participating effectively in civic life, of speaking well in the assembly and the law courts, of judging well in practical situations, and of conducting themselves with the composure and character appropriate to a free citizen. The aim was holistic formation, not cognitive benchmarking. When Menander wrote that *paideia* is the one possession no one can take away, he was describing something more like character than like a test score.

This means that the Greek educational tradition was, in the terms of this book, a thick tradition. It bundled together cognitive training, emotional formation, physical cultivation, civic preparation, and moral development into a single integrated project. It did not treat reasoning as a separable module that could be tested independently of judgment, character, and social competence. From the perspective of *paideia*, asking "how intelligent is this person?" without asking "what kind of person are they?" would have been unintelligible.

The Sophists: intelligence as trainable technique

In the fifth century BCE, a new kind of educator appeared in the Greek world: the Sophist. Figures like Protagoras, Gorgias, Hippias, and Prodicus offered, for a fee, to teach the skills needed for success in public life, especially rhetoric, argumentation, and persuasion. They were, in effect, the first professional teachers of what we might now call applied intelligence.

The Sophists are important for this book's argument for several reasons. First, they made an implicit

claim that would later become central to modern intelligence research: that cognitive skills can be trained and improved through deliberate instruction. Second, they focused on a specific kind of intellectual performance, persuasive argumentation, and developed explicit techniques for achieving it. This is an early form of domain-specific skill training, and it anticipated modern debates about whether intelligence is a general capacity or a bundle of specific trainable skills.

Third, and most importantly for the comparison with other traditions, the Sophists provoked a fierce backlash from Socrates and Plato, who argued that rhetorical skill without genuine understanding and moral commitment was not true wisdom but merely its imitation. This debate, between those who define intellectual excellence as effective performance and those who define it as deep understanding joined to moral character, has never been fully resolved. It continues in modern form in arguments about whether practical skills like persuasion and negotiation should count as "intelligence," and whether emotional and moral competence should be considered part of cognitive excellence or separate from it.

Socrates and Plato: the turn toward philosophical understanding

Socrates, as portrayed by Plato, redirected the question of intellectual excellence away from practical skill and toward a different kind of capacity: the ability to examine one's own assumptions, to follow an argument wherever it leads, and to pursue truth even when it is uncomfortable. The Socratic method, *elenchus*, is a technique for exposing contradictions in a person's beliefs, forcing them to confront the limits of what they think they know. Socrates' famous claim that his only wisdom consisted in knowing that he did not know was not false modesty. It was a redefinition of intellectual excellence: the wisest person is the one most aware of the boundaries of their own understanding.

Plato took this further. In dialogues like the *Republic* and the *Meno*, he argued that the highest form of knowledge is not empirical information or practical skill but philosophical understanding of the Forms: the eternal, unchanging structures that underlie the changing world of appearance. The famous allegory of the cave depicts most people as prisoners watching shadows on a wall, mistaking the shadows for reality. The philosopher is the one who turns around, sees the fire casting the shadows, and eventually climbs out of the cave into the sunlight of genuine understanding.

This is an extraordinarily ambitious vision of intellectual excellence, and it is worth noting what it does and does not include. It values abstract reasoning, dialectical skill, and the pursuit of deep structural understanding above practical competence. It is, in some ways, the closest ancient analogue to the modern psychometric emphasis on abstract reasoning and pattern detection. But Plato embedded this cognitive ideal within a broader ethical and political framework. The philosopher is not just clever; the philosopher is the person who, having seen the truth, is morally obligated to

return to the cave and govern wisely. Knowledge and political responsibility are inseparable. This is very far from treating intelligence as a value-neutral cognitive trait.

Aristotle: the five intellectual virtues

Aristotle, Plato's student, retained much of his teacher's commitment to systematic philosophical inquiry but rejected the theory of Forms and brought philosophy back into closer contact with the empirical and practical world. It is in Book VI of the *Nicomachean Ethics* that he laid out the most detailed ancient account of intellectual excellence.

Aristotle identified five states through which the mind attains truth:

Epistēmē (scientific knowledge): knowledge of what is necessary and universal, arrived at through demonstration from first principles. This is the kind of knowledge one has when one can not only state that something is the case but explain *why* it must be the case. It concerns things that cannot be otherwise.

Technē (art, craft, or skill): the capacity to make or produce something reliably under the guidance of reason. A skilled shipbuilder, physician, or sculptor has *technē*. It concerns things that can be otherwise, because production always involves contingent material and circumstances. Aristotle defined it as a productive state involving true reasoning.

Phronēsis (practical wisdom or prudence): the capacity for good deliberation about how to live well and act well in particular circumstances. This is the intellectual virtue most closely connected to moral virtue. A person with *phronēsis* can perceive the relevant features of a situation, deliberate about the best course of action, and act accordingly. It concerns things that can be otherwise, because practical situations are always particular and contingent.

Nous (intuitive reason or direct understanding): the capacity to grasp first principles directly, without demonstration. First principles cannot be proved by other first principles; they must be apprehended by a faculty that Aristotle calls *nous*. This is the starting point of all demonstrative knowledge.

Sophia (theoretical wisdom): the combination of *nous* and *epistēmē* directed at the highest and most fundamental objects of knowledge. *Sophia* is what the philosopher has when they understand not just particular domains but the deepest principles of reality.

Several things about this framework are striking from the standpoint of the modern intelligence debate.

First, Aristotle explicitly recognized that excelling at one intellectual virtue does not guarantee excelling at another. He noted that thinkers like Anaxagoras and Thales were credited with *sophia* (theoretical wisdom) but were said to lack *phronēsis* (practical wisdom). They could contemplate

the cosmos but could not manage their own affairs. This is a direct ancient anticipation of the observation that high abstract reasoning ability does not always predict practical effectiveness.

Second, *phronēsis* is not reducible to *epistēmē* or *technē*. It requires knowledge of universals (general principles about how to live) *and* knowledge of particulars (the specific features of this situation, here and now). Aristotle stressed that practical wisdom depends on experience, because experience gives familiarity with particulars that no amount of abstract knowledge can substitute. Young people can be mathematicians, he said, but they cannot be practically wise, because practical wisdom requires the kind of pattern recognition that only accumulated experience provides.

Third, the intellectual virtues are not isolated from the moral virtues. Aristotle argued that *phronēsis* is impossible without moral virtue, and moral virtue is impossible without *phronēsis*. The person who has good character but no practical wisdom will want to do the right thing but will not know how. The person who has practical cleverness but no moral character will be effective but will pursue the wrong ends. Genuine practical wisdom requires both: the right values *and* the ability to realize them in action. This is a deeply integrated account of excellence that resists any clean separation between cognition and character.

What Greek thought bundled together

Applying the four-layer template from Chapter 3:

Vocabulary. The Greek lexicon of intellectual excellence is rich and differentiated: *aretē* (excellence), *nous* (intuitive reason), *epistēmē* (scientific knowledge), *technē* (craft/skill), *sophia* (theoretical wisdom), *phronēsis* (practical wisdom), *mētis* (cunning intelligence), *paideia* (formation/education), *logos* (reason/account/speech).

Phenomena. Greek thought was tracking at least the following capacities: abstract reasoning, logical demonstration, craft skill, practical judgment in contingent situations, persuasive speech, self-knowledge, the ability to examine assumptions, and the integration of knowledge with moral character. It also recognized, in the figure of Odysseus and the tradition of *mētis*, a kind of situational intelligence that is closer to what modern psychology calls practical intelligence or adaptive expertise.

Norms. Different Greek thinkers valued these capacities differently. The Homeric tradition valued demonstrated performance and cunning. The Sophists valued persuasive technique. Plato valued abstract philosophical understanding. Aristotle valued the full set, but gave special attention to *phronēsis* as the intellectual virtue most closely connected to living well. Across these differences, there is a consistent thread: intellectual excellence is inseparable from how one lives. It is not a neutral

cognitive trait but a capacity directed at human flourishing.

Recognition. How was intellectual excellence identified and validated? Through public performance: in the assembly, in the law courts, in the gymnasium, in philosophical dialogue, and in the competitive display of rhetorical and argumentative skill. Greek education was not assessed by standardized tests. It was assessed by the quality of one's speech, the soundness of one's judgment, the effectiveness of one's action, and the conduct of one's life. The closest thing to a formal examination was the Sophistic practice of public disputation and display, and even that was more performative than psychometric.

What this means for the book's argument

The Greek case is instructive in two directions.

Looking backward from the modern present, it shows that the Western tradition's own deepest roots contain a model of intellectual excellence far richer and more plural than the one that modern psychometrics inherited. When IQ tests measure abstract reasoning and processing speed, they are capturing something roughly analogous to a thin slice of *nous* and *epistēmē*. They are largely ignoring *phronēsis*, *technē*, the integration of cognition with character, and the entire domain of practical, social, and civic judgment that the Greek tradition treated as central to being a well-formed human being. The narrowing was not inevitable. It was a choice, made for institutional reasons, and the Greek material shows how much the choice left behind.

Looking forward toward the non-Western traditions covered in later chapters, the Greek case also sets up an important comparison. Greek thought is unusual in the degree to which it differentiated and named distinct intellectual capacities. Other traditions, as we will see, often kept these capacities more tightly bundled. The Chinese *xin* integrates cognition and emotion in a way that Aristotle's framework does not. Buddhist *prajñā* connects wisdom to compassion in a way that even *phronēsis* does not. Akan *nyansa* foregrounds the social and communal dimensions of wisdom more than any Greek category. These are not deficiencies. They are different bundling choices, each illuminating something that the others underemphasize.

The Greek tradition, then, is not the origin of the correct account of intelligence. It is the origin of one extraordinarily influential decomposition of the space. And its most important lesson may be the one Aristotle himself drew: that there is no single intellectual excellence, but several, and that the practically wise person, not the abstractly brilliant one, is the one who lives well.

Chapter 5: Rome. Civic Judgment, Eloquence, and the Educated Person

Greece differentiated the intellectual virtues. Rome fused them back together in a new way, organized around a different central question. Where Aristotle asked, "What are the distinct ways the mind can arrive at truth?", the Roman tradition asked something more practical: "What kind of person should be entrusted with public responsibility, and how do we form them?"

The answer Rome gave is captured in a phrase attributed to Cato the Elder, the earliest formulation of a Roman educational ideal: *vir bonus dicendi peritus*, the good man skilled in speaking. Five words that contain an entire philosophy of human excellence. Note what the formula bundles together: moral character (*vir bonus*, the good man) and communicative competence (*dicendi peritus*, skilled in speaking). Not one or the other. Both, inseparably. A person who speaks brilliantly but lacks moral character is dangerous. A person of good character who cannot speak effectively is useless to the republic. The ideal is their union.

This chapter traces how that formula was developed, institutionalized, and transmitted. Rome did not produce a new theoretical decomposition of intellectual excellence to rival Aristotle's. What it produced was something arguably more consequential for the long history of education: a practical model of the formed person, embedded in a curriculum, that shaped Western schooling for nearly two thousand years.

Cicero: wisdom and eloquence reunited

Marcus Tullius Cicero, writing in the first century BCE, was the figure who most ambitiously developed the Roman ideal of the orator-statesman. His *De Oratore* (On the Ideal Orator), composed in 55 BCE during the crisis of the late Republic, is not merely a technical manual on rhetoric. It is a philosophical argument that wisdom and eloquence, which the Greek philosophical tradition had separated, must be reunited in the person of the ideal orator.

Cicero's complaint was that philosophy and rhetoric had been torn apart. The philosophers had retreated into their schools, pursuing abstract wisdom disconnected from public life. The rhetoricians had developed persuasive technique disconnected from genuine understanding. The result was that neither group could do what the republic most needed: communicate truth effectively in the service of governance. Cicero's ideal orator would heal this split. He would be a person of broad philosophical learning who could also speak with the force, clarity, and emotional power needed to move public assemblies and shape political decisions.

The demands Cicero placed on this ideal figure were extraordinary. In a famous passage, he wrote

that in the orator one must demand the subtlety of the logician, the wisdom of the philosopher, the language almost of poetry, the memory of lawyers, the voice of tragedians, and the bearing almost of the best actors. This is not a description of a specialist. It is a description of a generalist of the highest order: someone who integrates intellectual depth, verbal artistry, emotional intelligence, physical presence, and moral authority into a single public performance.

For the purposes of this book, several features of Cicero's model are significant.

First, Cicero explicitly argued that the orator must understand human emotion. Persuasion is not merely logical argument. It requires the ability to perceive what an audience feels, to calibrate emotional appeals, and to manage one's own emotional expression. Cicero insisted that the orator could not move others unless he was himself genuinely moved. This is, in modern terms, a demand for both emotional perception and emotional regulation, capacities that would not be formally named as "emotional intelligence" for another two millennia but that Cicero treated as central to intellectual excellence.

Second, Cicero's ideal is irreducibly practical. The orator is not a contemplative philosopher. He is a public actor: someone who must perform under pressure, in real time, before hostile audiences, with political stakes. The excellence Cicero admired was not the kind that shows up under controlled laboratory conditions. It was the kind that shows up in the Roman Forum, where a wrong word could end a career and the right one could save a republic. This is closer to what modern psychology calls situated performance or ecological validity than to standardized test-taking.

Third, Cicero understood that this kind of excellence could not be produced by narrow training. The orator needed philosophy, history, law, poetry, and experience. He needed what Cicero called *humanitas*: the broad cultivation that makes a person fully human, not merely technically competent. This concept of *humanitas* would later become the foundation of Renaissance humanism and, through it, of the modern liberal arts tradition.

Quintilian: the curriculum of the complete person

If Cicero articulated the ideal, Quintilian institutionalized it. Marcus Fabius Quintilianus, writing his twelve-volume *Institutio Oratoria* (The Education of the Orator) around 95 CE, produced the most comprehensive educational treatise of antiquity. It covers the entire arc of a person's formation, from infancy to the mature orator's retirement, and it remained a foundational text in European education from its rediscovery in 1416 until well into the modern period.

Quintilian's opening statement sets the tone: "My aim, then, is the education of the perfect orator. The first essential for such a one is that he should be a good man, and consequently we demand of

him not merely the possession of exceptional gifts of speech, but of all the excellences of character as well.” This is Cato’s formula made systematic. Quintilian did not treat moral goodness as a nice addition to rhetorical skill. He treated it as a prerequisite. His argument was that eloquence in the hands of a morally deficient person becomes a weapon against the common good. Only the good man can truly speak well, because only the good man deserves the trust that gives his words their force.

The curriculum Quintilian designed was accordingly broad. It began with careful attention to the child’s earliest environment: the quality of the nurses, the speech habits of the household, the selection of companions. Formal education proceeded through grammar (literary study, centered on close reading and composition), rhetoric (the theory and practice of persuasive speaking and writing), and a wide range of supporting disciplines including music, geometry, astronomy, and philosophy. Throughout, Quintilian emphasized that education was not the transmission of information but the formation of character. The student was being shaped into a certain kind of person, not merely equipped with a certain set of skills.

What is most striking about Quintilian for this book’s purposes is the sheer breadth of what he expected the educated person to command. His orator must master invention (the discovery of arguments), arrangement (the structuring of discourse), style (the selection of appropriate language), memory (the retention of material for extemporaneous delivery), and delivery (the control of voice, gesture, facial expression, and physical bearing). These five *canons of rhetoric* are not narrowly linguistic skills. They describe a complex integration of cognitive capacity, emotional control, physical training, and social perception.

Memory, in particular, deserves attention. In a world without teleprompters, the orator’s ability to retain and reproduce complex material was a cognitive achievement of the first order. Roman orators developed elaborate mnemonic techniques, including the *method of loci* (associating items to be remembered with locations in a familiar building), which represents one of the earliest systematic technologies for augmenting cognitive performance. This is a form of trained intelligence that modern testing does not capture but that Roman culture valued enormously.

Roman law and institutional judgment

Rome added something to the Greek inheritance that is easy to overlook but profoundly important: a developed legal system that demanded, and therefore cultivated, a specific kind of practical intelligence.

Roman jurists were required to interpret complex legal codes, apply general principles to particu-

lar cases, balance competing claims, and render judgments that would be accepted as fair by the community. This is a direct institutional embodiment of something close to Aristotelian *phronēsis*, but operating within a formal legal framework rather than the informal domain of personal ethics. Roman law required the capacity for analogical reasoning (applying precedents to novel cases), the ability to distinguish relevant from irrelevant features of a situation, and the judgment to balance the letter of the law against its spirit.

The Roman contribution to the history of "intelligence," then, is not a new philosophical theory but a new institutional context. Rome demonstrated that practical judgment could be formalized into professional practice, taught through apprenticeship, and applied at scale in the governance of an empire. When a young Roman advocate attached himself to an established jurist and followed him through the forum, observing how he argued cases and rendered judgments, that apprenticeship was a recognition method: a way of transmitting and validating a specific kind of intellectual excellence that no written test could capture.

The Stoic contribution: rational self-command

Any account of Roman intellectual ideals would be incomplete without mentioning Stoicism, which became the dominant philosophical framework among the Roman educated elite from the late Republic onward. Stoic philosophy contributed a distinctive element to the Roman model of excellence: the ideal of rational self-command.

For the Stoics, the sage was a person whose reason had achieved complete mastery over the passions. Virtue consisted in living according to reason and nature, accepting what is beyond one's control with equanimity, and exercising disciplined judgment in all circumstances. This ideal of the rationally self-governed person is directly relevant to what modern psychology calls self-regulation or executive control: the ability to direct one's own cognition, inhibit impulsive responses, maintain focus under stress, and pursue long-term goals despite short-term discomfort.

Cicero himself drew heavily on Stoic ethics even as he modified their rhetorical style. His *De Officiis* (On Duties) is essentially a Stoic moral handbook adapted for Roman practical life. And the later Roman Stoics, Seneca, Epictetus, and Marcus Aurelius, all developed the theme that the quality of a person's inner life, the degree to which reason governs emotion rather than being governed by it, is the truest measure of their excellence. This is not intelligence as cognitive performance. It is intelligence as self-governance. And it is a capacity that most modern IQ tests do not even attempt to measure.

What Roman thought bundled together

Applying the four-layer template:

Vocabulary. The key terms are *sapientia* (wisdom), *prudentia* (prudence, practical judgment), *eloquentia* (eloquence), *humanitas* (cultivated humanity, broad formation), *vir bonus dicendi peritus* (the good man skilled in speaking), *ars* (art, skill, craft), *memoria* (memory), *iudicium* (judgment), and the Stoic vocabulary of *ratio* (reason), *virtus* (virtue), and *officium* (duty).

Phenomena. Roman intellectual culture was tracking: persuasive communication, practical and legal judgment, moral character, emotional perception and regulation, trained memory, physical composure and presence, broad cultural formation, and rational self-command. It was a thick tradition that refused to separate cognitive skill from moral character, verbal ability from emotional intelligence, or intellectual formation from civic responsibility.

Norms. Excellence was valued not primarily for its own sake (as Aristotelian *sophia* might be) but for its service to the republic and to the community. The ideal was civic: the educated person is excellent *for* something, namely the governance and flourishing of public life. This instrumental-but-not-cynical valuation distinguishes the Roman model from both the Greek contemplative ideal and the modern psychometric emphasis on prediction.

Recognition. How was excellence validated? Through public performance in the forum, the law courts, the senate, and the school of rhetoric. Through apprenticeship to established practitioners. Through the judgment of peers, audiences, and patrons. And through the long-term track record of a person's career in public service. Quintilian's curriculum, with its structured progression from grammar through rhetoric to advanced practice, represents an early form of systematic educational assessment, though it was qualitative and situated rather than standardized and decontextualized.

What Rome means for the argument

The Roman case matters for this book in three specific ways.

First, it shows that the integration of cognitive skill with moral character is not a modern invention or a non-Western curiosity. The Western tradition's own deepest educational roots contain exactly this integration. Cato's formula, Cicero's ideal orator, and Quintilian's educational program all insist that separating intelligence from character produces something defective: the clever scoundrel or the well-meaning incompetent. When modern critics argue that IQ tests should be supplemented with measures of character, moral judgment, or emotional skill, they are not importing foreign ideas into the Western tradition. They are recovering something that the Western tradition itself once knew and then forgot.

Second, Rome demonstrates how an educational ideal shapes institutional recognition. The Roman curriculum of grammar and rhetoric, formalized by Quintilian and transmitted through the medieval trivium (grammar, rhetoric, logic) and quadrivium (arithmetic, geometry, music, astronomy), became the template for European education for over a millennium. What got measured, taught, and rewarded in Western schools for most of their history was not abstract reasoning ability but the broader formation captured by *humanitas*. The narrowing to cognitive benchmarking is a very recent development in this long story.

Third, the Stoic contribution introduces a dimension of excellence, rational self-governance, that will recur across multiple traditions in this survey. Confucian self-cultivation, Buddhist mindfulness, and modern executive-function research are all, in different ways, concerned with the same capacity: the ability to direct and regulate one's own cognitive and emotional processes. The Stoics named it clearly, practiced it systematically, and treated it as central to human excellence. Modern IQ tests leave it largely untouched.

Chapter 6: Classical China. Heart-Mind, Cultivation, Examination, and Office

The previous two chapters traced how Greece and Rome defined intellectual excellence. Both traditions were rich and plural, far more so than modern IQ thinking. But both operated within a framework that, at least in its philosophical articulation, tended to distinguish between different faculties of the mind. Aristotle separated *nous* from *epistēmē* from *phronēsis*. Cicero separated wisdom from eloquence, even as he argued they should be reunited. The distinctions were different in each case, but the impulse to differentiate was shared.

Classical Chinese philosophy begins from a fundamentally different starting point. The concept at its center, *xin* (心), refuses the distinctions that Greek and Roman thought took as natural. *Xin* is usually translated as "heart-mind," and the hyphen is doing real work. It is not that the Chinese had a concept for heart and a separate concept for mind and then combined them. It is that the single concept *xin* encompasses what Western languages split apart: thinking, feeling, desiring, intending, understanding, and morally evaluating are all functions of the same faculty. There is no separate cognitive module that operates independently of emotion, desire, and moral sensitivity. To have a well-functioning *xin* is to think, feel, and judge well simultaneously, not as three separate achievements but as one integrated capacity.

This chapter argues that classical Chinese thought offers not just a different answer to the question "What is intelligence?" but a different *framing* of the question itself. And it produced, in the imperial examination system, the most sustained and large-scale institutional attempt in pre-modern history to select for intellectual excellence.

Xin: the heart-mind as integrated faculty

The Stanford Encyclopedia of Philosophy's entry on mind in Chinese philosophy puts the point directly: *xin* is credited with thinking (*si* 思), understanding (*míng* 明), knowing (*zhi* 知), intention (*zhi* 志), felt moods and emotions (*qing* 情), and desire (*yu* 欲). The Routledge Encyclopedia adds that *xin* "preludes the assumption of distinctions between thinking and feeling, or idea and affect." Chinese philosophers, as Chad Hansen has argued, did not explain behavior by a combination of cognitive and affective factors. The *xin* guides action holistically, not through a Cartesian chain of belief leading to desire leading to action.

This is not a vague or underdeveloped position. It is a sophisticated philosophical stance with precise implications. For Mencius (Mengzi), the most influential early Confucian philosopher after Confucius himself, the heart-mind contains innate *duan* (端), seeds or sprouts of goodness. These sprouts

are simultaneously emotional and cognitive. The feeling of alarm and compassion upon seeing a child about to fall into a well is both a recognition (cognitive) that another person is about to suffer and a response (affective) of distress and compassion. For Mencius, the cognitive and affective components are not separate processes that happen to co-occur. They are aspects of a single, integrated response of the heart-mind.

The sprouts include compassion (which develops into the virtue of benevolence, *ren*), shame and dislike (which develop into righteousness, *yi*), deference and compliance (which develop into propriety, *li*), and the sense of right and wrong, *shi-fei* (which develops into wisdom, *zhi*). Notice that even wisdom, the capacity closest to what the Western tradition calls intelligence, arises here from an evaluative discrimination, a sense of right and wrong, that is as much moral and emotional as it is cognitive.

For the purposes of this book, the most important consequence of the *xin* framework is what it does to the concept of intelligence. If cognition and emotion are not separate faculties, then measuring cognitive ability independently of emotional and moral functioning is not just incomplete; it is based on a false division of the person. A Chinese thinker encountering a modern IQ test would not think, "This measures intelligence but leaves out emotion." They would more likely think, "This measures a fragment of the heart-mind while treating the fragment as though it were the whole."

This does not mean that Chinese thinkers were unaware of individual differences in cognitive ability. The Confucian tradition clearly recognizes that some people learn faster, understand more deeply, and judge more perceptively than others. But these differences are understood within the integrated *xin* framework. A person who reasons brilliantly but feels nothing, or who understands principles but cannot apply them in their relationships, does not have an excellent heart-mind. They have a disordered one.

Confucian cultivation: learning as self-transformation

If the heart-mind is an integrated faculty, then cultivating it requires an integrated practice. This is what the Confucian tradition calls self-cultivation (*xiuji* or *xiushen*), and it is one of the most developed traditions of deliberate human formation in world history.

Confucius (Kongzi, 551-479 BCE) taught that human beings are not born excellent. They become excellent through learning (*xue*), practice, and sustained effort at self-correction. The *Analekta* are filled with passages about the importance of study, reflection, and the constant work of aligning one's conduct with the standards set by the ancient sages and encoded in ritual (*li*). Learning, in the Confucian sense, is not the accumulation of information. It is the transformation of the self:

becoming a person who perceives correctly, feels appropriately, judges wisely, and acts well in all their relationships.

The ideal toward which this cultivation aims is the *junzi* (君子), the exemplary person or "gentleman" in older translations. The *junzi* is not primarily intelligent in the modern sense. The *junzi* is cultivated: formed through study, ritual practice, and moral effort into someone who acts with spontaneous appropriateness in all circumstances. This ideal is closer to Aristotle's *phronimos* (the person of practical wisdom) than to any modern concept of the intelligent person, but it goes further than *phronēsis* in integrating emotional attunement and relational skill into the core of excellence.

A critical feature of Confucian self-cultivation for this book's argument is its emphasis on *role-attunement*. The *junzi* does not exercise excellence in the abstract. Excellence is always exercised in the context of specific relationships: as a child, a parent, a sibling, a friend, a ruler, a subject. Each relationship has its own appropriate form of conduct, and the cultivated person knows how to adjust their behavior to the relational context. This is a form of social intelligence that is deeply embedded in the tradition's core concept of excellence, not bolted on as an afterthought.

Xunzi (c. 310-235 BCE), the other major early Confucian thinker, disagreed with Mencius about whether human nature is originally good or originally tending toward disorder, but agreed that cultivation is essential. For Xunzi, the heart-mind can be cultivated through deliberate effort, and the *xin* plays an active role in this process, selecting, evaluating, and directing one's development. This is an early articulation of something close to what modern psychology calls metacognition: the heart-mind's capacity to reflect on and regulate its own functioning.

The Daoist counterpoint

Not all Chinese traditions agreed with the Confucian model. Daoism, particularly as expressed in the *Daodejing* (attributed to Laozi) and the *Zhuangzi*, offered a pointed counter-ideal. Where Confucianism valued deliberate cultivation, ritual propriety, and role-attunement, Daoism valued spontaneity (*ziran* 自然), naturalness, and a kind of effortless skill that arises from dropping learned categories rather than accumulating them.

Zhuangzi's famous stories of the butcher Cook Ding, who carves an ox with supernatural precision because his spirit encounters the ox rather than his eyes seeing it, or the swimmer who moves through dangerous rapids because he has internalized the water's patterns rather than fighting them, describe a form of excellence that is neither cognitive in the modern sense nor emotional in the standard sense. It is embodied, spontaneous, and achieved by forgetting learned distinctions rather than mastering them.

For this book's purposes, the Daoist tradition is important because it offers a completely different model of what excellence looks like. If Confucianism is the closest Chinese analogue to a "cultivated intelligence" model, Daoism is closer to what modern psychology might call flow, embodied expertise, or ecological attunement. Both are Chinese traditions. Both are about what it means for a person to function excellently. But they disagree profoundly about how excellence is achieved and what it looks like in practice.

This internal diversity within Chinese thought is a useful reminder that "the Chinese view of intelligence" is not a monolith. Like Greek thought, Chinese thought contained vigorous internal debates about what kind of excellence matters most. The Confucian-Daoist disagreement is, in its own way, as significant as the Plato-Aristotle or Sophist-Socrates disagreements in the Greek world.

The imperial examination: merit at scale

China produced something that no other pre-modern civilization achieved at comparable scale: a formal, written, competitive examination system for selecting government officials on the basis of demonstrated intellectual and moral-literary formation. The imperial examination system (*keju*) was systematized during the Sui Dynasty (581-618 CE) and became the dominant pathway to office during the Song Dynasty (960-1279 CE). It lasted over 1,300 years, until its abolition in 1905. During the Qing Dynasty in the nineteenth century, two to three million candidates sat for the examinations annually, with a passage rate of roughly one percent.

The examinations were structured in ascending levels: county, provincial, metropolitan, and palace. Successful candidates at each level earned titles (*xiucaai*, *juren*, *gongshi*, *jinshi*) that conferred increasing social status and eligibility for government positions. The highest degree, *jinshi* (presented scholar), was earned by perhaps a few hundred candidates per cycle and virtually guaranteed a senior government appointment. The system produced more than 100,000 *jinshi* over its history and millions of *juren* (provincial degree holders).

What did the examinations test? Not abstract reasoning in the modern psychometric sense. The core content was mastery of the Confucian classics, literary composition (including poetry and the highly structured "eight-legged essay"), calligraphy, and, at various periods, knowledge of law, military strategy, revenue and taxation, agriculture, and geography. During the Ming and Qing dynasties, the content narrowed to the Four Books and Five Classics of Neo-Confucian orthodoxy, with composition required in a rigid formal structure.

Several features of the *keju* are significant for this book's argument.

First, the examinations were explicitly meritocratic in aspiration. In theory, any male adult in China,

regardless of wealth or social status, could become a high-ranking official by passing the examinations. In practice, wealth helped enormously (preparation required years of study and access to books and tutors), but the system genuinely produced significant social mobility. Sons of farmers and merchants did reach high office through the examinations. This makes the *keju* the earliest large-scale institutional embodiment of the idea that intellectual excellence can and should be measured, and that access to power should depend on demonstrated competence rather than birth.

Second, what was being measured was not raw cognitive ability but cultivated formation. The examinations tested whether a candidate had internalized the Confucian moral-literary tradition deeply enough to compose elegant, structured, and insightful responses under timed conditions. This is closer to what modern assessment might call crystallized knowledge combined with sophisticated literary production than to anything measured by a fluid reasoning test. The assumption was that a person who had mastered the classics at this level had undergone the moral-intellectual formation necessary to govern well.

Third, the *keju* system had real and widely recognized defects. Critics within the Chinese tradition itself argued that the examinations tested literary skill rather than practical governance ability, that the rigid eight-legged essay format rewarded conformity over originality, and that the system produced scholars who could quote Confucius but could not manage a granary. This internal criticism anticipates, by centuries, the modern critique that IQ tests measure a narrow band of cognitive performance that does not predict all forms of real-world competence.

Fourth, the *keju* influenced the development of civil service examinations worldwide. The British civil service examination, introduced in the mid-nineteenth century, was explicitly modeled in part on Chinese precedent. Modern civil service systems in many countries trace their lineage, directly or indirectly, to the Chinese example. When this book argues that measurement traditions are older and more diverse than modern psychometrics, the *keju* is the single most powerful exhibit.

What Chinese thought bundled together

Applying the four-layer template:

Vocabulary. *Xin* (heart-mind), *zhi* / (knowing/wisdom), *ren* (benevolence, humaneness), *li* (ritual propriety), *yi* (righteousness), *de* (moral virtue/power), *xue* (learning), *junzi* (exemplary person), *xiushen* (self-cultivation), *dao* (the Way), *ziran* (naturalness, spontaneity).

Phenomena. Chinese thought was tracking: an integrated cognitive-emotional-moral faculty (the heart-mind); the capacity for cultivated attunement to relational roles; literary and moral knowledge of the classical tradition; practical governance judgment; self-regulation and self-correction; sponta-

neous embodied skill (in the Daoist register); and the social-emotional capacities needed to maintain harmonious relationships. These overlap significantly with several of the dimensions this book will map in Movement II: practical judgment, emotional attunement, social-relational competence, self-regulation, moral discernment, and retained knowledge.

Norms. Excellence is not raw cognitive speed or abstract reasoning ability. It is cultivated attunement: the capacity, developed through sustained effort, to perceive, feel, judge, and act appropriately in the specific relational contexts one inhabits. The *junzi* ideal values breadth of formation over narrow specialization, moral character over technical brilliance, and relational harmony over individual achievement. This is a thick normative framework that resists any separation of intelligence from character.

Recognition. The primary institutional recognition method was the imperial examination system: a multi-level, competitive, written examination that tested mastery of the Confucian classics and literary composition. This was supplemented by teacher-student lineages, community reputation, and the long tradition of self-cultivation as ongoing work. The *keju* is remarkable both for its scale (millions of candidates over centuries) and for its explicit commitment to meritocratic selection. It is also remarkable for its acknowledged limitations: the system's own critics recognized that literary mastery does not guarantee governing competence.

What this means for the argument

The Chinese case is pivotal for this book in three respects.

First, it demonstrates that the separation of cognition from emotion is not a universal human intuition. It is a specific philosophical move, made in the Western tradition, that other traditions did not make. The *xin* framework treats thinking-feeling-intending as a single integrated process. This is not a failure to differentiate. It is a different, and arguably more accurate, model of how the mind actually works. Modern affective neuroscience and interoception research increasingly support the view that cognition and emotion are deeply intertwined, not cleanly separable modules. The Chinese tradition arrived at this conclusion two and a half millennia before the neuroscience confirmed it.

Second, the *keju* system proves that large-scale standardized assessment is not a Western invention. China was selecting officials through competitive written examinations a thousand years before Binet developed his intelligence test. The difference is in what was tested: the *keju* tested moral-literary formation, not cognitive ability abstracted from character. When modern psychometrics claims to have invented the idea of measuring intellectual excellence, it is making a historical error. What it invented was a particular kind of measurement, focused on a particular (narrow) band of human

capacity. The idea of measurement itself is much older and was first implemented at scale in China, not in the West.

Third, the Confucian-Daoist debate shows that even within a single civilization, there is no consensus about what excellence looks like. Confucians valued deliberate cultivation, ritual mastery, and role-attunement. Daoists valued spontaneity, embodied skill, and the dropping of learned categories. Both are coherent accounts of human excellence. Neither maps cleanly onto modern IQ. And the existence of their disagreement is itself evidence that "intelligence" is not a natural kind with a single correct definition, but a contested space where different traditions foreground different capacities.

Chapter 7: Classical India. Valid Knowing, Intellect, and the Sources of Knowledge

Every tradition discussed in this book so far has started from some version of the question: What makes a mind excellent? The Greek tradition asked what intellectual virtues a person can develop. The Roman tradition asked what kind of person should be entrusted with public responsibility. The Chinese tradition asked how the heart-mind should be cultivated for relational attunement. Classical Indian philosophy starts from a different question entirely, one that makes it unique in this survey:

What makes a cognition valid?

This shift in starting point has profound consequences. Where other traditions begin with the knower and ask what capacities they should develop, much of Indian epistemology begins with the act of knowing itself and asks under what conditions it produces genuine knowledge rather than error. The focus is not primarily on ranking minds but on understanding the structure of knowing. The result is one of the most rigorous and analytically precise traditions of epistemology ever developed, one that rivals and in some respects anticipates the concerns of contemporary analytic philosophy.

The *pramāṇa* framework: sources of valid knowledge

The concept at the center of Indian epistemology is *pramāṇa* (), which translates roughly as "source of valid knowledge," "means of knowing," or "epistemic instrument." A *pramāṇa* is not just any way of forming a belief. It is a reliable method that, when operating correctly, produces genuine knowledge (*pramā*) rather than error. The question that drove centuries of Indian philosophical debate was not "How intelligent is this person?" but "Through which sources does valid knowledge arise, and how can we distinguish genuine knowing from its counterfeits?"

Different Indian philosophical schools accepted different numbers and types of *pramāṇas*, and their disagreements about this were among the most sustained and technically sophisticated debates in the history of philosophy.

Pratyakṣa (perception) is the most widely accepted source. All major Indian schools, Hindu and Buddhist alike, grant that direct sensory contact with an object can produce valid knowledge. But even here the analysis is subtle. The Nyāya school distinguished between ordinary perception (where the senses contact objects in the usual way) and extraordinary perception (where contact occurs through unusual means, such as perceiving a universal through a particular instance). They also distinguished between non-conceptual perception (a raw, pre-linguistic awareness of an object) and conceptual perception (where the perceived object is categorized and named). These are distinctions that Western epistemology and cognitive science would not develop in comparable detail until the

twentieth century.

Anumāna (inference) is the second major source. Indian logicians developed a formal structure for inference that includes a five-step syllogism: the proposition to be proved, the reason, the example illustrating the general principle, the application of the principle to the case at hand, and the conclusion. The classic example: There is fire on the hill (proposition). Because there is smoke there (reason). Wherever there is smoke, there is fire, as in a kitchen (example with general principle). There is smoke on the hill (application). Therefore there is fire on the hill (conclusion). This structure was developed with extraordinary precision by the Nyāya school and later refined by Buddhist logicians like Dignāga and Dharmakīrti, who introduced new standards for what counts as a valid inferential sign.

Śabda (testimony or verbal knowledge) is accepted by most Hindu schools as a distinct, irreducible source of knowledge. Knowledge gained from the words of a reliable person (*āpta*) is not reducible to perception or inference; it is a separate epistemic channel. The Nyāya school argued that testimony is valid when the speaker is trustworthy and the statement is coherent. The Mīmāṃsā school went further, arguing that the Vedic scriptures are inherently valid (*apauruṣeya*, authorless and beginningless), not because they were composed by a reliable person but because they have no human author at all and therefore cannot contain the errors to which human cognition is prone.

Upamāna (comparison or analogy) is accepted by Nyāya and some other schools as a fourth distinct source. It is the knowledge that arises when one encounters a new object and recognizes it based on a prior description of its resemblance to something already known.

Two additional *pramāṇas* are accepted by some schools but disputed by others: **arthāpatti (postulation)**, which is knowledge of something that must be the case in order to explain an otherwise inexplicable observation, and **anupalabdhi (non-apprehension)**, which is knowledge of absence, knowing that something is not present by failing to perceive it when conditions for perception are met.

The fact that different schools accepted different numbers of *pramāṇas* (from two in Buddhism, to four in Nyāya, to six in Advaita Vedānta and certain Mīmāṃsā schools) is itself philosophically significant. It shows that even within a single broad tradition, the question of how many genuinely distinct ways there are of gaining knowledge was vigorously contested. This is a direct analogue, within epistemology, to the debate within modern intelligence research about how many genuinely distinct cognitive abilities there are.

Nyāya: the science of reasoning

The Nyāya school, founded by the sage Gautama (c. 200 CE, though drawing on much older traditions of debate), is sometimes described as the closest ancient Indian equivalent to contemporary analytic philosophy. The Internet Encyclopedia of Philosophy notes that Nyāya developed a sophisticated precursor to contemporary reliabilism (the view that knowledge is produced by reliable cognitive processes) and that its methods of analysis and argument resolution influenced much of classical Indian literary criticism, philosophical debate, and jurisprudence.

For Nyāya, logic is subsumed within epistemology. The purpose of reasoning is not merely formal validity but the production of genuine knowledge. This means that Nyāya logic has a strongly cognitive and practical orientation. One makes inferences for oneself (to arrive at knowledge) and then, in a separate step, formalizes those inferences for others (to demonstrate knowledge in debate). The five-step syllogism is not just a logical form; it is a map of how actual human reasoning moves from observation through generalization to conclusion.

Nyāya also developed an elaborate theory of error (*khyāti-vāda*), analyzing how and why cognition goes wrong. This is directly relevant to the question of intelligence, because any tradition that takes seriously the structure of valid knowing must also take seriously the structure of invalid knowing: the ways in which perception is distorted, inference is fallacious, testimony is unreliable, and reasoning is led astray. The Nyāya analysis of fallacies is as detailed as anything in Aristotle's *Sophistical Refutations* and in some respects more systematic.

What makes Nyāya distinctive in this book's comparative framework is that it treats intelligence not as a trait of persons but as a quality of cognitive processes. The question is not "Is this person smart?" but "Did this cognition arise from a valid source, through a correct process, in appropriate conditions?" A person can have excellent reasoning capacity and still produce false beliefs if they rely on faulty testimony or make an inferential error. Conversely, a person of ordinary reasoning capacity can arrive at genuine knowledge if they use the right methods in the right conditions. This is, in a sense, an epistemological democratization: it shifts the emphasis from innate cognitive power to the quality of one's epistemic practices.

Buddhi and the broader Indian vocabulary of mind

While the *pramāṇa* framework is the most distinctive Indian contribution to the topics covered by this book, classical Indian philosophy also had a rich vocabulary for the faculties of the mind.

Buddhi () is usually translated as intellect, understanding, or discernment. In the Sāṃkhya philosophical system, *buddhi* is the first and highest evolute of primordial nature (*prakṛti*), the faculty

through which discrimination and decision occur. It is closer to Aristotle's *nous* (direct apprehension of truth) than to modern "intelligence" in the psychometric sense, but it operates within a metaphysical framework very different from Aristotle's.

Jñāna () is knowledge or cognition in the broadest sense. It can refer to everyday perceptual knowledge, philosophical understanding, or the liberating knowledge (*mokṣa-jñāna*) that, in many Indian traditions, frees the knower from the cycle of rebirth. This last usage is crucial: in much of Indian thought, the highest form of knowing is not theoretical mastery or practical skill but liberating insight. Knowledge is ultimately valuable not because it lets you solve puzzles or govern a state, but because it frees you from suffering.

Manas () is often translated as "mind" in the sense of the internal organ that coordinates sensory input. In Nyāya, *manas* is the inner sense that mediates between the external senses and the self (*ātman*), enabling apperception (awareness of one's own mental states). It is not the seat of intelligence in the modern sense but a processing intermediary.

This vocabulary matters because it shows that Indian thought, like Greek thought, differentiated the faculties of mind rather than collapsing them into a single "intelligence." But the Indian differentiations are different from the Greek ones. Where Aristotle distinguished theoretical from practical from productive reasoning, Indian thought distinguished perception from inference from testimony from analogy. Where Aristotle's framework is organized around the relation of knowledge to action, the Indian framework is organized around the sources and validity of knowledge itself.

Debate as recognition: the culture of argument

How was intellectual excellence recognized in classical India? Not primarily through written examinations (as in China) or public oratory (as in Rome), but through **debate**.

The Indian philosophical tradition developed an elaborate culture of structured argumentation. The Nyāya school formalized the rules of debate (*vāda-śāstra*), distinguishing between honest debate aimed at truth (*vāda*), debate aimed at winning by any means (*jalpa*), and debate aimed at destroying the opponent's position without establishing one's own (*vitaṇḍā*). These distinctions show a sophisticated awareness that argumentative skill and intellectual honesty are not the same thing, and that a culture of debate needs rules to ensure that cleverness serves truth rather than substituting for it.

In practice, philosophical debate in India functioned as a public test of intellectual competence. Scholars would challenge each other in open forums, sometimes before royal or religious audiences. The ability to defend one's position against objections, to identify fallacies in an opponent's rea-

soning, and to construct sound arguments under pressure was the primary marker of philosophical excellence. This is a recognition method that directly tests reasoning, knowledge, and argumentative composure, though it does so in a situated, interactive, and social context very different from a written examination.

The later Buddhist epistemological tradition (Dignāga, Dharmakīrti) further refined the standards of philosophical argument, introducing new criteria for what counts as a valid reason and developing sophisticated accounts of perception and inference that engaged directly with Nyāya positions. This cross-school dialogue, conducted over centuries through texts, commentaries, and face-to-face debate, represents one of the most sustained intellectual conversations in human history. It is a form of collective intelligence: the quality of any individual thinker's work was partly a product of the tradition of rigorous criticism within which they operated.

What Indian thought bundled together

Applying the four-layer template:

Vocabulary. *Pramāṇa* (source of valid knowledge), *pratyakṣa* (perception), *anumāna* (inference), *śabda* (testimony), *upamāna* (comparison), *arthāpatti* (postulation), *anupalabdhi* (non-apprehension), *buddhi* (intellect), *jñāna* (knowledge/cognition), *manas* (mind/inner sense), *mokṣa* (liberation), *vāda* (honest debate), *tarka* (suppositional reasoning).

Phenomena. Indian epistemology was tracking: the structure and sources of valid cognition, the conditions under which perception and inference produce genuine knowledge, the reliability of testimony, the formal structure of logical argument, the analysis of error, and the cognitive conditions for liberating insight. These map onto several of the dimensions this book charts in Movement II: reasoning and abstraction, retained knowledge, and (in the liberating-knowledge tradition) a form of wisdom very different from the psychometric conception.

Norms. Two normative strands are visible. One, dominant in the *pramāṇa* tradition, values epistemic rigor: the ability to distinguish valid from invalid cognition, to reason correctly, and to identify fallacies. This is the strand most congruent with modern analytical reasoning. The other, present across many Indian philosophical and religious traditions, values liberating knowledge: cognition that frees one from suffering and the cycle of rebirth. In this second strand, the highest form of intellectual excellence is not analytical cleverness but transformative insight.

Recognition. Philosophical debate was the primary recognition method. Public argumentation tested reasoning ability, doctrinal knowledge, and argumentative composure in real time. This was supplemented by the *guru-śiṣya* (teacher-student) tradition, in which a student's competence was val-

idated by a recognized teacher, and by the commentary tradition, in which a thinker demonstrated mastery by writing detailed analyses of foundational texts. All three methods are situated and relational: they depend on the judgment of other experts, not on a standardized, decontextualized score.

What Indian thought leaves aside

Honesty requires noting what the classical Indian epistemological tradition does not emphasize, at least in its most technical philosophical registers. Compared to the Chinese heart-mind tradition, it does not foreground the integration of cognition with emotion. Compared to Akan and Yoruba wisdom traditions, it does not center practical communal judgment. Compared to Greek *phronēsis*, it does not focus heavily on the capacity for situated practical decision-making (though the *pramāṇa* framework is relevant to action insofar as correct knowledge is a precondition for correct action).

This is not a deficiency. It is a specialization. The Indian tradition chose to investigate the structure of knowing with extraordinary rigor, and in doing so it developed tools and concepts (the *pramāṇa* framework, the five-step syllogism, the theory of error, the analysis of testimony, the distinction between perception and inference) that remain philosophically productive. But it is a reminder that no single tradition covers the entire space of human excellence. The Indian tradition illuminates the reasoning and epistemic-validity dimensions with unmatched precision, while other traditions illuminate other dimensions more fully.

What this means for the argument

The Indian case contributes three things to this book's larger project.

First, it shows that "intelligence" can be approached not as a property of persons but as a quality of cognitive processes. The *pramāṇa* framework asks whether a particular act of knowing arose from a valid source through a correct process. This is a fundamentally different orientation from psychometrics, which asks how much cognitive ability a person possesses. Both approaches are legitimate. But they are not the same, and a comprehensive account of human intellectual excellence needs both: an account of what persons can do and an account of what makes any particular cognitive act trustworthy.

Second, the Indian debate tradition provides a powerful alternative model of intellectual assessment. Testing reasoning ability through structured, interactive, public argumentation is in some respects more ecologically valid than testing it through written multiple-choice items, because it requires the reasoner to respond to objections in real time, to adjust their arguments based on feedback, and to demonstrate not just logical competence but intellectual composure and honesty. This is closer to

what modern assessment researchers call performance-based or authentic assessment than to what IQ tests do.

Third, the Indian tradition's emphasis on liberating knowledge introduces a normative dimension that most other traditions in this survey share but that modern psychometrics largely ignores: the question of what knowledge is *for*. For the Nyāya school, correct knowledge is valuable because it is a precondition for successful action and ultimately for liberation. For the Buddhist epistemologists, the entire enterprise of reasoning is embedded in a soteriological (salvation-oriented) framework. Knowledge matters because it frees you from suffering. This teleological orientation, the idea that intellectual excellence is not an end in itself but a means toward human flourishing or liberation, connects the Indian tradition to the Greek concern with *eudaimonia*, the Confucian concern with harmonious social life, and the Akan concern with communal wellbeing, even though the specific telos differs in each case.

Chapter 8: Buddhist Traditions. Wisdom, Compassion, Realization, and Debate

The previous chapter described how classical Indian philosophy asked what makes a cognition valid. Buddhist traditions, which emerged from and in dialogue with that Indian philosophical world, ask something more radical: What is wisdom, and what is it for?

The answer, across nearly all Buddhist schools, is striking in its structure: wisdom (*prajñā*) is not the accumulation of information or the mastery of logical technique, though both can support it. Wisdom is direct insight into the nature of reality. And in the Mahāyāna traditions that came to dominate East and Central Asian Buddhism, wisdom is inseparable from compassion (*karuṇā*). Not merely paired with it as a complementary virtue, but structurally inseparable: wisdom without compassion is not yet wisdom, and compassion without wisdom is not yet compassion. This inseparability is one of the most radical claims any tradition in this survey makes about the relationship between cognition and affect, and it has direct implications for the question of whether "intelligence" can be meaningfully defined without reference to how it is oriented toward other beings.

At the same time, Buddhist traditions, particularly in Tibet, developed one of the most rigorous institutional systems for training and testing intellectual excellence ever devised. The Geshe degree in Tibetan Gelug monasteries requires roughly twenty years of study, memorization, and daily public debate. It is as intellectually demanding as any doctoral program in the modern world and far longer. This combination, of a philosophical tradition that defines wisdom as inseparable from compassion and an institutional tradition that tests reasoning with extreme rigor, makes the Buddhist case uniquely instructive.

Prajñā: wisdom as insight, not as data

The Sanskrit term *prajñā* (Pali: *paññā*) is usually translated as "wisdom," but in Buddhist usage it has a meaning quite different from common English senses of the word. Pāli scholars T. W. Rhys Davids and William Stede defined *paññā* as "intelligence, comprising all the higher faculties of cognition." But the Buddhist meaning goes beyond cognitive faculties. *Prajñā* is the understanding of the true nature of phenomena: their impermanence (*anicca*), their tendency to produce suffering when grasped at (*dukkha*), and their lack of fixed, independent selfhood (*anattā*). In Mahāyāna Buddhism, *prajñā* is most often associated with the realization of *śūnyatā* (emptiness): the insight that all phenomena, including the self, lack intrinsic, independent existence.

This is not an intellectual proposition that one assents to after reading a book. It is a transformation of how one perceives and relates to reality. The fifth-century scholar Buddhaghosa wrote that wisdom

”penetrates into dharmas as they are in themselves. It disperses the darkness of delusion, which covers up the own-being of dharmas.” The Prajñāpāramitā sūtras (Perfection of Wisdom sūtras), composed between roughly 100 BCE and 500 CE, describe wisdom as nonconceptual realization: not a thought about emptiness but a direct seeing of it.

For this book’s purposes, the crucial point is that *prajñā* is defined not as a quantity of cognitive power but as a quality of understanding. One does not have ”more” or ”less” *prajñā* in the way one might have a higher or lower IQ score. One either sees the nature of reality or one does not (with many gradations and depths of seeing along the path). This is a qualitative, transformative, and ultimately liberating kind of knowing, not a psychometric variable. It is closer to what this book will later call wisdom (Chapter 28) than to what modern psychology calls intelligence.

The Yogācāra tradition further analyzed *prajñā* into three progressive forms: wisdom arising from hearing or learning (*śrutamayī-prajñā*), wisdom arising from reflection and analysis (*cintāmayī-prajñā*), and wisdom arising from meditative cultivation and direct realization (*bhāvanāmayī-prajñā*). This progression, from received knowledge through analytical understanding to experiential insight, is itself a model of intellectual development that has no equivalent in modern psychometric thinking. It suggests that the deepest form of knowing is not analytical but experiential, and that analysis is a necessary stage on the way to something that transcends analysis.

The inseparability of wisdom and compassion

The Mahāyāna Buddhist tradition makes a claim that no other tradition in this survey makes with equal force: that wisdom and compassion are not separate virtues that happen to be valued together, but are structurally inseparable aspects of a single awakened mind.

The logic of this claim runs as follows. If *prajñā* is the insight that all beings lack independent self-existence, and that all beings are constituted by their interdependence with other beings, then genuinely seeing this must transform how one relates to others. A person who truly understands that the boundaries between self and other are less solid than they appear cannot remain indifferent to the suffering of others. Compassion arises naturally from the realization of interdependence. Conversely, a person who merely feels compassion without understanding the deeper structure of reality will be prone to what Buddhist teachers call ”idiot compassion”: well-meaning interventions that are misguided because they are not informed by clear seeing.

As one Buddhist source puts it: ”Genuine wisdom, genuine Prajñā, is always inseparable from Karuṇā in the same amount. If we feel like our wisdom is much less than our compassion, then this compassion is not Karuṇā.” Speaking of wisdom and compassion as two separate things that

need to be developed independently is, in this framework, a pedagogical convenience, not a description of reality. In practice, they arise together or not at all.

This has profound implications for the central question of this book. If wisdom and compassion are inseparable, then any definition of "intelligence" that is purely cognitive, that treats intellectual excellence as independent of one's orientation toward other beings, is not just incomplete. It is structurally wrong. It has separated something that, in the deepest analysis, cannot be separated without distorting both halves. The Chinese *xin* tradition made a similar move by refusing to separate cognition from emotion, but the Buddhist version goes further: it refuses to separate understanding from ethical-emotional orientation toward the world.

This does not mean that every Buddhist tradition reduces to a single teaching. Theravāda traditions tend to treat *prajñā* and *karuṇā* as distinct (though complementary) qualities to be cultivated. Mahāyāna and Vajrayāna traditions increasingly stress their inseparability. The bodhisattva ideal, central to Mahāyāna, is a being who vows to attain full awakening not for personal liberation alone but for the benefit of all sentient beings. In this framework, wisdom without the bodhisattva motivation is incomplete, and compassion without wisdom is blind.

The pāramitā framework: perfections, not traits

In Mahāyāna Buddhism, the path to awakening is structured around the cultivation of *pāramitās* (perfections): qualities that the bodhisattva develops over vast stretches of time. The standard list of six *pāramitās* is: generosity (*dāna*), ethical conduct (*śīla*), patience (*kṣānti*), effort (*vīrya*), concentration (*dhyāna*), and wisdom (*prajñā*). In some traditions, four more are added: skillful means (*upāya*), aspiration, power, and primordial wisdom.

What is striking about this framework is that wisdom is not the sole or even the primary perfection. It is the sixth, the culminating one, but it depends on and integrates the other five. A bodhisattva who has wisdom but lacks generosity, ethical conduct, patience, effort, and concentration is not yet on the path. Excellence, in this framework, is an integrated profile of qualities, not a single dominant trait.

This is directly relevant to the argument of this book. The *pāramitā* framework is, in effect, a multi-dimensional model of human excellence. It recognizes that different capacities are required and that they must develop together. It is closer to the "profile" model that Movement III will advocate than to the single-score model of modern IQ testing. And it was developed more than two millennia before psychometrics.

Tibetan monastic education: rigor at its most demanding

If the philosophical side of Buddhism stresses wisdom as transformative insight, the institutional side, especially in the Tibetan Gelug tradition, stresses intellectual rigor with an intensity that rivals any educational system in human history.

The Geshe degree (*dge bshes*, "virtuous friend") is the highest academic degree in the Gelug school. The curriculum derives from the Indian Buddhist monastic universities, particularly Nālandā, which flourished from roughly the seventh to twelfth centuries before their destruction by invaders. The Tibetan tradition continued and systematized this inheritance.

The Gelug Geshe curriculum centers on five major Indian Buddhist texts, studied through memorization, commentary, and dialectical debate over approximately twenty years. The five subjects are: *Pramāṇa* (logic and epistemology, based on Dharmakīrti's *Commentary on Valid Cognition*), *Prajñāpāramitā* (the perfection of wisdom, based on Maitreya's *Ornament of Realization*), *Madhyamaka* (the philosophy of emptiness, based on Chandrakīrti's *Supplement to the Middle Way*), *Abhidharma* (phenomenology and cosmology, based on Vasubandhu's *Treasury of Abhidharma*), and *Vinaya* (monastic discipline, based on Guṇaprabha's *Root Text on Vinaya*).

The pedagogical method is debate (*rtsod pa*). Monks engage in structured dialectical exchanges, with one serving as questioner and the other as defender. The questioner puts forth challenges, attempting to lead the defender into contradiction. The defender must maintain a thesis against all objections. Debate sessions occur daily, often lasting several hours, with gestures, clapping, and vocal exclamations adding a physical and theatrical dimension to the intellectual exchange. Georges Dreyfus, the first Westerner to earn the Geshe Lharampa degree (the highest rank), described the Tibetan monastic debate tradition as a sophisticated intellectual culture that has been at the center of traditional Tibetan life for centuries.

For the book's argument, several features of this system are important.

First, the Geshe degree is genuinely difficult to earn. The Lharampa degree requires roughly twenty years of study, with daily debate sessions, memorization of enormous bodies of text, and final examinations in which the abbot selects debate topics on the spot. Students cannot prepare for specific questions; they must command the entire curriculum deeply enough to respond to any challenge. This is a real test of integrated understanding, not memorization alone.

Second, the first subject in the curriculum is *pramāṇa*, logic and epistemology. This is a direct continuation of the Indian tradition discussed in Chapter 7. Dharmakīrti's epistemology, which systematized Buddhist responses to Nyāya, became the foundational intellectual language of Tibetan scholasticism. A Geshe begins by learning to reason precisely, and only then proceeds to the philo-

sophical content that reasoning will be applied to.

Third, the system integrates intellectual rigor with ethical and contemplative training. The Geshe is not merely a philosopher. He (and now she, since the Geshema degree was first awarded to women in 2016) is a monastic who lives under ethical discipline (*vinaya*), engages in contemplative practice, and pursues intellectual excellence within a framework that treats all of these as mutually supporting. This is the most extreme institutional case of what the book has been calling a "thick" tradition: one that refuses to separate cognitive skill from moral formation and contemplative practice.

Fourth, after completing the sutra curriculum and earning the Geshe degree, many graduates proceed to one of the tantric colleges for further study. The educational trajectory does not end with logical and philosophical mastery; it continues into contemplative and ritual domains. This is a concrete institutional embodiment of the three-stage wisdom model: learning, reflection, and meditative realization.

What Buddhist thought bundled together

Applying the four-layer template:

Vocabulary. *Prajñā* (wisdom/insight), *karuṇā* (compassion), *pāramitā* (perfection), *śūnyatā* (emptiness), *bodhicitta* (the awakening mind, the aspiration to attain enlightenment for all beings), *upāya* (skillful means), *śīla* (ethical conduct), *samādhi/dhyāna* (concentration/meditation), *pramāṇa* (valid cognition, in the Tibetan scholastic context), *tsod pa* (debate).

Phenomena. Buddhist traditions were tracking: transformative insight into the nature of reality, the integration of wisdom with compassionate orientation, ethical self-discipline, concentrated attention, logical and epistemological rigor, embodied debating skill, memorization and textual mastery, and the progressive development of qualities (the *pāramitā* framework). These map onto several dimensions in Movement II: wisdom, emotional attunement (through the compassion pairing), self-regulation (through meditation and ethical discipline), reasoning (through the *pramāṇa* and debate traditions), and moral discernment (through the *bodhisattva* ideal).

Norms. The highest form of excellence is awakening: the full realization of the nature of reality, inseparable from compassion for all beings. This is not cognitive achievement in any ordinary sense. It is a transformation of the entire person. Intellectual rigor is valued, but it is valued as a means, not as an end. The end is liberation and the capacity to benefit others. This teleological framing connects Buddhism to the Indian concern with *mokṣa*, to the Confucian concern with harmonious relational life, and to the Greek concern with *eudaimonia*, while differing from all of them in its specific content.

Recognition. In the Tibetan Gelug tradition, intellectual excellence is validated through the Geshe degree: approximately twenty years of study, daily debate, memorization, and final examinations before the monastic community. This is supplemented by teacher-student lineages, contemplative attainment recognized by one's teacher, and in some traditions specific markers of realization that are assessed by experienced practitioners. The system combines formal examination, public dialectical performance, and teacher validation in a way that is more multi-modal than either the Chinese imperial examination or modern IQ testing.

What this means for the argument

The Buddhist case adds three things to the book's project that no previous chapter has provided.

First, the inseparability of wisdom and compassion is the strongest argument in the survey so far against defining intelligence as a value-neutral cognitive trait. If the deepest form of understanding is structurally bound to how one relates to other beings, then treating intelligence as independent of moral and emotional orientation is not just a narrowing. It is a mistake about the nature of knowing itself. This is a more radical claim than the Chinese *xin* argument (which says cognition and emotion are aspects of one faculty) or the Greek *phronēsis* argument (which says practical wisdom requires moral virtue). It says that the very content of wisdom changes depending on whether it is oriented toward the welfare of others.

Second, the *pāramitā* framework is the clearest pre-modern articulation of a multi-dimensional model of human excellence. It explicitly names six (or ten) distinct qualities that must be cultivated together, and it treats wisdom as the culmination of a profile, not as a standalone capacity. This is a direct historical precedent for the dimensional framework that Movement III will propose.

Third, the Tibetan Geshe system demonstrates that extreme intellectual rigor and broad human formation are not incompatible. A tradition can demand twenty years of daily debate, precise logical analysis, and mastery of enormous textual corpora while simultaneously insisting that all of this is embedded in ethical discipline, contemplative practice, and the aspiration to benefit all beings. The modern assumption that rigor requires narrowness, that you can only test cognitive ability precisely if you strip away everything else, is contradicted by the existence of this system.

Chapter 9: Islamic and Persianate Traditions. Intellect, Wisdom, Cultivation, and Authority

The Islamic intellectual tradition presents a distinctive case in this survey. It is the only tradition covered so far that inherited two older traditions explicitly and self-consciously: the Greek philosophical heritage, which it received, translated, and transformed during the great translation movement of the eighth through tenth centuries, and the Qur'anic and Prophetic tradition, which provided its religious, ethical, and legal foundations. The result was not simply a blend but a creative synthesis that produced new concepts of intellect, new models of the cultivated person, and a unique system for authorizing the transmission of knowledge.

Three concepts anchor this chapter: *ʿaql* ,) intellect or reason), which the philosophers elaborated into a sophisticated metaphysics of mind; *adab* ,) cultivated comportment), which linked intellectual formation to ethical conduct, literary skill, and social refinement; and *ijazah* ,) authorization), which created a distinctive recognition system based on personal chains of teacher-to-student transmission rather than institutional examination.

ʿAql: intellect within a science of the soul

The Arabic word *ʿaql* is usually translated as "intellect" or "reason," but its range of meaning is broader than either English term suggests. In common usage, it conveyed something like sound judgment or discernment: the capacity to distinguish virtuous from vicious conduct through deliberation. Al-Fārābī, the philosopher known as the "Second Teacher" (after Aristotle), noted in his *Risālat fī al-ʿaql* (Epistle on the Intellect) that ordinary people used the term to describe the capacity that enables a person to discern which deeds are virtuous and which are not. This common-sense meaning is close to Aristotle's *phronēsis*, and it is no coincidence: al-Fārābī was working directly from Aristotle's texts.

But the philosophical tradition (*falsafa*) took *ʿaql* much further. Al-Fārābī classified the intellect into multiple types: potential intellect (the capacity to receive knowledge), actual intellect (the intellect in the act of knowing), acquired intellect (the intellect enriched by accumulated knowledge), and the Active Intellect (*ʿaql al-faʿāl*), a cosmic principle that actualizes human thought and connects individual minds to the intelligible order of reality. This hierarchy of intellects was not merely psychological; it was cosmological. Al-Fārābī situated the human intellect within a system of ten emanated intellects linking the divine to the material world. Human intellectual development, on this model, is a process of actualizing latent potential under the influence of a transcendent source

of intelligibility.

Ibn Sīnā (Avicenna, d. 1037), who built on al-Fārābī's work with extraordinary originality, developed this further. For Ibn Sīnā, the human intellect begins as a blank potentiality (*ʿaql hayūlānī*, material intellect) and develops through stages toward conjunction with the Active Intellect. Knowledge is not simply abstracted from sensory experience, as in the standard Aristotelian model; it is illuminated by the Active Intellect, which bestows intelligible forms on the prepared mind. The philosopher reaches truth through disciplined intellectual development. The prophet receives truth through a sudden effusion of wisdom from the Active Intellect, without needing to follow the progressive stages of philosophical training.

For this book's argument, several features of the *falsafa* tradition are significant.

First, it placed the development of the intellect at the center of human perfection. Al-Fārābī's *Enumeration of the Sciences* organized all knowledge into a comprehensive system, from linguistics and logic through mathematics and natural science to metaphysics and political science. The philosophical life was a life of progressive intellectual actualization, and the highest human achievement was the full development of the intellect to the point where it could grasp the most fundamental truths of reality. This is, in some ways, the most intellectualist tradition in the survey: it places the intellect (not practical judgment, not emotional attunement, not communal harmony) at the pinnacle of human excellence.

Second, however, this intellectualism was never fully separated from ethical and political concerns. Al-Fārābī was also the father of Islamic political philosophy, and his model of the ideal state was governed by a philosopher-king whose intellectual perfection qualified him to guide the community. Ibn Sīnā's theory of prophecy explicitly connected intellectual illumination with moral and political leadership. The intellect, in this tradition, is not a private cognitive engine. It is the faculty through which a person becomes fit to lead, to govern, and to guide others toward truth. This mirrors, in a different key, the Greek and Roman insistence that intellectual excellence is inseparable from civic responsibility.

Third, the *falsafa* tradition provoked a major counter-response from within Islamic thought. Al-Ghazālī (d. 1111), in his *Tahāfut al-Falāsifa* (The Incoherence of the Philosophers), attacked the philosophers' claims on multiple fronts, arguing that their metaphysical speculations could not deliver the certainty they promised and that genuine knowledge of ultimate realities came through religious experience and divine revelation, not through philosophical reasoning alone. This debate, between rationalist philosophy and experientially grounded religious knowledge, has structural parallels with the modern tension between psychometric measurement and contemplative or experiential accounts of wisdom. The Islamic tradition did not resolve this tension; it lived within it, producing

both great philosophers and great mystics, often in the same person.

Adab: the cultivated person as moral-literary ideal

If *‘aql* names the highest cognitive faculty, *adab* names the broader formation of the cultivated person. The term evolved across centuries. In pre-Islamic Arabia, it meant custom, habit, or norm of conduct. By the ninth century, it had expanded to encompass the entire range of qualities expected of a civilized and cultured Muslim: proper conduct, literary knowledge, rhetorical skill, poetic sensibility, historical awareness, and ethical refinement. The *adīb* (the person of *adab*) was someone with knowledge of history, poetry, ideas, proverbs, parallels, precedents, and the correct and pleasing use of language.

Adab is deeply relevant to this book because it represents an Islamic version of the thick formation ideal that we have already encountered in Greek *paideia*, Roman *humanitas*, and Confucian self-cultivation. Like those traditions, *adab* refuses to separate intellectual competence from moral character and social refinement. The learned person is not merely knowledgeable; they are well-mannered, ethically formed, socially graceful, and literarily accomplished. *Adab* is both knowledge and the embodiment of that knowledge in conduct. As the Encyclopedia Iranica notes, *adab* indicated a set of rules comprising practical ethics and the educational elements needed by a person who wanted to behave appropriately in all circumstances of life.

The *adab* tradition produced a vast literature: manuals of conduct for rulers, judges, scribes, physicians, and ordinary Muslims; anthologies of poetry, prose, and historical anecdotes organized for moral and literary edification; and works of fiction that provided moral exemplars within their stories. This literature served as a vehicle for the transmission of cultural ideals, much as Homer served in Greek *paideia* and the Confucian classics served in the *keju* system. The difference is that *adab* literature was more explicitly normative about conduct and manners, and more directly linked to professional competence. A good scribe, in the *adab* tradition, was not merely accurate; he was ethically formed, literarily skilled, and socially refined.

For Islamic ethicists like al-Ghazālī and al-Iṣfahānī, *adab* was the practical dimension of moral formation. Al-Ghazālī treated the cultivation of proper etiquette as essential for building character, and character as the foundation for genuine knowledge of God. The student’s manners toward the teacher, the conduct of scholarly discussion, the governance of one’s own appetites and emotions, these were not peripheral niceties but central practices through which virtue was cultivated and knowledge became transformative.

What *adab* bundles together, then, is literary knowledge, ethical conduct, social skill, emotional

composure, and professional competence. It is an integrative ideal, not a measurable trait. And it was assessed not by examination but by the quality of one's writing, speech, conduct, and reputation within the scholarly and administrative community.

Ijazah: knowledge authorized by persons, not by institutions

The most distinctive contribution of the Islamic tradition to the history of recognizing intellectual excellence is the *ijazah* system. An *ijazah* is a license authorizing its holder to transmit a certain text or body of knowledge, issued by someone who already possesses such authority. It is personal rather than institutional: the authorization comes from a specific teacher, not from a university or a state, and it carries the teacher's personal endorsement that the student has achieved sufficient mastery to teach and transmit the material to others.

The *ijazah* is typically accompanied by a *sanad* (chain of transmission), which traces the authority of the certification back through a succession of teachers to the original source, often the Prophet Muhammad himself in the case of Qur'anic recitation and hadith, or to the author of a text in the case of philosophical, legal, or literary works. This chain of transmission is not merely decorative. It is the *epistemic warrant* for the knowledge: the student's claim to legitimate knowledge rests on the fact that they received it face-to-face from a teacher who received it from their teacher, in an unbroken chain.

As Harvard scholar William A. Graham explained, this system of personal rather than institutional certification has served not only for hadith but also for transmission of texts of any kind, from history, law, or philology to literature, mysticism, or theology. The *ijazah* ideally reflects oral, face-to-face, teacher-to-student transmission validated by the teacher's personal authorization.

For this book's argument, the *ijazah* system is significant in several ways.

First, it is a recognition method that is fundamentally relational. Competence is validated not by an impersonal test but by a person who knows the student intimately and can attest to the quality of their understanding. This is closer to how apprenticeship-based traditions (Pacific wayfinding, craft guilds, medical residencies) validate competence than to how standardized tests do.

Second, the *ijazah* system places enormous weight on testimony and trust. The chain of transmission is only as strong as its weakest link, so the reliability and character of each transmitter matters. This connects directly to the Indian *pramāṇa* tradition's treatment of testimony (*śabda*) as a distinct source of knowledge: in both traditions, knowledge received from a trustworthy person is treated as a legitimate epistemic category, not a lesser substitute for direct observation or reasoning.

Third, the *ijazah* system was remarkably inclusive for its time. It was not limited by gender, age,

or social class in principle. Female hadith scholars like Zaynab bint al-Kamāl were known to earn *ijazahs* and instruct students. Children of scholarly families sometimes received *ijazahs* at young ages. The system was based on demonstrated competence, not on formal institutional enrollment.

Fourth, and most importantly for the comparative argument, the *ijazah* represents a completely different model of what it means to "measure" intellectual competence than either the Chinese imperial examination or modern standardized testing. The Chinese model is institutional, written, competitive, and anonymous (candidates were identified by number rather than name). The *ijazah* model is personal, oral, relational, and named (the specific teacher-student relationship is the foundation of the authorization). Modern psychometric testing is standardized, decontextualized, and quantitative. All three are genuine measurement traditions. They are radically different in form, and the differences reveal different assumptions about what knowledge is and how competence can be validated.

What Islamic thought bundled together

Applying the four-layer template:

Vocabulary. *ʿAql* (intellect/reason), *ḥikma* (wisdom), *adab* (cultivated comportment, literary refinement, ethics of conduct), *ʿilm* (knowledge), *ijazah* (authorization to transmit), *sanad* (chain of transmission), *falsafa* (philosophy in the Greek-influenced sense), *akhlāq* (ethics, moral character), *nafs* (soul), *ʿaql al-faʿcāl* (Active Intellect).

Phenomena. The Islamic tradition was tracking: the progressive development of the intellect toward truth, the moral-literary formation of the cultivated person, the reliability of knowledge transmission through personal chains of authority, the relationship between philosophical reasoning and religious revelation, and the integration of intellectual competence with ethical conduct and social refinement. These map onto several dimensions in Movement II: reasoning and abstraction (through the *falsafa* tradition), retained knowledge (through the *ʿilm* and *ijazah* traditions), moral discernment (through *akhlāq* and *adab*), and social-relational competence (through the *adab* ideal of appropriate conduct in all circumstances).

Norms. Two normative strands are visible. The philosophical strand values intellectual actualization as the highest human achievement, with the fully developed intellect approaching the Active Intellect. The *adab* strand values integrated moral-literary-social formation, treating intellectual competence as inseparable from ethical refinement and proper conduct. The *ijazah* tradition adds a third normative element: knowledge is legitimate only when it can be traced to a trustworthy source through an unbroken chain of personal transmission. All three strands resist the modern separation of cognitive ability from character, conduct, and relational accountability.

Recognition. The primary recognition methods are the *ijazah* (personal authorization from a qualified teacher), the *sanad* (chain of transmission validating the authorization), and the judgment of scholarly communities. The *falsafa* tradition also recognized intellectual excellence through the production of philosophical treatises and commentaries. The *adab* tradition recognized it through literary production, rhetorical skill, and social reputation. None of these is a standardized test in the modern sense, but all are genuine methods for selecting, validating, and certifying competence.

What this means for the argument

The Islamic case contributes three distinctive elements to the book's project.

First, it shows what happens when a tradition self-consciously inherits and transforms another tradition's philosophical vocabulary. The Islamic philosophers did not merely repeat Aristotle; they embedded his analysis of the intellect within a cosmological framework that made the development of the intellect a spiritual and soteriological project, not just a cognitive one. When al-Fārābī classified the intellect into four types and situated them within a hierarchy of emanated intellects, he was doing something genuinely new: connecting Greek epistemology to a Neoplatonic-Islamic cosmology. This demonstrates that intellectual traditions do not merely "have concepts of intelligence." They actively construct, revise, and transform them in response to their own philosophical and religious commitments.

Second, the *adab* tradition provides the strongest Islamic analogue to what this book has been calling thick formation. Like Greek *paideia*, Roman *humanitas*, Confucian self-cultivation, and Buddhist integrated training, *adab* refuses to separate what you know from how you behave, how you speak from how you treat others, intellectual competence from moral character. This refusal is so consistent across traditions that it is beginning to look less like a cultural preference and more like a structural insight about human excellence that the modern psychometric tradition chose to abandon.

Third, the *ijazah* system introduces a model of knowledge validation that is entirely absent from modern psychometrics: personal, relational, chain-based authorization. Knowledge is legitimate because a trustworthy person transmitted it to you, and they received it from a trustworthy person, and so on. This is not a primitive system awaiting replacement by standardized testing. It is a distinct epistemic architecture, built on fundamentally different assumptions about what makes knowledge reliable. Modern science operates on a version of this logic when it values peer review, replication, and mentorship. The *ijazah* makes the logic explicit and traces it through named individuals across centuries.

Chapter 10: Akan, Yoruba, and Related African Traditions. Wisdom, Proverb, Character, and Conduct

With this chapter, the survey crosses a significant threshold. The traditions discussed so far, Greek, Roman, Chinese, Indian, Buddhist, and Islamic, all produced large bodies of written philosophical text. They are well represented in global academic philosophy, and their concepts can be studied through canonical works that have been translated, annotated, and debated for generations. The traditions discussed in this chapter and the next three are different. They were transmitted primarily through oral means: proverbs, stories, songs, ritual speech, adjudication, apprenticeship, and the living example of recognized elders and sages. Their philosophical content is no less sophisticated for being oral. But it requires a different kind of interpretive attention, and the book must be honest about the challenges this poses.

The Akan and Yoruba traditions of West Africa offer accounts of human excellence that are among the most directly relevant to this book's central argument. They define wisdom not as abstract reasoning but as the ability to make good use of what one knows, demonstrated in speech, conduct, judgment, and the maintenance of communal harmony. They refuse to separate intellectual capacity from moral character. And they provide, in the concepts of *nyansa* and *omolúàbí*, ideals of the excellent person that are as philosophically precise and as practically demanding as anything in Aristotle or Confucius.

Akan *nyansa*: wisdom as the good use of knowledge

The Akan people of Ghana, one of West Africa's largest ethnic groups, have a philosophical tradition that has been the subject of significant scholarly attention, particularly through the work of Kwame Gyekye and Kwasi Wiredu, two of the most important figures in modern African philosophy.

The Akan concept most relevant to this book is *nyansa*, which is usually translated as "wisdom." But the translation, as always, risks assimilating the concept to its English equivalent. In Akan thought, *nyansa* is not primarily theoretical knowledge or abstract reasoning capacity. It is the ability to make good practical use of what one knows. The wise person is not the one who knows the most facts or reasons the most quickly, but the one who applies knowledge well: who speaks appropriately, judges soundly, acts with discernment, and contributes to the harmony of communal life.

This emphasis on application rather than accumulation is captured in Akan proverbs, which function not as decorative sayings but as vehicles of condensed philosophical reasoning. Proverbs are used in adjudication, in counsel, in education, and in daily discourse. A person's ability to deploy proverbs

aply, to select the right proverb for the situation and apply it with timing and judgment, is itself a recognized mark of wisdom. This is not a trivial skill. It requires extensive knowledge of the proverbial corpus, understanding of the situation at hand, judgment about which principle applies, and the social intelligence to know when and how to invoke it. It is, in effect, a form of case-based practical reasoning that parallels Aristotelian *phronēsis* in its structure even though it operates through a completely different cultural medium.

The Akan *nyansapo* (wisdom knot), one of the Adinkra symbols that encode philosophical concepts in visual form, represents the integration of knowledge with its application. The symbol itself embodies the idea that wisdom is not a linear accumulation but a complex knot: a pattern of interconnected judgments that cannot be reduced to a simple formula. Education scholars have used *nyansapo* as a framework for thinking about indigenous Akan philosophies of education, arguing that Akan educational ideals center on producing not merely knowledgeable individuals but wise ones, people capable of sound judgment in the practical circumstances of communal life.

What *nyansa* bundles together, then, is knowledge, practical judgment, speech competence, social perception, and moral character. A person who knows many things but uses them badly, who speaks clumsily, who judges poorly, or who acts against the community's interests does not have *nyansa*. This is a thick concept, and its thickness is philosophically significant. It represents an independent arrival at the conclusion, shared by the Greek, Chinese, and Islamic traditions, that intellectual excellence without moral-practical integration is not excellence at all.

Yoruba *ogbón* and the *omolúàbí* ideal

The Yoruba people, numbering over forty million across West Africa and the diaspora, possess one of the most philosophically elaborated oral traditions in the world. At its center is the Ifá divination system, a vast corpus of oral literature preserved and transmitted by the *babaláwo* (father of secrets, the divination priest). Ifá contains poetry, narrative, moral instruction, cosmological theory, and practical wisdom organized into 256 *odù* (chapters), each containing multiple verses. It is, in effect, a library encoded in human memory and transmitted through oral practice.

The Yoruba concept of wisdom is *ogbón*, and it has dimensions that anticipate modern debates about emotional intelligence in striking ways. Some contemporary Yoruba scholarship distinguishes between *ogbón-ori* (intellectual wisdom, wisdom of the head) and *ogbón-inú* (emotional or interior wisdom, wisdom of the belly/inner being). As the Ifá tradition puts it, the person who fails to make use of their *ogbón-inú* becomes a fool, regardless of their intellectual knowledge. This distinction between a cognitive and an affective dimension of wisdom, made within an oral philosophical tradition, is a direct parallel to the modern distinction between cognitive intelligence and emotional

intelligence, arrived at independently and expressed in a completely different cultural register.

The concept of *ori* (head) adds further depth. In Yoruba metaphysics, *ori* is both a physical organ and a spiritual entity: it is the seat of destiny, consciousness, and personal identity. The quality of one's *ori* shapes one's capacity for wisdom, success, and good judgment. This is not a crude determinism; the Ifá tradition teaches that one's *ori* can be strengthened through proper conduct, ritual, and moral effort. But it means that Yoruba thought locates wisdom not in a detachable cognitive module but in the whole person, including their spiritual constitution, their destiny, and their relationship to the divine.

The ideal of human excellence in Yoruba culture is captured in the concept of *omolúàbí*. The word itself encodes its meaning: *omo* (child) + *ti* (that) + *olu-ìwà* (chief/master of character) + *bí* (born). An *omolúàbí* is literally a child born of the master of character: a person whose entire formation has been oriented toward moral excellence. The end of Yoruba traditional education, as scholars have noted, is to make every individual an *omolúàbí*. To be an *omolúàbí* is to be of good character.

The virtues that define the *omolúàbí* are enumerated with philosophical precision. They include *ìwà pẹ̀lẹ̀* (good character, gentle comportment), *ọ̀rọ̀ síṣọ* (proper speech, the Yoruba accord great respect for intelligent and expert use of language), *ìtẹ́ríba* (respect), *inú rere* (goodwill, having a good mind toward others), *ọ̀títọ́* (truth), *isẹ́* (hard work), *àkínkanjú* (courage), and *ọgbón* (wisdom/intelligence). Notice that *ọgbón* (intelligence) is one virtue among many, not the master category. The *omolúàbí* is intelligent, but intelligence alone does not make an *omolúàbí*. What makes one is the full integration of intelligence with character, speech, social conduct, courage, and hard work.

The parallels with other traditions covered in this book are extensive and cannot be coincidental. The *omolúàbí* is structurally similar to Aristotle's *phronimos* (the person of practical wisdom), to the Confucian *junzi* (the exemplary person), and to the Islamic *adīb* (the person of cultivated comportment). All four ideals integrate cognitive excellence with moral character, social skill, and practical competence. All four resist reducing human excellence to a single cognitive dimension. And all four arose independently, in civilizations that developed without direct contact with each other. This convergence is one of the strongest pieces of evidence in the book for the claim that the modern psychometric narrowing is a departure from a near-universal human insight, not a refinement of it.

Proverbs as philosophy

A methodological note is necessary here. It is tempting for scholars trained in the Western literate tradition to treat proverbs as "folk wisdom," something less rigorous and less philosophical than systematic treatises. This temptation must be resisted.

Akan and Yoruba proverbs are not casual sayings. They are carefully constructed, memorized, and transmitted formulations of general principles that are applied to particular situations through judgment. The ability to use proverbs well requires knowledge of the corpus, understanding of the principles encoded in each proverb, perception of the relevant features of the current situation, and judgment about which principle applies. This is a form of reasoning that is analogical, contextual, and practical. It is closer to what legal reasoning does with precedents than to what formal logic does with syllogisms. And it is no less cognitively demanding for being oral rather than written.

Gyekye argued explicitly that Akan proverbs should be treated as reflections of individual philosophical thought within a collective context, not as mindless repetitions of communal platitudes. Proverbs are authored, even if the author is usually anonymous. They encode real insights about human nature, social dynamics, moral principles, and practical reasoning. And their deployment in context requires exactly the kind of situated practical judgment that this book argues is one of the most universally valued dimensions of human excellence.

Oruka's sage philosophy project

The Kenyan philosopher Henry Odera Oruka (1944-1995) undertook a project of direct relevance to this chapter. He sought to demonstrate that individual critical philosophical thinking existed in traditional African communities by identifying and interviewing people recognized by their communities as sages. His "philosophic sagacity" project involved recording dialogues with these sages, documenting their reflections on fundamental philosophical questions about knowledge, reality, justice, and the good life.

Oruka's project is important for this book because it challenges the assumption that philosophy requires literacy. The sages he interviewed were not academics. They were farmers, elders, and community leaders who had spent decades thinking carefully about fundamental questions and who were recognized by their communities for the depth and originality of their reflection. Their philosophical activity was not less rigorous for being oral, conversational, and embedded in practical life rather than published in academic journals.

The sage philosophy project also demonstrates that the Akan and Yoruba traditions are not unique cases. Across sub-Saharan Africa, communities have recognized individuals of exceptional wisdom, judgment, and reflective capacity. The methods of recognition differ from modern testing (they depend on community judgment, track records of good advice, and demonstrated practical wisdom over decades), but they are genuine methods of identifying and validating intellectual excellence.

What these traditions bundled together

Applying the four-layer template:

Vocabulary. Akan: *nyansa* (wisdom), *nyansapo* (wisdom knot), *onipa* (person), *suban* (character/disposition). Yoruba: *ogbón* (wisdom/intelligence), *ogbón-ori* (intellectual wisdom), *ogbón-inú* (emotional/interior wisdom), *ori* (head/destiny/consciousness), *ìwà* (character), *ìwà pẹ̀lẹ̀* (good character), *omolúàbí* (ideal person of character), *Ifá* (divination corpus and wisdom tradition), *babaláwo* (father of secrets, divination priest).

Phenomena. These traditions were tracking: practical judgment (the ability to apply knowledge well in context), moral character (the integrated virtues that define a good person), speech competence (the ability to use language wisely, including proverb deployment), emotional-interior wisdom (the affective dimension of good judgment), social harmony (the capacity to contribute to communal wellbeing), and the integration of all these into a single ideal of the excellent human being. These map onto multiple dimensions in Movement II: practical judgment, emotional attunement, social-relational competence, moral discernment, and wisdom.

Norms. Excellence is defined not by cognitive performance in isolation but by the full integration of knowledge, judgment, character, speech, and social conduct. The *omolúàbí* ideal treats intelligence (*ogbón*) as one virtue among several, not as the master category. What is most valued is the whole person: someone who knows well, speaks well, judges well, acts well, and lives in harmony with their community. This is among the thickest normative frameworks in the survey.

Recognition. Wisdom is recognized through community judgment over time: through the quality of one's speech in councils and adjudications, the soundness of one's advice, the aptness of one's proverb use, the track record of one's practical decisions, and the testimony of those who have observed one's conduct over years and decades. The Ifá system provides an additional layer: the *babaláwo* is trained through long apprenticeship and recognized by the quality of their divination and counsel. There is no written examination, no standardized test, and no single moment of certification. Recognition is earned through sustained performance in the contexts that matter to the community.

What this means for the argument

The West African traditions covered in this chapter contribute four things to the book's project.

First, they demonstrate that the practical-judgment dimension of excellence, the capacity to apply knowledge well in real-world situations, is valued independently across civilizations that had no contact with each other. Aristotelian *phronēsis*, Confucian cultivated judgment, and Akan *nyansa*

are not the same concept, but they are pointing at overlapping phenomena. This convergence is strong evidence that practical judgment is a real and important dimension of human excellence, not just a Western philosophical construct.

Second, the Yoruba distinction between *ogbón-ori* (intellectual wisdom) and *ogbón-inú* (emotional/interior wisdom) shows that the modern invention of "emotional intelligence" as a separate construct has a precedent in a tradition that drew the same distinction centuries earlier. The Yoruba case suggests that the cognitive-emotional split in modern Western psychology is not the only way to handle this distinction, and that treating emotional wisdom as a dimension of wisdom itself (rather than as a separate construct called "emotional intelligence") may be more natural and more accurate.

Third, the *omolúàbí* ideal, with its explicit enumeration of virtues (character, speech, respect, goodwill, truth, hard work, courage, wisdom), provides the most detailed indigenous profile model of human excellence in the survey so far. It is a multi-dimensional ideal that treats intelligence as one dimension among several, integrated with and subordinate to character as a whole. This directly anticipates the dimensional framework proposed in Movement III.

Fourth, the proverb tradition demonstrates that philosophical reasoning can be conducted in oral, analogical, and contextual modes rather than in written, syllogistic, and abstract modes, and that this difference in medium does not imply a difference in cognitive sophistication. If anything, the ability to select, from a vast corpus of memorized principles, exactly the right one for this particular situation, and to apply it with appropriate timing and social sensitivity, is a more demanding cognitive task than solving a well-defined problem on a written test. The oral medium makes the reasoning less visible to outsiders, but it does not make it less real.

Chapter 11: Ubuntu / Hunhu / Botho. Personhood Through Others

Every tradition covered in this book so far has shared one assumption, even as they disagreed about nearly everything else: they assumed that intellectual excellence is a property of individual persons. The Greek *phronimos*, the Confucian *junzi*, the Yoruba *omolúàbí*, the Indian philosopher, the Buddhist sage: all are individuals whose excellence is located within them, even when that excellence is defined in relational or communal terms. The traditions disagreed about what excellence consists of, how it should be cultivated, and how it should be recognized. But they all started with the individual as the unit of analysis.

The Ubuntu tradition of Southern Africa challenges this assumption at its root. *Umuntu ngumuntu ngabantu*: a person is a person through other persons. This is not a slogan. It is a philosophical claim about the nature of personhood itself, and its implications for the question of human excellence are profound. If personhood is constituted through relationships, then excellence cannot be fully located in any individual taken in isolation. The unit of analysis must include the person *and* the relational web through which they become who they are.

This chapter is shorter than the preceding ones because Ubuntu's contribution to the survey is more focused. It does not offer a detailed taxonomy of intellectual virtues or a new theory of knowledge. It offers something more radical: a challenge to the individualist framework within which all theories of intelligence, Western and non-Western alike, have been constructed.

What Ubuntu claims

Ubuntu (Zulu, Xhosa), Hunhu (Shona), and Botho (Sotho, Tswana) are cognate terms found across the Bantu-speaking peoples of Southern Africa. The word itself encodes its meaning: the prefix *ubu-* signals an abstract quality or state of being, and *-ntu* refers to a person or being. Ubuntu is thus the quality or condition of being human. But in this philosophical tradition, being human is not a biological given. It is an achievement, an ongoing process of becoming that is realized through one's relationships with others.

The philosopher Mogobe Ramose developed Ubuntu as a philosophical system spanning metaphysics, epistemology, and ethics, positioning it as foundational to African thought. For Ramose, Ubuntu expresses an ontology in which being is becoming and personhood is inherently relational and dynamic. Community is not a collection of pre-formed individuals who choose to cooperate. Community is the matrix within which individuals come into being as persons. As Augustine Shutte, an influential scholar of Ubuntu, put it: the moral life is seen as a process of personal growth, and

our deepest moral obligation is to become more fully human, which means entering more and more deeply into community with others.

John Mbiti articulated this most memorably: "I am because we are, and since we are, therefore I am." This inverts the Cartesian formula. Where Descartes grounded certainty in the isolated thinking subject (*cogito ergo sum*, I think therefore I am), Mbiti grounds existence in communal belonging. The self is not prior to the community; the community is the condition for the self.

This does not mean that Ubuntu denies individuality. Scholars like Kwame Gyekye and more recently Thaddeus Metz have argued that Ubuntu includes an individualist dimension: personal moral agency, autonomy, and a developmental view of personhood. The point is not that individuals do not exist, but that their humanity, their personhood, their excellence are co-constituted through relationships of care, recognition, and mutual obligation. As one scholar noted, each individual sees every other individual member of the community as another self. My humanity is not mine alone; it is co-substantively bestowed upon the other and me.

What this means for "intelligence"

The implications for the central question of this book are striking.

If personhood is constituted through relationships, then what counts as "excellence" cannot be fully specified without reference to relationships. An individually brilliant person who contributes nothing to communal life, who fails to recognize the humanity of others, who does not sustain the web of relationships through which persons become persons, is not, in Ubuntu terms, excellent. They may be cognitively powerful, but they are not yet fully a person, because personhood requires the relational dimension that their individual brilliance ignores.

This is not merely a normative claim (that we should value social skill alongside cognitive ability). It is a constitutive claim: being human, in the full sense, *is* being-in-relationship. Excellence that is not relational is not human excellence, because humanity itself is relational.

Consider what this means for IQ testing. An IQ test measures the performance of an individual, in isolation, on a standardized set of tasks, under controlled conditions designed to minimize social interaction. From an Ubuntu perspective, this is measuring something, but it is measuring it in a context that systematically excludes the dimension through which personhood and therefore genuine excellence are constituted. It is like measuring a musician's excellence by testing how well they play alone in a soundproof room, when what matters is how they play with others, how they listen, how they respond, how they contribute to an ensemble. The measurement is not wrong. It is radically incomplete, and its incompleteness is not incidental but structural.

This connects to a dimension of excellence that this book tracks in Movement II under the heading of "collective and distributed intelligence" (Chapter 31). But Ubuntu's challenge goes deeper than the claim that group intelligence exists alongside individual intelligence. It challenges the assumption that individual and collective intelligence are separate things that can be measured independently. If personhood is relational, then the individual's capacities are always already shaped by, and expressed through, relationships. Extracting the individual from the relational context and measuring them alone is not neutral isolation; it is distortion.

Ubuntu in practice: recognition and reconciliation

How does the Ubuntu tradition recognize excellence? Not through examinations, debates, or formal certification, but through the judgment of the community over the course of a life. A person demonstrates Ubuntu through their conduct: through hospitality to strangers, care for the vulnerable, willingness to share resources, participation in communal decision-making, restraint in the exercise of power, and the quality of their relationships over time.

The most visible institutional expression of Ubuntu in recent history is its role in South Africa's Truth and Reconciliation Commission, chaired by Archbishop Desmond Tutu after the end of apartheid. The Commission's approach to justice was explicitly informed by Ubuntu: rather than pursuing purely retributive punishment, it sought to restore relationships through truth-telling, acknowledgment, and the possibility of forgiveness. This is a form of justice that treats the repair of the relational web as more important than the punishment of individuals, and it reflects Ubuntu's foundational commitment to the primacy of communal bonds.

For this book's purposes, the Commission is significant because it demonstrates that Ubuntu is not merely an abstract philosophical position. It has real institutional consequences. When Ubuntu shapes how a society handles its deepest conflicts, it shows that relational personhood is not a decorative cultural value but a working political principle.

What Ubuntu bundled together

Applying the four-layer template:

Vocabulary. *Ubuntu* (Zulu/Xhosa: humanness, the quality of being a person through others), *Hunhu* (Shona equivalent), *Botho* (Sotho/Tswana equivalent), *umuntu ngumuntu ngabantu* (a person is a person through other persons), *ukama* (relatedness).

Phenomena. Ubuntu tracks: relational personhood (the process of becoming fully human through relationships), mutual recognition and care, communal harmony, the quality of interpersonal conduct

over time, the capacity to sustain relationships of reciprocity and solidarity, and the moral dimension of all human excellence. These map onto the social-relational competence, moral discernment, and collective capability dimensions in Movement II.

Norms. Excellence is relational, communal, and processual. One becomes excellent by deepening relationships of care and mutual recognition, by contributing to communal wellbeing, and by treating the humanity of others as co-substantive with one's own. Individual cognitive power, unmoored from relational conduct, does not count as excellence. This is the most communitarian normative framework in the survey.

Recognition. Excellence is recognized through community judgment over the course of a life: through the quality of one's conduct, relationships, hospitality, contribution to communal decisions, and care for others. There is no examination, no formal certification, and no single moment of assessment. Recognition is longitudinal and relational, depending on the sustained testimony of those who have lived alongside the person being assessed.

What this means for the argument

Ubuntu's contribution to the book is not a new theory of cognition or a new model of knowledge. It is a challenge to the framework within which all other theories in this survey operate.

First, it challenges the individualist assumption. Every cognitive test, from IQ to EI to SJTs, measures individual performance. Ubuntu suggests that this is a distortion: that human excellence is always relational, and that measuring individuals in isolation misses the most important dimension of what makes them excellent.

Second, it connects to a growing body of research on collective intelligence, distributed cognition, and team performance. Modern organizational psychology has begun to discover empirically what Ubuntu articulates philosophically: that the intelligence of a group is not simply the sum of the intelligences of its members, and that relational dynamics (trust, communication patterns, equitable participation) predict group performance better than average individual IQ.

Third, Ubuntu challenges the Western philosophical tradition's dominant assumption that the self is prior to the community. This assumption, running from Descartes through Locke to contemporary liberal individualism, has shaped not only political philosophy but also the entire architecture of cognitive testing. If the self is prior to the community, then measuring the self in isolation is natural and appropriate. If the community is constitutive of the self, then measuring the self in isolation is systematically misleading.

The Ubuntu tradition does not replace the other traditions in this survey. It does not deny that individ-

uals differ in cognitive ability, practical judgment, or emotional attunement. But it insists that these individual differences are always expressed within, and partly constituted by, relational contexts. Any framework for human excellence that ignores this is not merely incomplete. It has misunderstood its subject.

Chapter 12: Māori, Inuit, Pacific, and Indigenous Traditions. Memory, Ecology, Collective Life, and Culturally Grounded Assessment

This chapter covers multiple Indigenous knowledge traditions in a single treatment. This is not because any of them is less important than the traditions that received individual chapters. It is because each one, while distinctive, contributes to a shared pattern: the insistence that knowing is inseparable from place, practice, ancestry, ecological attunement, and collective continuity. Together, they illuminate dimensions of human excellence that the previous chapters have touched on but not centered, particularly embodied ecological skill, trained memory of extraordinary depth, and culturally grounded models of assessment that challenge the universality of Western psychometric tools.

A methodological caution stated in Chapter 3 becomes especially important here. These traditions encoded their knowledge primarily in oral practice, apprenticeship, ritual performance, and living transmission. The written sources available to this book are ethnographic studies, philosophical reconstructions by Indigenous and non-Indigenous scholars, community accounts, and contemporary Indigenous-led scholarship. These are legitimate evidence, but they are mediated in ways that texts by Aristotle or Zhu Xi are not. The chapter treats them with the same philosophical seriousness given to every other tradition in the survey, while being honest about the nature of the archive.

Māori: the baskets of knowledge and the houses of learning

In Māori tradition, knowledge has a sacred origin. According to one widely known account, the god Tāne ascended to the twelfth and highest heaven, Te Toi-o-ngā-rangi, and returned with three *kete wānanga* (baskets of knowledge). These baskets, which were brought to earth and taught in *whare wānanga* (houses of learning), represent not just information but ordered, sacred, authoritative knowledge whose proper transmission is a matter of spiritual significance.

The concept of *mātauranga Māori* spans knowledge, culture, values, and worldview. As researchers at the University of Auckland have noted, *mātauranga Māori* includes knowledge generated using techniques consistent with scientific methods, but explained according to a Māori worldview. The point is not that Māori knowledge is "really" Western science in disguise. It is that systematic observation, pattern recognition, prediction, and practical application were developed independently within a Māori framework that connects them to *whakapapa* (genealogy), *tikanga* (correct procedure), and the spiritual dimensions of the natural world.

The *whare wānanga* were selective institutions where carefully chosen students received instruction in high-level knowledge: cosmological narratives, genealogies, ritual, astronomy, and other sub-

jects critical to the tribe's identity and survival. Participants were selected for aptitude and lineage. Learning was conducted in a state of *tapu* (sacredness), often away from the village, during the winter months. Students were required to memorize vast bodies of material: *waiata* (songs), *karakia* (prayers), *whakapapa* (genealogies), and oral histories. The cognitive demands were enormous. The knowledge experts (*tohunga*) who led these institutions used whakapapa as an organizing framework for memorization, a mnemonic architecture as sophisticated as the Roman method of loci discussed in Chapter 5.

What matters for this book is not just the content of what was learned but what counted as excellence. In the Māori context, an excellent mind was one that could retain, organize, and transmit large bodies of culturally essential knowledge accurately and with appropriate understanding of its significance. This is a form of intelligence that centers trained memory, cultural literacy, and the responsibility of transmission. It is quite distant from modern IQ testing, but it is cognitively demanding in ways that no standardized test captures: the integration of enormous bodies of information with their genealogical, spiritual, and practical contexts, maintained without written records, and transmitted across generations through performance.

The last traditional whare wānanga were held in the second half of the nineteenth century. But the concept has been revived in modern form: contemporary wānanga are Māori tertiary institutions that provide education grounded in Māori knowledge, language, and values. Kaupapa Māori assessment frameworks explicitly argue that assessment should be grounded in Māori values, language, and knowledge rather than imported from Eurocentric assumptions. This is one of the clearest contemporary examples of an Indigenous community building its own measurement tools rather than accepting externally imposed ones.

Inuit: knowledge, values, and working together well

Inuit Qaujimagatuqangit (IQ), often translated as "traditional Inuit knowledge," is broader than the English word "knowledge" suggests. It encompasses societal values, practical skills, perceptual acuity, social norms, and ways of living and deciding together. It is not a body of propositions to be learned but an integrated system for living well in an extreme environment.

One of the IQ principles most relevant to this book is *Piliriqatigiingniq*, which has been translated as "working together for a common cause" or "working together well, motivated by shared values and goals." This is not a description of teamwork as a corporate buzzword. It is a deep principle about how knowledge, decision-making, and competence are distributed across a community. The good hunter is not merely the one who can track caribou; he is the one whose conduct sustains the relationships and shared practices that make the community's survival possible.

Contemporary Nunavut educational frameworks explicitly tie IQ principles to curriculum competencies. This means that in at least one modern jurisdiction, an Indigenous knowledge system is being used to define what students should know and be able to do, not as a supplement to a Western curriculum but as a foundational framework in its own right. The significance for this book is that IQ demonstrates a living tradition of collective, practice-based, value-embedded intelligence that is now being translated into formal educational standards.

What IQ bundles together is striking in its breadth: environmental knowledge (weather, ice, animal behavior, geography), practical skills (hunting, navigation, tool-making), social values (sharing, consensus-building, elder respect), emotional regulation (patience, composure in dangerous situations), and collaborative competence (the ability to contribute to group functioning). No single dimension of modern psychometric testing captures this bundle. It cuts across cognitive ability, practical judgment, emotional regulation, social competence, ecological skill, and moral conduct simultaneously.

Pacific wayfinding: embodied intelligence at its most demanding

The traditions of Polynesian and Micronesian navigation, often called wayfinding, represent perhaps the most vivid case in this entire survey of a form of intelligence that modern testing cannot capture and that no standardized instrument has ever measured.

Before the invention of the compass, sextant, or GPS, Pacific navigators crossed thousands of miles of open ocean using observations of stars, sun, moon, ocean swells, wave patterns, wind, clouds, bird flight, and marine life. The Hawaiian star compass, a mental construct (not a physical instrument) developed by master navigator Nainoa Thompson based on traditional principles, divides the visual horizon into thirty-two houses, each separated by 11.25 degrees of arc. Navigators memorize the rising and setting points of hundreds of celestial bodies within this framework and use them, together with swell reading and environmental observation, to maintain course and estimate position across voyages lasting weeks.

The cognitive demands of this practice are extraordinary. The navigator must simultaneously hold in mind a mental model of the sky (which changes through the night and across seasons), a model of the ocean's swell patterns (which are affected by islands, weather, and current interactions), a dead-reckoning estimate of distance traveled, and a continuously updated integration of environmental cues (cloud formations, bird behavior, water color, wave refraction patterns). This is multimodal sensory integration of a complexity that rivals anything studied in modern cognitive neuroscience. Yet because it is embodied, situated, and ecological rather than abstract and decontextualized, no IQ test or cognitive battery has ever attempted to measure it.

The transmission of navigational knowledge followed a strict apprenticeship model. Master navigators selected apprentices for aptitude and trained them over years, sometimes decades. Knowledge was transmitted orally, often in the form of chants and songs that encoded navigational information. The apprentice demonstrated competence not by passing a written examination but by successfully navigating a voyage. Success was the test. Arrival at the intended destination, safely, with crew and cargo intact, was the proof of competence.

As navigator Lehua Kamalu put it: "We are the compass." The navigator is not a person who uses instruments. The navigator is the instrument: a human being whose trained perception, memory, judgment, and embodied skill are the means by which a community crosses the ocean. This is intelligence in the most demanding and consequential sense, and it is completely invisible to any measurement tradition that equates intelligence with performance on written tests of abstract reasoning.

Contemporary Indigenous-led assessment: building tools from the ground up

A development of direct relevance to this book's argument is the contemporary movement, particularly in Australia and New Zealand, to develop assessment tools grounded in Indigenous knowledge frameworks rather than imported from Western psychometrics.

In Australia, First Nations scholars and communities have argued that standard assessment tools embed Western assumptions and miss local strengths. The response has been to co-design and statistically validate tools for child learning, youth wellbeing, and cultural learning "by us, for us." These tools measure what the communities themselves consider important: connection to country, cultural knowledge, kinship competence, and social-emotional wellbeing defined in community terms.

In New Zealand, Kaupapa Māori assessment explicitly argues that assessment should be grounded in Māori values, language, and knowledge. This is not merely adding Māori content to Western tests. It is rethinking what assessment is for and what counts as evidence of competence, starting from a Māori philosophical foundation.

These projects are important for the book because they represent real-time examples of the plural assessment model that Movement III will advocate. They demonstrate that it is practically possible, not just theoretically desirable, to build rigorous assessment tools that start from non-Western foundations. They also demonstrate the challenges: co-design is slow, funding is limited, and the tools must be robust enough to withstand scrutiny while remaining true to the values they are designed to serve.

What these traditions bundled together

Applying the four-layer template across the traditions covered in this chapter:

Vocabulary. Māori: *mātauranga* (knowledge/understanding), *māramatanga* (enlightenment/wisdom), *wānanga* (sacred knowledge, learning institution, method of knowledge transmission), *kete wānanga* (baskets of knowledge), *whakapapa* (genealogy), *tohunga* (knowledge expert). Inuit: *Qaujima-jatuqangit* (traditional knowledge system), *Piliriqatigiingniq* (working together well). Pacific navigation: no single standardized vocabulary, but a rich technical language for stars, swells, winds, and navigational signs, much of it specific to particular island groups.

Phenomena. These traditions were tracking: trained memory of enormous cultural and environmental bodies of knowledge, ecological attunement (reading stars, weather, ocean, land, and animal behavior), embodied skill (navigation, hunting, building, weaving), collaborative competence (the ability to contribute to group survival and functioning), cultural transmission (the responsibility of passing knowledge accurately to the next generation), and the integration of knowledge with spiritual, genealogical, and ethical frameworks. These map most directly onto the embodied-ecological skill, collective capability, retained knowledge, and practical judgment dimensions in Movement II.

Norms. Excellence in these traditions is inseparable from responsibility. To know is to be responsible for transmitting what one knows. To be skilled is to be responsible for using one's skill in the service of the community. To be a navigator is to be responsible for the lives of everyone on the canoe. This normative framework treats knowledge as a communal resource held in trust, not as an individual asset to be measured and ranked.

Recognition. The primary recognition methods are apprenticeship and demonstrated performance. The master navigator recognizes the apprentice's competence by entrusting them with responsibility for a voyage. The *tohunga* recognizes the student's competence by admitting them to the whare *wānanga* and, eventually, by entrusting them with the transmission of sacred knowledge. The community recognizes the elder's wisdom through decades of observed conduct, judgment, and contribution. In the contemporary context, Indigenous-led assessment projects represent a new form of recognition: measurement tools designed by communities to validate what they consider important, using methods that reflect their own values.

What this means for the argument

The Indigenous traditions covered in this chapter contribute four distinctive elements to the book.

First, they center embodied ecological skill as a genuine form of intelligence. Pacific wayfinding is not a metaphor for intelligence. It is intelligence: a cognitive achievement of extraordinary so-

phistication that integrates perception, memory, judgment, and embodied skill in real time under high-stakes conditions. Its absence from any modern psychometric battery is not evidence of its insignificance. It is evidence of the battery's narrowness.

Second, they demonstrate that trained memory, when operating at the scale required by the whare wānanga or the navigational tradition, is a cognitive capacity of the first order. The modern psychometric tradition treats memory as a supporting function of intelligence, subordinate to reasoning. These traditions treat it as central: the capacity to retain, organize, and transmit vast bodies of knowledge accurately, across generations, without writing, is the foundation of cultural continuity and communal survival.

Third, the Inuit IQ framework and the contemporary Indigenous-led assessment projects demonstrate that it is practically possible to build measurement from non-Western foundations. These are not theoretical proposals. They are working programs that define competence in community terms, design tools to assess it, and validate those tools empirically. They are existence proofs for the plural assessment model that Movement III will propose.

Fourth, these traditions collectively challenge the assumption that intelligence is primarily about solving novel problems. In these contexts, intelligence is at least as much about maintaining continuity: preserving knowledge, transmitting it accurately, applying it in the contexts where it has been validated over generations, and adapting it carefully as conditions change. This is a dimension of excellence that modern psychometrics, with its emphasis on novelty and fluid reasoning, systematically undervalues.

Chapter 13: Nahua, Japanese, and Other Traditions. Breadth Beyond the Major Cases

This chapter serves a different function from those that precede it. Chapters 4 through 12 provided sustained analyses of individual traditions, each given enough space to develop its vocabulary, phenomena, norms, and recognition methods in detail. This chapter surveys more briefly several traditions that reinforce the pattern established by the preceding chapters and that deserve, in a longer work, their own sustained treatment. Its purpose is not to suggest that these traditions are less important than those already covered. It is to demonstrate that the pattern identified so far, the near-universal refusal to reduce human excellence to a single cognitive dimension, extends well beyond the traditions examined in depth. It also names, honestly, what the survey has not been able to cover.

The Nahua *tlamatini*: giving wisdom to the face and strength to the heart

The Nahua (Aztec) philosophical tradition of pre-Columbian central Mexico offers one of the most vivid cases in the survey of a civilization that defined human excellence in explicitly multi-dimensional terms and built an educational system around cultivating it.

The scholar Miguel León-Portilla, working from over ninety original Nahuatl-language documents, demonstrated that the Nahua developed a sophisticated philosophical tradition. At its center was the *tlamatini* (plural: *tlamatinime*), a figure whose name means "one who knows something" or "knower of things." The *tlamatinime* were teachers, moralists, cosmologists, psychologists, and humanists. They were responsible for preserving and transmitting the sacred songs, poetry, astronomical knowledge, and cultural narratives in which Nahua learning was encoded. León-Portilla argued that they constituted a genuine philosophical class, comparable in function (though not in content) to Greek philosophers or Indian *rishis*.

The Nahua educational system had two primary institutions. The *Calmécac*, attended primarily by the nobility, provided what we would now call higher education: instruction in the *tecpilatlatolli* (lordly or elevated language), sacred songs (*teocuicame*), *in xōchitl in cūcatl* ("flower and song," the Nahuatl expression for poetry and truth-seeking), astronomy, calendar interpretation, history, and religious ritual. The *Telpochcalli* ("house of youth"), open to commoners, focused on military training, civic duties, and practical skills, though it too included instruction in moral conduct and cultural knowledge.

What makes the Nahua case relevant to this book is the Nahuatl philosophy of education itself. Two Nahuatl terms for education encode its purpose: *tlacahuapahualiztli* (the act of strengthening or

bringing up persons) and *neixtlamachiliztli* (the act of giving wisdom to the face). The educational goal, as León-Portilla interpreted the Nahuatl sources, was to endow the face with wisdom and the heart with strength. *Ixtli* (face) and *yollotl* (heart) together formed the Nahuatl concept of personality or character. To give someone a face was to cultivate their individuality, self-knowledge, and public identity. To strengthen their heart was to develop their will, moral resolve, and emotional fortitude. Education was the process of forming both simultaneously.

The *Huehuetlatolli* ("ancient words" or "discourses of the elders") were the oral teaching instruments through which this formation was transmitted. These were carefully constructed speeches, memorized verbatim and delivered by elders, parents, and *tlataminime* at key life transitions: birth, coming of age, marriage, assumption of office. They are not casual advice. They are formal philosophical discourses on how to live, how to respect others, how to seek what is good and righteous, and how to avoid evil. Their memorization and performance was itself a form of intellectual training.

The Nahua tradition also held that "the only truth on earth" was found in *in xōchitl in cuīcatl*, "flower and song," the poetic-philosophical expression through which the highest truths could be apprehended. This is a striking claim: it locates ultimate truth not in syllogistic reasoning or empirical observation but in poetic intuition. The *tlatamini* was someone whose heart had been "deified" through contemplation and creative expression, enabling them to perceive and communicate truths inaccessible to ordinary cognition.

For the book's argument, the Nahua case reinforces the pattern in three ways. First, it provides yet another independently developed ideal of the excellent person (the *tlatamini*) that integrates intellectual capacity with moral character, emotional strength, aesthetic sensibility, and social responsibility. Second, it demonstrates that formal educational institutions aimed at cultivating thick human excellence are not unique to the Eurasian traditions; they existed independently in Mesoamerica before European contact. Third, the "face and heart" formulation is one of the most elegant expressions in the entire survey of the multi-dimensional ideal: excellence is not a single trait but the integrated cultivation of public identity (face) and interior strength (heart).

Japanese traditions: knowing through doing

Japanese philosophical traditions present a distinctive approach to human excellence that differs from both the Western and Chinese models, despite deep historical connections with the latter.

Several features of the Japanese intellectual tradition are relevant to this book.

First, Japanese thought, particularly in the Zen Buddhist, Confucian, and martial arts traditions, has consistently emphasized the unity of knowledge and action. The Confucian thinker Wang Yangming

(whose influence in Japan was enormous, particularly through the *Yōmeigaku* school) formulated this as the unity of knowing and acting (*chi-kō gōitsu* in Japanese). This principle holds that genuine knowledge is inseparable from practice. To know something and not act on it is, in this framework, not yet to know it. This directly challenges the Western philosophical assumption, dominant from at least Descartes onward, that knowledge is primarily theoretical and that action is a separate domain.

Second, Japanese aesthetic and martial traditions developed the concept of *dō* ("the way"): a disciplined practice pursued over a lifetime through which the practitioner achieves a form of mastery that integrates bodily skill, mental discipline, aesthetic refinement, and moral formation. *Chadō* (the way of tea), *kendō* (the way of the sword), *judō* (the way of flexibility), *kadō* (the way of flowers, or ikebana), and *shodō* (the way of the brush) are all examples. In each case, excellence is not a cognitive state but an embodied, practiced, relational achievement that takes years or decades to cultivate. The tea master, the calligrapher, the martial artist: all are figures whose excellence is visible in what they *do*, not in what they *know* abstractly. This connects directly to the embodied-ecological skill dimension discussed in Chapter 12 and to the book's general argument that intelligence expressed through practice is no less cognitively demanding than intelligence expressed through abstract reasoning.

Third, the Japanese concept of *kata* (form, pattern) embodies a distinctive theory of learning. In martial arts, calligraphy, and other *dō* disciplines, the student begins by imitating the master's form precisely. Only after years of disciplined imitation does the student begin to develop their own expression. The progression is often described as *shu-ha-ri*: protect/obey the form, break from the form, transcend the form. This developmental model, from faithful imitation through creative departure to spontaneous mastery, is structurally similar to the Buddhist three-form *prajñā* model (hearing, reflection, cultivation) and to the Confucian progression from learning through practice to embodied virtue. It suggests a cross-civilizational convergence on the idea that expertise develops through stages, and that the highest form of mastery transcends the rules it began by learning.

Fourth, the Japanese philosophical tradition has been notably resistant to the mind-body dualism that characterizes much of Western thought. In the Zen tradition, the body is not an obstacle to knowing; it is the medium through which knowing occurs. Sitting meditation (*zazen*) is not preparation for insight; it *is* insight in practice. The martial artist's trained reflexes are not mindless automation; they are a form of intelligence expressed through the body. This anti-dualism connects to the broader argument of this book: that the Western separation of "intelligence" (located in the mind) from "skill" (located in the body) is a philosophical choice, not a universal feature of human cognition.

Other traditions: gesturing at what this survey cannot cover

An honest survey must name its gaps. This book has not been able to provide sustained treatment of many traditions that would reinforce and complicate its argument.

Southeast Asian traditions, including Thai, Burmese, Khmer, and Vietnamese philosophical and pedagogical systems, developed their own accounts of excellence under the combined influence of Indian (Hindu and Buddhist), Chinese, and Indigenous sources. The Theravāda Buddhist traditions of mainland Southeast Asia, with their emphasis on *sīla* (moral conduct), *samādhi* (concentration), and *paññā* (wisdom), deserve their own chapter. The Javanese concept of *rasa* (refined inner feeling as the basis of wisdom and social competence) is particularly relevant to the emotional attunement dimension.

Central Asian traditions, including the nomadic philosophical cultures of the Mongol, Turkic, and Tibetan-influenced peoples of the steppe, developed forms of excellence centered on ecological skill, animal husbandry, oral epic poetry (the *Manas* tradition, for example, involves texts of extraordinary length memorized and performed by specialist bards), and consensus-based decision-making. These traditions are largely absent from global philosophy but represent genuine accounts of human excellence.

The Andean traditions of the Quechua and Aymara peoples developed concepts of *sumak kawsay* ("good living" or "living well") that integrate ecological balance, communal harmony, and spiritual relationship with the land. This is an account of excellence that is explicitly anti-individualist and anti-accumulative, defining the good life not as the maximization of individual achievement but as the maintenance of balanced relationships.

The Jewish philosophical tradition, with its emphasis on *chokhma* (wisdom), *talmud Torah* (study as a lifelong religious obligation), *mussar* (ethical self-improvement), and the *talmid chacham* (the wise student) as an ideal of human excellence, deserves more attention than this survey has given it. The Talmudic tradition of dialectical reasoning and the yeshiva system of paired study (*chavruta*) represent distinctive models of both intellectual excellence and its recognition.

These gaps are not fatal to the book's argument, but they constrain it. The pattern identified in Chapters 4 through 12, the near-universal refusal to reduce human excellence to a single cognitive dimension, is supported by every tradition examined. It would be strengthened further by the addition of these other cases, all of which, on preliminary investigation, reinforce rather than contradict the pattern.

What this means for the argument

This chapter contributes three things to the book's project.

First, it demonstrates that the pattern is broader than the major cases. The Nahua *tlamatini* with his integrated "face and heart," the Japanese *dō* practitioner with their embodied mastery, the Andean *sumak kawsay* with its relational ecology: all point in the same direction as Aristotle, Confucius, the Yoruba *omolúàbí*, and the Buddhist *pāramitā* framework. The pattern is not an artifact of the cases selected. It is robust across civilizations, continents, and millennia.

Second, the Nahua case provides the first Mesoamerican example in the survey, demonstrating that the pattern extends to the pre-Columbian Americas. This matters because it eliminates the hypothesis that the pattern is an artifact of Old World cultural contact. The Nahua developed their philosophical tradition independently of any Eurasian influence, and they arrived at a structurally similar conclusion: excellence requires the integrated cultivation of multiple dimensions, not the maximization of a single one.

Third, the honest naming of gaps demonstrates the book's intellectual seriousness. A survey that claimed completeness would not be credible. A survey that acknowledges its limits while identifying a robust cross-civilizational pattern earns the reader's trust precisely by admitting what it cannot yet prove.

Chapter 14: Oral Traditions, Sages, and the Politics of What Counts as Knowledge

The preceding ten chapters have surveyed more than a dozen traditions across six continents. Before the survey turns to the Western narrowing that produced the modern concept of intelligence, it is necessary to pause and address a problem that has been present throughout: the global record of human thought about excellence is structurally uneven, and this unevenness is not neutral. It has shaped which traditions are taken seriously as philosophy, which are dismissed as folklore, and which definitions of excellence have been treated as natural and which as exotic. Any honest survey must name this asymmetry and reckon with its consequences.

The archival asymmetry

Literate state traditions left texts. Greece left Aristotle's *Nicomachean Ethics*, China left the Confucian *Analects*, India left the *Nyāya Sūtras*, and the Islamic world left al-Fārābī's *Enumeration of the Sciences*. These texts can be studied, translated, debated, and assigned in university courses. They have generated centuries of commentary. They are treated, rightly, as philosophy.

Oral traditions left something different: proverbs, songs, chants, genealogies, ritual performances, apprenticeship practices, and living knowledge held by recognized experts. The Yoruba Ifá corpus, the Māori *whakapapa* framework, the Inuit *Qaujimaqatuqangit*, the Polynesian navigational tradition: all of these contain philosophical content as sophisticated as anything in the written traditions. But their medium of preservation makes them structurally harder for text-based academic disciplines to engage with. They do not fit neatly into library catalogs, citation indexes, or philosophy syllabi. As a result, they have been systematically undervalued.

This is not merely an oversight to be corrected by adding a few non-Western readings to philosophy curricula, though that would help. It is a structural feature of how knowledge production has been organized globally since the colonial era. The institutions that decide what counts as "philosophy" (university philosophy departments, academic publishers, peer-reviewed journals) were built on European models, using European languages, applying European standards of what constitutes philosophical rigor. When a tradition's knowledge is preserved in written treatises organized by argument and counter-argument, it looks like philosophy. When it is preserved in proverbs deployed in council, in chants memorized by apprentices, in navigational skills demonstrated on the open ocean, or in elder judgments rendered over decades, it does not look like philosophy to institutions trained to recognize only the textual form.

The consequence for this book's project is significant. The traditions covered in Chapters 4 through

9 (Greece, Rome, China, India, Buddhism, Islam) have been studied by academic philosophers for decades or centuries. Their concepts have been translated, their arguments reconstructed, and their implications debated. The traditions covered in Chapters 10 through 13 (Akan, Yoruba, Ubuntu, Māori, Inuit, Pacific, Nahua, Japanese) have received serious philosophical attention more recently, often through the work of scholars from those traditions who have had to fight for recognition within disciplines that initially regarded their subject matter as anthropology, folklore, or area studies rather than philosophy.

This asymmetry does not mean that the text-based traditions are more philosophical than the oral ones. It means that the evidence base is different, and the interpretive challenges are different. When Aristotle's argument for *phronēsis* is evaluated, the reader can examine the text, consider the premises, and assess the reasoning directly. When the Akan *nyansa* tradition is evaluated, the evidence comes through ethnographic observation, philosophical reconstruction by scholars like Gyekye and Wiredu, and the analysis of proverbs, adjudication practices, and community testimony. Both forms of evidence are legitimate. But the latter requires a broader definition of what counts as philosophical evidence than many academic philosophers have traditionally accepted.

Recognition methods revisited: measurement by other names

One of the most consequential effects of the archival asymmetry is the tendency to treat written examinations and standardized tests as "real" measurement, while treating apprenticeship assessment, elder judgment, proverb competence, and demonstrated mastery as mere "cultural practice." This distinction is indefensible.

Consider what the traditions surveyed in this book actually *do* when they identify excellence.

The Chinese imperial examination (*keju*) selected candidates through a formal written examination: essays on classical texts, evaluated by trained examiners, with results that determined bureaucratic appointment. This is obviously measurement: it selects, ranks, and validates competence through a systematic procedure. No one disputes that it is a recognition system.

The Tibetan Geshe system selects candidates through memorization, years of daily debate, and final examinations where the abbot chooses topics on the spot. This too is obviously measurement: it takes twenty years, it is brutally selective, and failure is real. It measures something different from the *keju* (dialectical skill rather than essay composition), but it is plainly a system for selecting and validating competence.

Now consider the *ijazah* system. A teacher who knows the student personally, who has assessed their learning face-to-face over years, grants formal authorization to transmit a specific body of knowledge.

This is personal rather than institutional, but it is still a recognition system. It selects (not everyone receives an *ijazah*), it validates (the teacher attests to competence), and it has consequences (the recipient is now authorized to teach).

The Polynesian navigational tradition selects apprentices for aptitude, trains them over years, and validates competence through the ultimate test: successful navigation of an ocean voyage. This is measurement in the most consequential sense. The test is not graded on a curve. It is pass/fail, and failure can mean death.

Akan and Yoruba communities recognize wisdom through decades of observed conduct, the quality of speech in councils, the aptness of proverb deployment, and the track record of sound judgment. This is slower than a written examination, but it is arguably more valid: it assesses performance in the actual contexts where competence matters, over a time span long enough to distinguish genuine wisdom from lucky guessing or rhetorical flair.

Ubuntu traditions recognize personhood and excellence through the sustained quality of relational conduct: hospitality, care, reciprocity, and contribution to communal life. This is the least test-like of all the recognition methods in the survey, but it is still a form of social selection: communities distinguish between those who embody Ubuntu and those who do not, and this distinction has real consequences for social standing, leadership, and trust.

What all of these systems share is that they are methods for selecting and validating competence through systematic procedures, applied over time, with consequences for the person being assessed. They differ in their medium (written vs. oral vs. behavioral), their time scale (hours vs. years vs. decades), their degree of formalization (examination vs. apprenticeship vs. community judgment), and their criteria (textual knowledge vs. dialectical skill vs. navigational competence vs. relational conduct). But they are all doing the same fundamental thing: distinguishing, within a population, those who have achieved a particular form of excellence from those who have not.

The modern psychometric tradition, which will be the subject of Chapters 16 and 17, is one member of this family of recognition systems. It is not the only one, and it is not self-evidently the best one. It has specific advantages (standardization, scalability, statistical analysis) and specific limitations (decontextualization, narrowness, cultural bias). Understanding it as one recognition system among many, rather than as the definitive method for measuring intelligence, is essential for the argument of this book.

The sage and the philosopher

Henry Odera Oruka's sage philosophy project, introduced briefly in Chapter 10, deserves further attention here because it addresses the archival asymmetry directly.

Oruka argued that African philosophy could not be reduced to what he called "ethnophilosophy" (the systematic description of communal belief systems) or "nationalistic-ideological philosophy" (the deployment of philosophical language for political purposes). He proposed a third category: "philosophic sagacity," the individual critical philosophical thinking of persons recognized by their communities as sages. These were not academics. They were people, often farmers, elders, or community leaders, who had spent decades reflecting on fundamental questions and who were recognized by their neighbors for the depth and originality of their thought.

Oruka's project matters because it challenges the assumption that philosophy requires literacy. The sages he documented were oral thinkers. Their philosophical activity was conducted through conversation, debate, adjudication, and counsel rather than through the writing and publication of texts. But the quality of their thought, its rigor, its originality, its engagement with fundamental questions about knowledge, justice, reality, and the good life, was undeniable to anyone who engaged with it directly.

The sage, as a figure, appears in many of the traditions surveyed in this book. The Greek *sophos*, the Chinese *shēngrén* (sage-king), the Indian *ṛṣi*, the Yoruba *babaláwo*, the Māori *tohunga*: all are figures recognized for wisdom that goes beyond ordinary knowledge. What makes them sages is not that they have passed a test. It is that their understanding has been validated over time through the quality of their judgment, their counsel, their teaching, and their lives. This is recognition through track record rather than through examination, and it is a form of measurement that is, in many ways, more ecologically valid than a standardized test: it assesses competence in the contexts that matter, over time scales that allow genuine wisdom to be distinguished from mere cleverness.

What "counts" as philosophy, and why it matters

The politics of what counts as philosophy is not a side issue for this book. It is central to the argument. If philosophy is defined as the production of written, argumentative texts by individual thinkers working within an institutional tradition of commentary and debate, then philosophy is, by definition, limited to literate traditions. This definition excludes most of humanity's intellectual history. It treats the philosophical achievements of the Akan, Yoruba, Inuit, Māori, and Pacific traditions as invisible, not because they lack philosophical content but because they lack the institutional form that the definition requires.

If philosophy is defined more broadly, as sustained critical reflection on fundamental questions about knowledge, reality, value, and the good life, then philosophy has been practiced in every human culture. It has been practiced in different media (text, speech, song, proverb, ritual, practice) and through different institutional forms (academies, monasteries, councils, apprenticeships, sacred houses of learning). The content differs across traditions. The form differs. But the activity, the sustained, critical, reflective engagement with fundamental questions, is human rather than specifically Western.

This book adopts the broader definition, not because the narrower one is indefensible, but because the narrower one would make the book's central project impossible. If only written philosophical traditions count, then the survey would cover only half of human civilization and would miss precisely the traditions whose accounts of excellence are most challenging to the modern psychometric consensus. The purpose of the survey is to discover what happens when all of humanity's traditions are included. This requires a definition of philosophy, and of evidence, broad enough to include them.

The transition to the Western narrowing

With this chapter, the cross-civilizational survey is complete. Chapters 4 through 14 have examined more than a dozen traditions across six continents and several millennia. The pattern that has emerged is remarkably consistent.

Almost every tradition surveyed refuses to reduce human excellence to a single cognitive dimension. Almost every tradition integrates intellectual capacity with moral character, practical judgment, emotional attunement, social skill, or some combination of these. Almost every tradition has developed methods for recognizing excellence that go well beyond written examinations or standardized tests. And almost every tradition treats the excellent person as someone whose cognitive power has been cultivated within, and directed toward, a broader human formation.

The modern concept of intelligence, which emerged in Western Europe and North America between roughly 1850 and 1950, departed from this pattern. It stripped away the moral, emotional, social, and practical dimensions. It defined intelligence as a measurable cognitive trait. It built an institutional apparatus (IQ tests, scholastic aptitude tests, standardized assessments) to measure this trait at scale. And it exported this framework globally, so that today, in virtually every country on earth, children are assessed primarily on cognitive dimensions that represent only a fraction of what their own cultural traditions consider important.

The next four chapters (15-18) trace how this narrowing happened. They are not a story of villains.

They are a story of specific historical pressures, institutional needs, and philosophical choices that produced an outcome that, viewed against the full scope of human thought about excellence, looks like a radical departure from a near-universal consensus.

Chapter 15: Early Modern Europe. Method, Reason, Certainty, and the Beginning of the Narrowing

The cross-civilizational survey is complete. The pattern is clear: across more than a dozen traditions and several millennia, humanity defined excellence as something broader than abstract reasoning alone. Excellence was integrated with moral character, practical judgment, emotional attunement, social competence, ecological skill, or some combination of these. Education aimed at cultivating whole persons, not at maximizing a single cognitive variable.

Something changed. Between roughly 1600 and 1800, European philosophy took a series of steps that would eventually make it possible to conceive of intelligence as a measurable, value-neutral cognitive trait. These steps were not taken all at once, and none of the thinkers involved intended the outcome that eventually followed. But together they produced a philosophical environment in which the later invention of IQ testing became thinkable, even natural. This chapter traces those steps.

Bacon: knowledge as power, method as the great equalizer

Francis Bacon (1561-1626) did not define intelligence. He did something more consequential: he redefined the purpose of knowledge. For the traditions surveyed in Chapters 4 through 13, knowledge was inseparable from its human context: it was for living well, for governing justly, for achieving liberation, for sustaining community. Bacon reoriented knowledge toward a different end. Knowledge was for power: power over nature, power to improve the human condition, power to extend the reach of human capability.

This reorientation had a crucial consequence. If knowledge is for living well, then the knower's character matters. A person of bad character cannot live well, no matter how much they know. But if knowledge is for power over nature, then the knower's character is irrelevant to the knowledge itself. A morally deficient person can still build a better pump, identify a new compound, or predict an eclipse. The separation of knowledge from character, which every tradition in the survey resisted, becomes possible once the purpose of knowledge shifts from human formation to natural mastery.

Bacon also championed method as the key to reliable knowledge. His *Novum Organum* (1620) proposed a systematic procedure for investigating nature that would compensate for the weaknesses of human perception and reasoning. The "Idols" he identified (Idols of the Tribe, Cave, Marketplace, and Theater) were systematic biases in human cognition that method could correct. This is important for the book's argument because it treats human cognitive faculties as unreliable in their natural state. The senses deceive. Reasoning is distorted by prejudice, language, and received doctrine. Only

method, the disciplined procedure of observation, experiment, and induction, can produce reliable knowledge.

The implicit picture of the human mind here is strikingly different from the one found in, say, the Confucian tradition. For the Confucians, the heart-mind (*xin*) is naturally oriented toward moral goodness and needs only cultivation to achieve its potential. For Bacon, the mind is naturally prone to error and needs method to overcome its deficiencies. This is not a theory of intelligence, but it is a theory of cognition that treats the mind as a problem to be solved rather than a potential to be cultivated. It points toward a future in which cognitive capacity will be measured precisely because it is seen as a variable, something that differs across individuals and can be compensated for by technique.

Descartes: the thinking subject and the separation of mind from body

René Descartes (1596-1650) made two moves that are foundational for the story this book tells.

First, his method of radical doubt (*cogito ergo sum*, I think therefore I am) established the isolated thinking subject as the foundation of certainty. In the *Meditations* (1641), Descartes stripped away everything that could be doubted: the senses, the body, the external world, even mathematics. What remained was the pure act of thinking. The self that thinks is the most certain thing in the universe. For the book's argument, this move is significant because it locates the essential self in cognition alone. The body, the emotions, social relationships, practical judgment, ecological attunement, and all the other dimensions that the traditions in Chapters 4 through 13 treated as central to human excellence are, in Descartes's framework, separable from the thinking self. This does not mean Descartes denied their existence. But his philosophical method established a hierarchy: thought is primary, everything else is secondary. The long-term consequence, which Descartes did not intend, was a philosophical framework in which it made sense to measure "intelligence" as a purely cognitive trait, detached from body, emotion, character, and relationship.

Second, Descartes formalized the mind-body distinction as a fundamental metaphysical division. *Res cogitans* (thinking substance) and *res extensa* (extended substance) are ontologically distinct. The mind is not the body. Thought is not a physical process. This dualism, however controversial it became, shaped the conceptual architecture of Western psychology for centuries. When psychometricians later set out to measure "intelligence," they were measuring a mental capacity understood as distinct from physical ability, emotional sensitivity, and social skill. This separation was not inevitable. It was a philosophical choice, made by Descartes and inherited by his successors, that carved the mind out of the integrated human being and treated it as an independent domain.

Chapter 11 noted that Ubuntu inverts the Cartesian formula ("I am because we are" rather than "I think therefore I am"). Chapter 13 noted that Japanese philosophical traditions reject the mind-body dualism. These are not merely cultural differences. They are philosophical disagreements about the structure of human existence, and the Cartesian position is one side of a live argument, not a settled truth.

Locke and the empiricists: the mind as blank slate

John Locke (1632-1704) replaced Descartes's innate ideas with the *tabula rasa*: the mind begins as a blank slate, and all knowledge comes from experience. Locke's *Essay Concerning Human Understanding* (1690) proposed that the mind's operations (perception, memory, abstraction, reasoning, judgment) act on sensory input to produce knowledge. As noted in Chapter 1, Locke used "understanding" rather than "intelligence" for these operations, deliberately avoiding the scholastic connotations of the older term.

Locke's contribution to the narrowing is subtle but important. By treating the mind as a set of operations that process experience, he made it possible to think of mental capacities as decomposable: perception is one operation, memory another, reasoning a third. This decomposition is the ancestor of the factor-analytic approach to intelligence that would emerge two centuries later with Spearman and Thurstone. If the mind is a set of distinct operations, then those operations can in principle be measured separately, and their relationships can be analyzed statistically.

Locke also made a move that would prove consequential for the politics of intelligence: he argued that differences in understanding between people are due primarily to differences in experience and education, not to innate differences in mental capacity. This egalitarian implication of empiricism sits in tension with the hereditarian tradition that would later dominate IQ research. The tension between "intelligence is shaped by environment" and "intelligence is largely innate" runs through the entire history of psychometrics, and its roots are here, in the seventeenth-century debate between nativism and empiricism.

Hume and the sentimentalist counter-current

David Hume (1711-1776) represents a counter-current within the early modern tradition that is crucial for the book's argument. Hume argued that reason alone cannot motivate action and that moral judgments are grounded in sentiment, not in rational demonstration. "Reason is, and ought only to be, the slave of the passions," he wrote in the *Treatise of Human Nature* (1739-40).

This is important because it preserves, within the Western tradition, a version of the insight that the

Chinese, Buddhist, and Yoruba traditions articulated from the start: that cognition and affect are not fully separable, and that practical and moral judgment depend on emotional as well as rational capacities. Hume's sentimentalism is a direct ancestor of the modern emotional intelligence construct: the idea that emotional perception, understanding, and regulation are forms of competence that contribute to good judgment and effective action.

But Hume's insight was largely sidelined in the tradition that produced psychometrics. The dominant stream of Western thought about mental capacity, running from Locke through the Scottish Enlightenment to the associationist psychologists and on to Galton and Spearman, treated intelligence as primarily rational and cognitive. Hume's insistence on the centrality of sentiment was acknowledged in moral philosophy but not incorporated into the emerging science of mental measurement. When IQ testing was invented, it measured reasoning, not feeling; cognition, not affect. The sentimentalist counter-current had to wait until the 1990s (Salovey and Mayer, Goleman) to re-enter the measurement conversation as "emotional intelligence."

Kant: judgment, autonomy, and the limits of measurement

Immanuel Kant (1724-1804) occupies a complex position in this story. On one hand, his philosophy represents the highest achievement of the Enlightenment ideal of rational autonomy: the mature, self-governing use of reason, freed from dependence on authority and tradition. His famous definition of Enlightenment ("the emergence of humanity from its self-imposed immaturity") places the exercise of independent rational judgment at the center of human dignity.

On the other hand, Kant preserved a richer account of judgment than the psychometric tradition that followed him. In the *Critique of Judgment* (1790), he analyzed aesthetic and teleological judgment as forms of cognition that are irreducible to the application of rules. Judgment, for Kant, is the capacity to subsume a particular case under a general rule, and this capacity cannot itself be taught by rules, because any rule for applying rules would itself require judgment to apply. This is a philosophical insight of the first order: judgment is a cognitive capacity that exceeds any formal procedure. It is what allows a person to act wisely in novel situations where no algorithm is available.

Kant's analysis of judgment is directly relevant to the modern Situational Judgment Test (SJT) literature, which attempts to measure practical judgment in context. It also connects to the Greek *phronēsis* tradition and to the Akan *nyansa* tradition: all three hold that the highest form of practical cognition is not rule-following but the capacity to see what a situation requires. This capacity resists standardized measurement precisely because it exceeds any fixed procedure.

What was left behind

The early modern period did not kill the thicker traditions. Rhetoric, civic formation, moral philosophy, and the cultivation of character continued to be taught in European universities, churches, and aristocratic households throughout the seventeenth and eighteenth centuries. The ideal of the cultivated gentleman (or, more rarely, gentlewoman) persisted well into the nineteenth century, and it incorporated many of the same elements found in the *adab*, *paideia*, and *junzi* traditions: literary knowledge, social grace, moral refinement, and practical judgment.

But the philosophical foundations had shifted. Knowledge was being separated from character (Bacon). The thinking self was being separated from the body, the emotions, and the social world (Descartes). Mental operations were being decomposed into measurable components (Locke). And reason was being elevated above sentiment as the primary faculty of the excellent mind (the mainstream Enlightenment, against Hume's dissent).

None of these thinkers intended to produce the IQ test. None of them would have recognized psychometric *g* as a natural consequence of their philosophy. But together they created the intellectual conditions in which the next step, the systematic measurement of individual cognitive capacity as a single, quantifiable trait, became possible. That step is the subject of Chapter 16.

Chapter 16: The Birth of Measurable Intelligence

Chapter 15 traced the philosophical moves that made the modern concept of intelligence thinkable: the separation of knowledge from character, of mind from body, of reason from sentiment. But philosophy alone does not create measurement systems. The transformation of "intelligence" from a philosophical concept into a measurable, scorable, rankable trait required specific institutional pressures, specific scientific innovations, and specific individuals who brought them together. This chapter covers the period from roughly 1869 to 1927, during which the modern intelligence test was invented, refined, and institutionalized.

The story is not simple. It involves a hereditarian ideologue who got the science wrong, a French clinician who got it largely right but whose warnings were ignored, a British statistician who discovered an important empirical pattern and over-interpreted it, and a set of institutional forces (mass schooling, military mobilization, immigration control) that adopted intelligence testing not because the science demanded it but because the tests were useful.

Galton: hereditary genius and the wrong measurement

Francis Galton (1822-1911), Charles Darwin's half-cousin, is the founding figure of psychometrics. His 1869 book *Hereditary Genius* was the first systematic attempt to study intellectual eminence scientifically. His central claim was that intellectual abilities are inherited, distributed unequally across the population, and measurable. He was right about the distribution (intelligence, however defined, does vary across individuals). He was right that statistical methods could illuminate these differences. He was wrong about almost everything else.

Galton's approach to measuring intelligence rested on a specific hypothesis: that intelligence is a function of sensory acuity. He reasoned that individuals with sharper senses would perceive the world in greater detail and therefore reason better. In 1884, he established the world's first Anthropometric Laboratory at the International Health Exhibition in London, where over 9,000 visitors paid a small fee to have their visual acuity, auditory accuracy, reaction time, grip strength, breathing capacity, and other physical and sensory qualities measured.

The hypothesis was wrong. By 1901, a series of studies by James McKeen Cattell (who had coined the term "mental tests" based on Galton's approach) and his students showed no significant relationship between sensory discrimination and academic performance. Sensory acuity does not predict what most people mean by intelligence. The Galtonian approach to measurement was largely abandoned.

But Galton's contributions to the infrastructure of measurement survived. He pioneered correlation and regression to the mean as statistical techniques. He introduced the use of questionnaires as research instruments. He popularized the phrase "nature versus nurture." And he established the basic framework within which all subsequent intelligence research would operate: the assumption that intelligence is a natural trait, that it varies across individuals, that it is substantially inherited, and that it can be measured through standardized procedures.

Galton also coined the term "eugenics" and advocated for the selective breeding of human beings to enhance desirable traits. This connection between intelligence measurement and eugenic ideology is not incidental. From the very beginning, the measurement of intelligence was intertwined with programs for ranking, sorting, and controlling human populations. This does not invalidate the science of mental measurement, but it means that the institutional history of intelligence testing cannot be told as a story of pure scientific discovery. It was always, from its inception, a tool of social policy.

Binet: the right measurement for the wrong reasons

The story pivots to Paris. Alfred Binet (1857-1911) took a fundamentally different approach from Galton. Where Galton measured sensory acuity, Binet argued that intelligence involved higher-order mental processes: attention, memory, imagination, comprehension, and judgment. Where Galton sought to measure innate endowment, Binet sought to identify children who needed additional educational support.

In 1904, Binet was appointed by the French minister of public instruction to a commission charged with identifying children who were failing in the regular school system and needed special educational programs. Together with Théodore Simon, Binet developed the Binet-Simon Scale (1905): a set of thirty tasks arranged in order of increasing difficulty, designed to distinguish children with normal cognitive development from those who were developing more slowly. The 1908 revision introduced the concept of "mental age": a child whose performance matched the average of, say, seven-year-olds was assigned a mental age of seven, regardless of their chronological age.

For this book's argument, several features of Binet's work are significant.

First, Binet's definition of intelligence was remarkably broad. He described it in terms that sound more like the traditions surveyed in Chapters 4 through 13 than like the IQ tests that followed: judgment, good sense, practical sense, initiative, the faculty of adapting oneself to circumstances. He explicitly rejected the reduction of intelligence to a single faculty, writing that intelligence is not a single thing but a collection of faculties that work together.

Second, Binet's test was designed for a specific practical purpose: identifying children who needed help. It was not designed to rank all children on a single scale of cognitive ability, and Binet explicitly warned against using it that way. He cautioned that the scale should not be used to label children as permanently deficient, that intelligence is malleable and can be improved with proper instruction, and that the scores represent a snapshot of current performance, not a fixed trait.

Third, Binet's warnings were almost entirely ignored. When the Binet-Simon Scale crossed the Atlantic, it was adapted by Henry Goddard (who used it to classify immigrants at Ellis Island) and by Lewis Terman (who created the Stanford-Binet test in 1916, introducing the "intelligence quotient" or IQ as the ratio of mental age to chronological age multiplied by 100). In Terman's hands, the test became precisely what Binet had cautioned against: a ranking system applied to all children, used to sort them into educational tracks, and interpreted as measuring a largely innate and fixed trait.

The irony is profound. Binet, the inventor of the test that became the foundation of IQ measurement, defined intelligence as fundamentally about judgment and practical adaptation. The testing apparatus that followed his work increasingly left judgment behind, focusing on abstract reasoning, vocabulary, spatial manipulation, and processing speed. The very qualities that Binet thought were most important (good sense, adaptability, initiative) proved hardest to measure in a standardized format and were gradually dropped from the tests that bore his name.

Spearman: the discovery of *g*

Charles Spearman (1863-1945) made the next decisive move. In 1904, the same year Binet was appointed to his commission, Spearman published a landmark paper arguing that the positive correlations among different cognitive tests could be explained by a single underlying factor, which he called *g* (for general intelligence).

Spearman's observation was empirical and replicable. When people take multiple cognitive tests (vocabulary, spatial reasoning, arithmetic, memory, etc.), their scores tend to be positively correlated: people who do well on one test tend to do well on others. Factor analysis, a statistical technique Spearman developed for this purpose, extracts the common variance shared across tests and identifies it as a general factor. This is *g*.

The existence of positive correlations among cognitive tests is one of the most robust findings in all of psychology. It has been replicated thousands of times across cultures, age groups, and test formats. The book does not dispute this finding. But it is important to understand what *g* is and what it is not.

g is a statistical pattern: the common variance across a set of cognitive tests. It is not, despite

what some of its advocates have claimed, a single thing in the brain, a biological substrate, or a causal entity. The fact that tests correlate does not prove that a single underlying capacity causes the correlation. The correlation could be produced by multiple overlapping capacities, by shared environmental influences (such as education), by test design choices (tests that are constructed to correlate will correlate), or by some combination of these. The question of what *g* is, ontologically, remains open and contested.

Spearman's discovery was followed by decades of debate about the structure of intelligence. Louis Thurstone proposed seven primary mental abilities rather than a single *g*. Raymond Cattell distinguished fluid intelligence (*Gf*, the ability to solve novel problems) from crystallized intelligence (*Gc*, the ability to apply learned knowledge). John Carroll's three-stratum model proposed a hierarchy with *g* at the top, broad abilities in the middle, and narrow abilities at the bottom. The Cattell-Horn-Carroll (CHC) model, which synthesized these approaches, is currently the dominant structural model in psychometrics.

All of these models share one feature: they treat intelligence as a cognitive trait measured through performance on standardized tests. They differ about the internal structure of this trait (one factor or many, hierarchical or flat). But they all inherit the basic framework established by Galton (intelligence is a natural, measurable, variable trait), refined by Binet (it should be measured through cognitive tasks rather than sensory acuity), and formalized by Spearman (the positive manifold among tests reflects an underlying structure that can be analyzed statistically).

The institutional drivers

The intelligence test did not succeed because the science was irresistible. It succeeded because it was useful.

Mass public schooling, which expanded rapidly across Europe and North America in the late nineteenth and early twentieth centuries, created an administrative problem: how to sort millions of children into educational tracks. Intelligence tests offered a solution: a standardized, apparently objective procedure for identifying which children should receive which kind of instruction.

Military mobilization during the First World War created another problem: how to sort millions of recruits into military roles. The U.S. Army's Alpha and Beta tests, administered to approximately 1.75 million recruits between 1917 and 1919, were the first mass application of intelligence testing. The Alpha (for literate recruits) and Beta (for illiterate or non-English-speaking recruits) were designed to identify officer candidates and assign soldiers to appropriate roles. They demonstrated that intelligence testing could operate at scale and that the results were useful for institutional purposes.

Immigration policy created a third application. Henry Goddard administered Binet-type tests to immigrants arriving at Ellis Island, and his results were used to argue that certain national and ethnic groups were intellectually inferior. This application was scientifically indefensible (the tests were administered in English to non-English speakers, under stressful conditions, with no controls for education or cultural background), but it was politically influential. The Immigration Restriction Act of 1924 was partly justified by intelligence-testing data.

These institutional applications drove the proliferation of intelligence testing far more than the science itself did. The tests spread because they solved administrative problems: sorting, ranking, selecting, and placing large numbers of people into institutional categories. The fact that the tests measured something real (cognitive ability does vary across individuals) made them credible. But the reason they became ubiquitous was not that they captured the full range of human excellence. It was that they were efficient tools for institutional sorting.

What had been accomplished, and at what cost

By the 1920s, the basic architecture of modern intelligence testing was in place. The assumptions were established (intelligence is a measurable cognitive trait that varies across individuals), the tools were built (standardized tests administered under controlled conditions), the statistical framework was developed (factor analysis, correlation, the normal distribution), and the institutional applications were thriving (schools, military, immigration).

This was a genuine achievement. The ability to measure cognitive performance reliably and at scale is not trivial. It made possible scientific research on individual differences, educational interventions for children with learning difficulties, and fair(er) selection procedures for institutions that had previously relied on patronage, social class, or arbitrary judgment.

But the achievement came at a cost that the next two chapters will examine in detail. The cost was narrowing. Binet's original definition of intelligence (judgment, good sense, practical sense, initiative, adaptation) was broader than the tests that bore his name. Spearman's *g* was narrower than any of the concepts surveyed in Chapters 4 through 13. And the institutional adoption of IQ testing created a feedback loop: once intelligence was defined operationally as "what IQ tests measure," the tests became self-validating. Anything not captured by the tests was, by definition, not intelligence. The broad, thick, multi-dimensional accounts of human excellence that had characterized virtually every prior tradition were compressed into a single score.

Chapter 17 will present the strongest honest case for what the psychometric tradition got right. Chapter 18 will name what it amputated.

Chapter 17: Psychometrics, g , IQ, and What Was Gained

This chapter owes the psychometric tradition an honest hearing. The preceding chapters have built a case that the modern concept of intelligence is narrower than what most human traditions cared about. That case is strong. But it would be intellectually dishonest, and strategically foolish, to pretend that the narrowing produced nothing of value. It produced something remarkable: a body of scientific findings about human cognitive variation that is among the most robust, replicable, and practically consequential in all of psychology.

This chapter presents the strongest honest case for psychometric intelligence. Not as a defense of the claim that IQ captures all of human excellence (it does not, and the next chapter will say so explicitly), but as a recognition that the psychometric tradition discovered real things about real cognitive variation, and that any successor framework must account for those discoveries rather than dismissing them.

What IQ tests actually measure

Modern intelligence tests (the Wechsler Adult Intelligence Scale, the Stanford-Binet, Raven's Progressive Matrices, the Woodcock-Johnson, and others) are not single tests but batteries of subtests, each measuring a somewhat different cognitive ability. Typical subtests assess vocabulary knowledge, verbal comprehension, arithmetic reasoning, spatial visualization, working memory, processing speed, and abstract pattern recognition.

The subtests are designed to be diverse: they sample different cognitive domains and use different task formats. But when large numbers of people take these batteries, a consistent empirical pattern emerges: performance on the subtests is positively correlated. People who do well on vocabulary tend to do well on spatial reasoning, and those who do well on both tend to do well on working memory. This positive manifold, as psychometricians call it, is one of the most replicated findings in psychological science.

Factor analysis extracts the common variance shared across the subtests and identifies it as g , the general factor of intelligence. In the Cattell-Horn-Carroll (CHC) model, which is the dominant structural model in contemporary psychometrics, g sits at the top of a hierarchy. Below it are broad abilities: fluid reasoning (Gf), crystallized knowledge (Gc), short-term memory (Gsm), processing speed (Gs), visual-spatial processing (Gv), and others. Below those are narrow abilities: specific skills measured by individual subtests.

The IQ score that a person receives is a composite: a weighted combination of subtest scores, normed

against a reference population so that the average is 100 and the standard deviation is 15. An IQ of 115 means a person scored higher than approximately 84% of the reference population. An IQ of 85 means they scored higher than approximately 16%.

What the tradition got right

The psychometric tradition has produced five findings that any successor framework must take seriously.

First, reliability. Modern IQ tests are extraordinarily reliable as measuring instruments. Test-retest correlations are typically above 0.90, meaning that a person who takes the same test twice (or takes two well-constructed tests) will receive very similar scores. This level of measurement precision is rare in the social sciences and is a genuine technical achievement.

Second, predictive validity. IQ scores predict a wide range of life outcomes. The correlations are not perfect, but they are real and substantial. IQ predicts academic performance (correlations around 0.50-0.60), job performance across occupations (correlations around 0.25-0.55, higher for more cognitively complex jobs), income, educational attainment, health outcomes, and even longevity. These are not trivial predictions. They mean that cognitive ability, as measured by IQ tests, captures something about a person that matters for how their life unfolds across many domains.

Third, the positive manifold. The fact that cognitive abilities correlate positively with each other is a robust empirical finding. It means that cognitive ability is not a collection of entirely independent modules (as some strong versions of multiple intelligences theory have claimed). There is real common variance, something that cuts across different cognitive tasks, and whatever that something is, it matters.

Fourth, partial heritability. Twin studies and adoption studies consistently show that IQ is substantially heritable, with heritability estimates typically ranging from 0.50 to 0.80 in adults. This means that a significant portion of the variation in IQ within populations is associated with genetic variation. Heritability does not mean immutability (IQ is also influenced by environment, and population-level IQ has risen substantially over the twentieth century, a phenomenon known as the Flynn effect). But it means that individual differences in cognitive ability are not purely the product of educational opportunity or environmental advantage. There is a biological component.

Fifth, practical utility. IQ tests have proven practically useful for clinical diagnosis (identifying intellectual disabilities and learning disorders), educational placement, personnel selection, and research. Their utility is not evidence that they measure everything that matters, but it is evidence that they measure something that matters.

The APA consensus and Gottfredson's definition

In 1996, following decades of public controversy about intelligence research, the American Psychological Association published a task force report, *Intelligence: Knowns and Unknowns*, that attempted to summarize the scientific consensus. The report acknowledged that IQ tests measure important cognitive abilities, that scores are reliable and predict important outcomes, that they are substantially heritable, and that group differences in scores exist but are not well explained by any single factor. The report also acknowledged significant limitations: IQ tests do not capture all forms of intelligence, scores are influenced by education and culture, and the meaning of group differences is contested.

In 1997, Linda Gottfredson published a definition of intelligence signed by 52 researchers: "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings." This definition is worth quoting because it is simultaneously honest and revealing. It is honest because it describes something real: a general cognitive capacity that goes beyond mere test performance. It is revealing because, despite its breadth, it is still entirely cognitive. It mentions reasoning, planning, problem-solving, and comprehension. It does not mention judgment, character, emotional attunement, relational competence, ecological skill, or moral discernment. The things that virtually every tradition in Chapters 4 through 13 placed at the center of human excellence are absent from this definition, not because the signatories denied their importance, but because the psychometric tradition treats them as outside the scope of "intelligence."

What the fair observer must concede

A book that argues for a broader framework must be willing to concede the following points to the psychometric tradition.

IQ tests measure something real. The positive manifold is not an artifact of test construction. It reflects genuine shared variance in cognitive ability. Calling g a statistical artifact or dismissing IQ as "just a test score" is not honest.

IQ scores predict important outcomes. A framework that cannot account for the predictive validity of IQ is incomplete. The successor model proposed in Movement III must explain why cognitive ability predicts academic and occupational outcomes, and it must be able to capture the same predictions at least as well as IQ does.

The heritability of IQ constrains what education alone can accomplish. While environment matters enormously (the Flynn effect proves this), individual differences in cognitive ability are not purely environmental. A framework that assumes all differences in intellectual performance are due to education, culture, or opportunity is empirically wrong.

The psychometric tradition's methodological rigor is a genuine contribution to science. Factor analysis, item response theory, reliability analysis, validity assessment: these are powerful tools for understanding human variation, and they can and should be applied to a broader range of human capacities than the psychometric tradition has traditionally studied.

What the fair observer must also note

Having conceded all of the above, the fair observer must also note the following.

Predictive validity is not the same as conceptual validity. IQ predicts academic performance partly because academic tasks are cognitively demanding, but also partly because IQ tests and academic assessments are constructed from the same cultural-cognitive matrix. Both value abstract reasoning, verbal comprehension, and the ability to perform under timed, decontextualized conditions. The correlation between IQ and academic performance is partly a reflection of shared cognitive demands, but it is also partly a reflection of shared cultural assumptions about what counts as important cognitive work.

The things IQ does not predict are as important as the things it does. IQ is a weak predictor of leadership effectiveness, marital satisfaction, ethical conduct, creative achievement (beyond a threshold), practical problem-solving in everyday contexts, and what most people would recognize as wisdom. These are not trivial omissions. They are dimensions of human functioning that matter enormously for individual and collective wellbeing, and the fact that IQ misses them is evidence of its narrowness, not evidence of their unimportance.

The positive manifold does not prove that g is a single natural kind. The correlations among cognitive tests could be produced by multiple overlapping capacities, by shared environmental influences, or by the way tests are constructed. The move from "tests correlate" to "there is a single thing called general intelligence" is an inference, not an observation, and it is contested within the psychometric tradition itself.

Heritability does not mean genetic determination. Heritability is a population-level statistic that describes how much of the variance in a trait within a specific population, at a specific time, under specific environmental conditions, is associated with genetic variation. It says nothing about the malleability of the trait, the potential for intervention, or the causes of differences between popula-

tions. The frequent confusion of heritability with immutability has caused enormous harm and is scientifically unjustified.

The honest summary

The psychometric tradition discovered that cognitive ability varies across individuals, that this variation is partly heritable, that different cognitive abilities correlate positively with each other, and that cognitive ability as measured by IQ tests predicts important life outcomes. These are real, replicable, important findings. They are the foundation on which any successor framework must build.

But the psychometric tradition also made a choice: it defined intelligence as the kind of thing that its tests could measure. It then used the success of its measurements to argue that intelligence *is* what its tests measure. This circular logic, however institutionally successful, does not establish that *g* is the whole of intelligence, or even its most important dimension. It establishes only that there is a measurable cognitive trait that matters. The question, to which this book has been building since Chapter 1, is whether that trait is the whole story.

It is not. Chapter 18 will show what was left out.

Chapter 18: What Was Amputated

The previous chapter presented the strongest honest case for what the psychometric tradition achieved. IQ tests measure something real, reliable, heritable, and predictive. This chapter presents the other side of the ledger: what the psychometric tradition left out.

The word "amputated" is chosen deliberately. The dimensions of human excellence described in the following pages were not refuted by the psychometric tradition. They were not shown to be illusory or unimportant. They were simply excluded from the operational definition of intelligence, not because the science demanded their exclusion, but because they were difficult to standardize, hard to score on a single scale, and inconvenient for the institutional purposes (mass sorting, selection, and placement) that drove the adoption of intelligence testing. The exclusions were administratively convenient, not scientifically mandated.

This chapter inventories what was lost. Each section names a dimension of excellence that was central to multiple traditions in the cross-civilizational survey and shows that it is absent from, or marginal within, the standard IQ framework. The inventory is not exhaustive: Movement II will develop each dimension in full. But it is sufficient to demonstrate the scale of what the narrowing cost.

Practical judgment

Aristotle's *phronēsis*, Akan *nyansa*, Yoruba *ogbón*, Confucian cultivated judgment, and Binet's own definition of intelligence all center the capacity to judge well in real situations: to perceive the relevant features of a context, to weigh competing considerations, and to act appropriately. This is not abstract reasoning. It is situated, contextual, and dependent on experience, character, and perceptual sensitivity.

IQ tests do not measure this. They present well-defined problems with single correct answers under controlled conditions. Real-world judgment operates on ill-defined problems with multiple possible responses under conditions of uncertainty, time pressure, and incomplete information. The Situational Judgment Test (SJT) literature has attempted to measure practical judgment, and SJTs do predict job performance with incremental validity above IQ. But SJTs remain peripheral to the psychometric mainstream, and no SJT is included in any standard IQ battery.

Binet knew this. He defined intelligence as "judgment, good sense, practical sense, initiative, the faculty of adapting oneself to circumstances." The testing apparatus that followed his work measured reasoning, vocabulary, and processing speed instead.

Moral discernment

Every tradition surveyed in Chapters 4 through 13 treated moral discernment as either a component of excellence or as inseparable from it. The Greek *phronimos* is morally virtuous by definition. The Confucian *junzi* integrates intellectual capacity with ethical conduct. The Yoruba *omolúàbí* lists character (*ìwà*) as the foundational virtue. The Buddhist tradition insists that wisdom without compassion is not yet wisdom. The Islamic *adab* tradition treats ethical refinement as inseparable from intellectual cultivation.

IQ tests are value-neutral by design. They do not measure whether a person uses their cognitive ability well or badly, wisely or foolishly, for the benefit or the harm of others. This was a deliberate choice: the psychometric tradition sought a measure of cognitive capacity that was independent of moral content, so that the same scale could be applied to a saint and a sociopath. From the perspective of virtually every tradition in this survey, this is not neutrality. It is a mutilation of the concept. Intelligence without moral direction is not, in these traditions, excellence. It is mere cleverness, and mere cleverness can be dangerous.

Emotional attunement

The Yoruba *ogbón-inú* (interior/emotional wisdom), the Chinese *xin* (heart-mind that integrates cognition and affect), the Buddhist inseparability of wisdom and compassion, and Hume's insistence that reason is the slave of the passions: all point to an emotional dimension of excellent functioning that IQ tests do not capture.

The emotional intelligence (EI) construct, proposed by Salovey and Mayer in 1990 and popularized by Goleman in 1995, was an attempt to recover this dimension. EI is explicitly presented as a capacity that IQ does not measure: the ability to perceive, understand, use, and manage emotions in oneself and others. The fact that EI had to be invented as a separate construct, outside the IQ framework, is itself evidence of what the narrowing excluded. The Yoruba tradition did not need a separate construct for emotional wisdom; it was already included in *ogbón*. It was the psychometric narrowing that forced the emotional dimension out and then required a new label to bring it back.

Relational and social competence

Ubuntu holds that personhood is constituted through relationships. The Confucian tradition treats the five relationships (*wulun*) as the structure within which virtue is cultivated and expressed. The Yoruba *omolúàbí* includes *òrọ̀ síṣọ̀* (expert speech), *ìtẹ́ríba* (respect), and *inú rere* (goodwill) as core

virtues. Across the survey, the capacity to sustain, repair, and contribute to relationships emerges as a dimension of excellence that is irreducible to cognitive ability.

IQ tests measure individual performance in isolation. They are designed to minimize social interaction: the test-taker works alone, under standardized conditions, with no opportunity for collaboration, negotiation, or relational display. This is the "musician in a soundproof room" problem identified in Chapter 11. The measurement is not wrong. But it systematically excludes the relational dimension that many traditions consider central. Social intelligence research (Thorndike, 1920; Cantor and Kihlstrom, 1987) and interpersonal skill assessment have attempted to fill this gap, but they remain marginal to the psychometric mainstream.

Embodied and ecological skill

Pacific wayfinding, Māori trained memory within the *whare wānanga*, Inuit environmental knowledge, Japanese *dō* disciplines: all treat embodied, situated, ecological skill as a genuine form of intelligence. The navigator who reads stars, swells, wind, and bird flight to find a small island across thousands of miles of open ocean is performing a cognitive feat of extraordinary sophistication. The tea master whose movements embody decades of disciplined practice is expressing a form of intelligence through the body.

IQ tests measure nothing of this. They are paper-and-pencil (or now computer-based) assessments of abstract cognitive performance. They assume that intelligence is expressed through symbolic manipulation (answering questions, solving puzzles) rather than through embodied action in the world. This assumption is not universal. It is a product of the Western philosophical tradition that separated mind from body (Descartes) and treated thought as the essential human activity. The Japanese, Polynesian, and many Indigenous traditions disagree.

Cultivated character and self-governance

The Greek *paideia*, Roman *humanitas*, Confucian self-cultivation, Islamic *adab*, Buddhist ethical training, and Yoruba traditional education all aim at producing a whole person: someone whose cognitive abilities are integrated with self-discipline, emotional composure, moral commitment, and social grace. This is what the book has called "thick formation," and it was the dominant model of education across civilizations for millennia.

IQ tests measure a thin slice of this. They assess cognitive performance at a single moment, under controlled conditions, without reference to how the person has developed over time, what character they have formed, how they conduct themselves in daily life, or what kind of person they are becom-

ing. Executive function research, grit research, and self-regulation research have all attempted to capture aspects of this dimension, but they operate as separate constructs, not as part of a unified intelligence framework. The psychometric tradition's decision to strip intelligence down to cognitive performance, leaving character, self-discipline, and moral formation as separate topics for separate research programs, was the most consequential narrowing of all.

Wisdom

The most conspicuous absence. Wisdom is central to Greek, Buddhist, Confucian, Akan, Yoruba, and Islamic accounts of excellence. It is the culminating perfection in the Buddhist pāramitā framework, the highest intellectual virtue in Aristotle's hierarchy, and the defining quality of the Confucian sage.

Psychometric *g* is not wisdom. It is not even close. Wisdom involves the integration of knowledge with experience, the capacity to act well under uncertainty, an awareness of the limits of one's own knowledge, concern for consequences beyond the immediate situation, and (in many traditions) an orientation toward the wellbeing of others. None of this is captured by IQ tests. The Berlin Wisdom Paradigm, Sternberg's balance theory of wisdom, and other wisdom-research programs have attempted to study this capacity, but wisdom remains almost entirely outside the scope of standard psychometric assessment.

The compensatory inventions

The pattern is clear: for nearly every dimension that the psychometric narrowing excluded, a compensatory construct has been invented to recover it. Emotional intelligence attempts to recover the emotional dimension. Situational judgment tests attempt to recover practical judgment. Executive function research attempts to recover self-regulation. Creativity research attempts to recover generative capacity. Wisdom research attempts to recover integrative judgment. Social intelligence and interpersonal competence research attempt to recover the relational dimension.

These compensatory inventions are not failures. Many of them have produced genuine scientific insights. But their very existence is evidence of the narrowing's cost. If "intelligence" had been defined broadly enough to include judgment, emotion, character, and relationship from the start, these constructs would not have needed to be invented separately. They are the phantom limbs of the amputated dimensions, attempts to feel what was cut away.

The feedback loop, revisited

Chapter 2 introduced the feedback loop argument: the five-step sequence from statistical discovery through naming, measurement apparatus, institutional adoption, to cultural absorption, after which the construct becomes so familiar that its historical contingency is forgotten. Chapter 16 showed how this loop operated in the case of IQ. Chapter 17 showed what the loop produced: a robust, reliable, predictive measurement system.

This chapter has shown what the loop excluded. Once intelligence was operationally defined as "what IQ tests measure," the tests became self-validating. Anything not captured by the tests was, by definition, not intelligence. It became "personality," "character," "emotion," "wisdom," "social skill," or simply "culture." The word "intelligence" was reserved for the measurable cognitive residue, and everything else was exiled to other research programs, other departments, other disciplines.

This is not a conspiracy. It is how operationalism works: define a term by your measurement procedure, and the measurement becomes the reality. But viewed from the perspective of the cross-civilizational survey in Chapters 4 through 14, the result is astonishing. Virtually every tradition in the survey defined excellence as an integrated bundle of cognitive, moral, emotional, social, practical, and (in some cases) embodied-ecological capacities. The psychometric tradition stripped this bundle down to a single cognitive dimension, gave that dimension the name "intelligence," and then treated everything else as if it belonged to a different topic.

The next thirteen chapters (Movement II) will map the full range of dimensions that the traditions tracked, comparing them across cultures and connecting them to modern measurement where it exists. Movement III will then ask whether these dimensions can be reassembled into a framework that is as empirically rigorous as psychometrics but as humanly complete as the traditions that psychometrics displaced.

Chapter 19: Predecessors, Allies, and What This Book Adds

The argument developed across the preceding eighteen chapters, that modern intelligence testing captures only a fraction of what human civilizations have recognized as excellence, is not an argument made for the first time in this book. Several important thinkers and research programs have noticed parts of the problem, proposed partial solutions, and built frameworks that overlap with the one this book will construct. Intellectual honesty requires naming them, acknowledging what they accomplished, and clarifying where this book's contribution begins.

The psychometric reformers: Sternberg, Gardner, Stanovich

The most sustained challenge to the single-score model of intelligence has come from within psychology itself.

Robert Sternberg's triarchic theory (1985) distinguished three aspects of intelligence: analytical (the kind measured by IQ tests), creative (the ability to generate novel and useful ideas), and practical (the ability to adapt to, shape, and select environments). His later work on wisdom proposed that wisdom is the application of intelligence and creativity toward a common good, mediated by values. In *Wisdom, Intelligence, and Creativity Synthesized* (2003), Sternberg argued explicitly that wisdom, not IQ, should be the goal of education. Sternberg has been the most persistent and productive critic of the *g*-only model for four decades. His insistence that practical intelligence and wisdom are genuine cognitive capacities, distinct from and at least as important as analytical intelligence, closely parallels the argument of this book. Where this book differs: Sternberg works within the Western psychological tradition. His evidence base is experimental and psychometric. He does not undertake the cross-civilizational survey that forms the foundation of this book, and his framework is a reform of psychometrics from within rather than a reconstruction grounded in global philosophical traditions. The convergence between his conclusions and the cross-civilizational evidence strengthens both projects.

Howard Gardner's theory of multiple intelligences (1983) is the most famous alternative to *g*. He proposed that human cognitive competence is better described as a set of relatively independent intelligences: linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, intrapersonal, naturalist, and possibly existential. Gardner's framework was enormously influential in education and popular culture. It was also criticized by psychometricians for lacking empirical support: the positive manifold (the fact that cognitive abilities correlate positively) suggests that the intelligences are not as independent as Gardner claimed, and few of his proposed intelligences have been operationalized in ways that yield the kind of psychometric evidence that the field requires.

Where this book differs: Gardner generated his list from Western cognitive science and developmental psychology, supplemented by case studies and neurological evidence. This book arrives at its dimensional framework through a completely different method (a survey of what the world's traditions actually valued and recognized), and the resulting framework differs from Gardner's in important ways. For example, this book treats moral discernment, wisdom, and relational competence as dimensions of excellence, none of which appear in Gardner's classification. More fundamentally, Gardner proposed that his intelligences are biologically based cognitive modules. This book's framework makes no such claim; it proposes dimensions of human excellence that traditions have valued, leaving the question of their neural architecture for empirical investigation.

Keith Stanovich (*What Intelligence Tests Miss*, 2009) offered a precision critique from within psychometrics. He argued that IQ tests miss what he called "rational thinking dispositions": the tendency to think critically, to calibrate beliefs to evidence, to avoid cognitive biases, and to engage in reflective rather than intuitive judgment. He coined the term "dysrationalia" to describe the phenomenon of intelligent people who think irrationally. Stanovich's work is important because it demonstrates, using the psychometric tradition's own methods, that IQ captures only part of what matters about cognitive functioning. Where this book differs: Stanovich's critique is narrower than this book's. He focuses on rational thinking dispositions within the cognitive domain. This book argues that the amputated dimensions extend well beyond rationality to include moral discernment, emotional attunement, relational competence, embodied skill, and wisdom.

The virtue and character tradition: Peterson, Seligman, MacIntyre

A second family of predecessors approached the problem from the direction of character and virtue rather than intelligence.

Christopher Peterson and Martin Seligman's *Character Strengths and Virtues: A Handbook and Classification* (2004) is the most direct structural parallel to this book. They surveyed virtues across cultures and historical traditions, identified twenty-four character strengths organized under six broad virtues (wisdom, courage, humanity, justice, temperance, and transcendence), and built assessment tools (the VIA survey) to measure them. Their project explicitly modeled itself as a counterpart to the *Diagnostic and Statistical Manual of Mental Disorders*: where the DSM classifies what can go wrong with people, the VIA classifies what can go right. Peterson and Seligman's cross-cultural survey of virtues found remarkable convergence across traditions, a finding that directly parallels this book's discovery of the thick-formation convergence. Where this book differs: Peterson and Seligman explicitly excluded intelligence and cognitive ability from their classification, treating them as "talents and abilities" rather than character strengths. This book argues that this separation is

itself a product of the very narrowing it seeks to overcome. The traditions surveyed in Chapters 4 through 13 do not separate cognitive excellence from moral character. The *phronimos*, the *junzi*, the *omolúàbí*, and the *tlamatini* integrate both. This book's framework puts cognitive ability back together with character, treating them as dimensions of a single multi-dimensional profile of human excellence rather than as separate topics for separate disciplines.

Alasdair MacIntyre's *After Virtue* (1981) argued that modern moral philosophy lost its coherence when it abandoned the Aristotelian framework of virtues, practices, and teleological reasoning. The result, MacIntyre claimed, was a moral vocabulary that retained the surface forms of ethical language but had lost the philosophical framework that made those forms intelligible. This book makes a structurally parallel argument about intelligence: the modern psychometric tradition retained the word "intelligence" but lost the philosophical framework (thick formation, integrated excellence, practical judgment embedded in character) that the word originally pointed toward. MacIntyre's diagnosis of moral philosophy's predicament illuminates the parallel predicament in the psychology of intelligence.

The capability theorists: Sen, Nussbaum

Amartya Sen's capability approach, developed across several decades of work in economics and philosophy, argued against single-index measures of human welfare (such as GDP per capita) in favor of attention to multiple, irreducible dimensions of human functioning. What matters, Sen argued, is not income or utility but the real freedoms people have to live the kinds of lives they have reason to value. Martha Nussbaum (*Creating Capabilities*, 2011) developed this into a list of ten central capabilities (including practical reason, affiliation, bodily integrity, and play) that she argued should be guaranteed as a matter of justice.

The capability approach is philosophically allied with this book in two ways. First, it insists on irreducible plurality: human flourishing cannot be captured by a single score, and the dimensions of flourishing are qualitatively distinct and cannot be collapsed into each other. Second, it draws on cross-cultural philosophical traditions (Nussbaum engages Aristotle, Indian philosophy, and others) rather than relying solely on Western empirical psychology. Where this book differs: the capability approach is a theory of justice (what should societies guarantee?) rather than a theory of psychological measurement (what are the dimensions of human excellence and how can they be assessed?). This book's project is closer to measurement than to political philosophy, though the two are not unrelated.

The wisdom researchers: Baltes, Ardelt, Grossmann

A dedicated research tradition has attempted to study wisdom as a psychological construct. Paul Baltes and the Berlin Wisdom Paradigm defined wisdom as "expert knowledge about the fundamental pragmatics of life" and developed performance-based measures of wisdom-related knowledge. Monika Ardelt proposed a three-dimensional wisdom model (cognitive, reflective, and affective). Igor Grossmann developed the "wise reasoning" construct, emphasizing epistemic humility, perspective-taking, and recognition of uncertainty.

These programs are directly relevant to Chapter 28 (Wisdom and Uncertainty Management) in Movement II. They demonstrate that wisdom can be studied empirically without reducing it to IQ. They also demonstrate how difficult it is: wisdom research remains a small field compared to intelligence research, its measures have lower reliability and predictive validity than IQ tests, and there is no consensus definition of wisdom comparable to the Gottfredson definition of intelligence. The difficulty is itself informative: it suggests that wisdom, as the traditions in Chapters 4 through 13 understood it, may be genuinely harder to measure than cognitive ability, which would explain (though not justify) why the psychometric tradition focused on what was easier to measure.

The critique tradition: Gould, Flynn, cross-cultural psychologists

Stephen Jay Gould's *The Mismeasure of Man* (1981, revised 1996) is the classic critique of IQ as reification: the fallacy of treating a statistical abstraction as a concrete thing. Chapter 2 of this book engages Gould's argument fairly, noting both its strengths (the reification point is valid) and its limitations (Gould sometimes overstated the case and his critics have identified real errors). James Flynn, discoverer of the Flynn effect (the steady rise in IQ scores across the twentieth century), argued in *What Is Intelligence?* (2007) that IQ gains reflect changing cognitive habits shaped by modernity, not biological evolution, and that the meaning of IQ scores is culturally contingent.

Cross-cultural psychologists, including Patricia Greenfield, Robert Serpell, and Michael Cole, have for decades documented the cultural specificity of cognitive testing and the ways in which Western tests embed assumptions about what counts as important cognitive work. Serpell's studies in Zambia showed that children who performed poorly on Western tests of spatial intelligence performed excellently on locally valued tasks (wire-bending models) and vice versa. This work directly supports the argument of Chapter 14: recognition methods that seem universal are often culturally specific, and what counts as intelligent performance depends on what a culture values.

What this book adds

Having acknowledged these predecessors, the book's distinctive contribution can be stated precisely.

No previous work has combined all three of the following moves in a single project:

First, a genuine cross-civilizational survey of what human traditions valued as excellence, covering literate and oral traditions across six continents, treating each with philosophical seriousness and using a consistent analytical framework (the four-layer template of vocabulary, phenomena, norms, and recognition).

Second, an honest engagement with psychometrics that concedes what it got right (reliability, predictive validity, the positive manifold, partial heritability) while demonstrating, through the cross-civilizational evidence, that what it measures is a fraction of what human traditions have recognized as important.

Third, a proposed replacement framework (Movement III) that is both empirically grounded and civilizationally informed: a dimensional model of human excellence that accounts for the psychometric findings while recovering the amputated dimensions, and that is accompanied by proposals for assessment that are as rigorous as IQ testing but as humanly complete as the traditions that IQ testing displaced.

Sternberg and Gardner reformed psychometrics from within but did not undertake the cross-civilizational survey. Peterson and Seligman surveyed virtues cross-culturally but separated character from intelligence. Gould critiqued the single-score model but did not rebuild. Nussbaum and Sen worked in political philosophy, not psychological measurement. The wisdom researchers studied one amputated dimension but not the full range. This book is the first attempt to do all three simultaneously, and the cross-civilizational evidence, developed across Chapters 4 through 14, is the foundation on which everything else rests.

Chapter 20: Reasoning and Abstraction

Movement II maps the recurring dimensions of human excellence across traditions and connects them to modern measurement. Each chapter asks four questions about a single dimension: What did the traditions track? How did they value it relative to other dimensions? What does modern science say about it? And what measurement exists, with what limitations?

Reasoning and abstraction is the natural starting point because it is the dimension that the psychometric tradition has studied most thoroughly and that comes closest to what IQ tests actually measure. If this book is to be taken seriously, it must begin by engaging the dimension where the psychometric tradition is strongest.

What the traditions tracked

Nearly every tradition surveyed in Movement I recognized something we can broadly call reasoning: the capacity to draw valid inferences, detect patterns, analyze complex situations, and move from particular observations to general principles. But the forms this recognition took varied enormously, and several traditions developed formal reasoning systems of remarkable sophistication, independently of each other.

Greek philosophy placed reasoning near the center of its account of excellence. Aristotle distinguished *nous* (intuitive grasp of first principles), *epistēmē* (demonstrative knowledge derived through syllogistic reasoning), and *technē* (productive knowledge). His *Prior Analytics* formalized the syllogism as a structure for valid deductive inference, establishing what became the dominant logical tradition in the West for two millennia.

The Nyāya school of classical Indian philosophy developed an independent formal logic of comparable sophistication. The Nyāya five-step syllogism (*avayava*) proceeds through thesis (*pratijñā*: "There is fire on the hill"), reason (*hetu*: "Because there is smoke there"), illustration of concomitance (*udāharaṇa*: "Wherever there is smoke, there is fire, as in a kitchen hearth and unlike a lake"), application (*upanaya*: "This hill is likewise smoky"), and conclusion (*nigamana*: "Thus, there is fire on the hill"). This structure combines induction and deduction, moving from particular to particular via generality. The Nyāya also developed a detailed analysis of inferential fallacies (*hetvābhāsa*), classifying errors in reasoning with a precision that rivals Aristotle's treatment of fallacies. As the Internet Encyclopedia of Philosophy notes, the Nyāya five-step argument was employed for both private reasoning (*svārthānumāna*) and public demonstration aimed at convincing others (*parārthānumāna*), a distinction that anticipates the modern difference between reasoning as a cog-

nitive process and argumentation as a social practice.

The later Mohist school in China, active during the Warring States period (roughly 479-221 BCE), developed what has been recognized as one of the three great classical logical systems of the world, alongside Greek Aristotelian logic and Indian *Hetuvīdyā*. The Mohist Canons contain a theory of analogical argumentation based on *fǎ* (models, standards, paradigms): to determine whether something belongs to a category, one compares it to a standard and checks for relevant similarity. The Stanford Encyclopedia of Philosophy notes that the Mohists developed a pragmatic, non-representational theory of knowledge grounded in practical ability: knowing something is having the capacity to recognize and respond to it appropriately. Their conception of reasoning stressed analogical comparison over deductive syllogism, and their approach to knowledge treated it as a type of reliable capacity rather than a set of justified beliefs. This is a fundamentally different architecture of reasoning from the Greek deductive model, yet equally systematic and equally concerned with distinguishing valid from invalid inference.

Tibetan Buddhist scholasticism made logic and epistemology (*pramāṇa*) the foundation of the Geshe curriculum: students spent three years learning the structure and techniques of formal debate before proceeding to philosophical content. Dharmakīrti's epistemological system, which built on and transformed the Nyāya tradition, became the intellectual language of Tibetan scholarship for centuries.

Islamic *falsafa* gave reasoning a cosmic significance: al-Fārābī's hierarchy of intellects treated the development of rational capacity as the process through which human beings approach the divine Active Intellect. Al-Fārābī was known as the "Second Teacher" after Aristotle precisely because of his elaboration of logical method within an Islamic philosophical framework.

These traditions took reasoning seriously. They developed formal techniques for conducting it, institutional settings for training it, and philosophical frameworks for understanding its place in human cognition. Three independent civilizations (Greek, Indian, Chinese) created formal logical systems without knowledge of each other's work. This is strong evidence that formal reasoning is a genuine human capacity, not a cultural invention of any single tradition. The book does not dispute that reasoning is real, important, and measurable.

The range of valuations

What differs sharply across traditions is how reasoning is valued relative to other dimensions of excellence.

At one end of the range, some strands of the Greek and Islamic traditions treated reasoning as the

highest human capacity: the faculty that distinguishes humans from animals and that, at its fullest development, connects the human mind to the intelligible order of reality. The Enlightenment mainstream (Chapter 15) inherited this elevation of reason and made it the defining feature of human dignity: Kant's mature, self-governing use of reason was the essence of Enlightenment.

At the other end, several traditions treated reasoning as useful but subordinate. The Confucian tradition valued reasoning but embedded it within the broader project of self-cultivation: a person of brilliant reasoning who lacked moral character was not a *junzi* but a clever opportunist. Notably, the Mohists, China's most rigorous logicians, grounded their entire logical system in ethical concerns: the purpose of correct reasoning was to identify standards (*fǎ*) that would guide action toward collective welfare. Reasoning was not valued for its own sake but for its practical consequences.

The Buddhist tradition treated logical analysis as a necessary stage in the development of wisdom (*cintāmayī-prajñā*, wisdom from reflection) but subordinate to meditative realization (*bhāvanāmayī-prajñā*), which transcends conceptual analysis altogether. The Nyāya school, despite producing India's most sophisticated formal logic, held that the ultimate purpose of correct reasoning was *mokṣa* (liberation from suffering): knowledge of the sixteen categories was pursued not for intellectual satisfaction but because, as the *Nyāya-sūtras* state, salvation is attained through right knowledge. Even India's most rigorously logical tradition embedded reasoning within a soteriological framework.

The Akan and Yoruba traditions valued good reasoning as part of practical wisdom (*nyansa*, *ogbón*) but did not isolate it as a separate faculty to be cultivated independently of character, speech, and social conduct. The Ubuntu tradition questioned whether reasoning (or any cognitive capacity) can be meaningfully assessed in an individual extracted from their relational context.

The range, then, runs from "reasoning is the essence of human excellence" to "reasoning is one component, valuable but incomplete without moral, emotional, and relational dimensions." The psychometric tradition's decision to make reasoning the centerpiece of intelligence testing places it at one end of this range, not in the middle.

What modern science says

Modern cognitive science has extensively studied reasoning, and several findings complicate the picture that the psychometric tradition presents.

First, human reasoning is not a single, unified capacity. Dual-process theories, developed by Keith Stanovich and Richard West and popularized by Daniel Kahneman in *Thinking, Fast and Slow* (2011), distinguish between two types of cognitive processing. Type 1 (System 1) is fast, automatic, intuitive, and often emotionally inflected. Type 2 (System 2) is slow, deliberate, effortful, and

dependent on working memory. IQ tests primarily measure System 2 capacity: the ability to engage in deliberate, effortful reasoning under controlled conditions. But much of what people do in daily life depends on System 1: pattern recognition, emotional appraisal, rapid social judgment, and the deployment of learned expertise.

Stanovich's concept of "dysrationalia" demonstrates that System 2 capacity (as measured by IQ) does not guarantee rational behavior. People of high IQ can be systematically irrational: they can fail to calibrate their beliefs to evidence, fall prey to cognitive biases, and make poor decisions because they lack what Stanovich calls "rational thinking dispositions." As Evans and Stanovich (2013) caution, "perhaps the most persistent fallacy in the perception of dual-process theories is the idea that Type 1 processes are responsible for all bad thinking and that Type 2 processes necessarily lead to correct responses." Both systems can produce rational or irrational outputs, depending on the context.

Second, abstract reasoning ability (fluid intelligence, *Gf*) is distinguishable from knowledge-based reasoning (crystallized intelligence, *Gc*). *Gf* peaks in early adulthood and declines gradually thereafter. *Gc* continues to increase into middle age and declines later, if at all. This developmental pattern suggests that what people ordinarily call "getting wiser with age" involves the growth of *Gc* (accumulated knowledge, experience, and judgment) even as *Gf* (raw processing speed and pattern detection) declines. The traditions that emphasized wisdom over raw reasoning were tracking something that the fluid/crystallized distinction confirms empirically.

Third, reasoning performance is strongly influenced by context, framing, and content domain. People who reason well about familiar content often reason poorly about unfamiliar content. Experts in a domain reason far better within that domain than outside it. This domain-specificity of reasoning is consistent with the traditions (Akan *nyansa*, Aristotelian *phronēsis*, Mohist *fǎ*-based reasoning) that treated good reasoning as inseparable from experience and situational knowledge rather than as an abstract, domain-general capacity.

Fourth, the positive manifold (the fact that different cognitive tests, including reasoning tests, correlate positively) is real and robust. This means that reasoning ability is not entirely independent of other cognitive capacities. But the correlations are far from perfect, leaving substantial room for distinct dimensions. A person can be an excellent abstract reasoner and a poor practical judge, a brilliant pattern-detector and an emotionally obtuse colleague. The positive manifold establishes that cognitive abilities share common variance; it does not establish that a single ability explains all cognitive performance.

Current measurement

Reasoning is the best-measured dimension in the entire framework this book proposes. Raven's Progressive Matrices, the gold standard of fluid intelligence testing, measures abstract pattern detection using nonverbal, culture-reduced stimuli. The WAIS (Wechsler Adult Intelligence Scale) includes multiple reasoning subtests: matrix reasoning, figure weights, arithmetic reasoning, and similarities. The Woodcock-Johnson Tests of Cognitive Abilities include fluid reasoning clusters.

These instruments are reliable (test-retest correlations above 0.85), well-normed, and widely used. The measurement is genuine. Abstract reasoning ability varies across individuals, can be assessed reliably, and predicts important outcomes. This book does not dispute any of this. What the book disputes is the inference that because reasoning is well measured, it is the most important dimension of human excellence, or that a framework adequate to reasoning is adequate to excellence as a whole. It is worth noting that the three great classical logical traditions (Greek, Indian, Chinese) developed strikingly different approaches to reasoning. Greek logic prioritized deductive syllogism. Indian Nyāya logic combined induction and deduction in a five-step structure that includes empirical illustration. Chinese Mohist logic prioritized analogical reasoning based on standards and resemblance. Modern psychometric reasoning tests are almost entirely in the Greek deductive tradition: they test pattern completion, logical inference, and abstract rule application. They do not test analogical reasoning of the Mohist type, nor the empirically grounded inference of the Nyāya type. Even within the dimension of reasoning, the measurement is narrower than the cross-civilizational evidence suggests it should be.

Frontier questions

Several frontier developments are worth noting.

AI benchmarking has forced the question of what "reasoning" means in a new way. Large language models can now pass many tests of abstract reasoning that were once considered measures of human intelligence. This does not necessarily mean that AI systems reason in the way humans do (there are vigorous debates about whether pattern-matching constitutes genuine reasoning), but it challenges the assumption that high scores on reasoning tests are a unique marker of human cognitive achievement.

Computational approaches to cognitive modeling, including predictive processing frameworks and Bayesian models of inference, offer more precise accounts of what the brain does when it "reasons." These models suggest that reasoning is not a single operation but a family of processes (hypothesis generation, evidence evaluation, belief updating, causal inference) that can be dissociated and

measured separately.

The rational reinterpretation of dual-process theory, developed by Lieder, Griffiths, and colleagues, provides a normative justification for why humans have two reasoning systems. Their computational analysis shows that in variable environments, the optimal cognitive architecture involves one system that is accurate but costly to use and another that is fast but sometimes less accurate, mirroring the Type 1/Type 2 distinction. This suggests that "irrational" behavior may sometimes reflect optimal resource allocation under computational constraints, not cognitive failure.

Summary for the dimensional framework

Reasoning and abstraction is a real, important, measurable dimension of human excellence. It is the dimension that the psychometric tradition has studied most thoroughly and measures most effectively. The cross-civilizational evidence confirms its importance: at least three civilizations independently developed formal logical systems, and every tradition recognized some form of reasoning as valuable.

But the cross-civilizational evidence also places reasoning within a larger context. Most traditions valued reasoning as one component of excellence, not as its essence. Many traditions (including the Nyāya and Mohist logicians themselves) treated reasoning as subordinate to practical judgment, moral discernment, or integrative wisdom. Even the most rigorously logical traditions embedded reasoning within ethical or soteriological frameworks, refusing to treat it as a value-neutral cognitive engine. And modern dual-process theory confirms that the deliberate, effortful reasoning measured by IQ tests is only one type of cognitive processing, operating alongside a faster, intuitive system that is at least as important for real-world functioning.

Reasoning is the first dimension in the framework, not the only one, and not necessarily the most important. The next eleven chapters will map the dimensions that the psychometric tradition either marginalized or excluded entirely.

Chapter 21: Learning and Plasticity

Chapter 20 established that reasoning is real, important, and well measured. But reasoning tests, like all standard intelligence tests, measure what a person can do right now, under standardized conditions, at a single moment. They do not measure how quickly or flexibly that person can learn something new. This distinction, between current performance and learning potential, is one of the most important in the psychology of intelligence, and it maps onto a dimension of excellence that nearly every tradition in the survey recognized but that the psychometric mainstream has largely neglected.

What the traditions tracked

The capacity to learn, to change, to develop new competencies, and to adapt to novel challenges was valued across the traditions surveyed in Movement I, though it was rarely isolated as a separate dimension. Instead, it was embedded in broader frameworks of human development.

Confucian self-cultivation (*xiuyang*) is fundamentally a theory of learning and development. The *junzi* is not born excellent; they become excellent through decades of disciplined study, reflection, and practice. The Confucian tradition emphasizes that the capacity for moral and intellectual growth is universal (Mencius's four sprouts), that learning requires sustained effort and proper guidance (the teacher-student relationship), and that the process is lifelong. The Imperial Examination system (*keju*) was, among other things, a massive institutional bet on the idea that learning capacity matters: it assumed that anyone who studied the classics deeply enough could develop the judgment needed to govern, regardless of birth.

The Buddhist three-form *prajñā* model (Chapter 8) is explicitly developmental: wisdom progresses from hearing or learning (*śrutamayī*) through reflection (*cintāmayī*) to meditative realization (*bhāvanāmayī*). This is not merely a classification of types of wisdom. It is a developmental sequence in which earlier forms are necessary foundations for later ones, and in which the highest form (experiential realization) transcends the cognitive processes (analysis, reflection) that enabled it.

The Japanese *shu-ha-ri* model (Chapter 13) describes a parallel developmental progression in embodied skill: protect the form, break from the form, transcend the form. The apprentice begins by imitating the master exactly. Only after years of disciplined imitation does creative departure become possible. And only after creative departure has been mastered does spontaneous, rule-transcending mastery emerge.

The Māori *whare wānanga* and the Polynesian navigational apprenticeship (Chapter 12) were ex-

tended learning systems that selected students for aptitude and trained them over years, testing progress through demonstrated performance rather than written examination. The Geshe degree (Chapter 8) required approximately twenty years of progressive learning, with each stage building on the last.

What these traditions share is an emphasis on the *process* of learning rather than on a static snapshot of current ability. They cared about how people develop over time, how they respond to instruction, and how they progress through stages of mastery. Standard IQ tests capture none of this.

The Vygotsky revolution: learning potential and the zone of proximal development

The most important theoretical framework for understanding learning as a dimension of excellence comes from the Soviet psychologist Lev Vygotsky (1896-1934). Vygotsky argued that static tests of current performance miss the most important thing about a learner: their potential for development. He proposed the concept of the Zone of Proximal Development (ZPD), defined as the distance between what a learner can do independently and what they can achieve with appropriate guidance from a more knowledgeable other.

The ZPD is not just a pedagogical tool. It is a philosophical claim about the nature of intelligence: that what a person can learn (with proper support) is as important as what they already know, and that intelligence is better understood as a capacity for development than as a fixed endowment. This aligns with the developmental emphasis found in the Confucian, Buddhist, and apprenticeship traditions, all of which treated the capacity for growth as central to what they valued.

Reuven Feuerstein extended Vygotsky's insight into practical assessment through his Learning Potential Assessment Device (LPAD) and his theory of Structural Cognitive Modifiability. Feuerstein argued that intelligence is dynamic and modifiable, not static. His assessment procedure uses a test-teach-retest format: the examiner first determines what the student can do independently, then provides mediated learning experiences (guided instruction, hints, prompts, and scaffolding), and finally reassesses to see how much the student has learned. The difference between pre-teaching and post-teaching performance is the measure of learning potential.

Dynamic assessment, as this approach is now called, has been found to be more accurate than static IQ tests in reflecting children's learning potential, especially for minority, linguistically diverse, and learning-disabled children. This is significant because it directly addresses one of the most serious criticisms of standard IQ testing: that static tests penalize children whose prior educational experiences have been limited, confusing lack of opportunity with lack of ability.

Why learning potential is neglected in psychometrics

If learning potential is important and measurable, why is it not part of standard intelligence testing?

The answer is partly methodological and partly institutional. Dynamic assessment is harder to standardize than static testing. It requires trained examiners who can adjust their mediation to the individual learner. Scoring depends on the quality of the examiner-learner interaction, which introduces variability that threatens the reliability that psychometricians prize. From a psychometric standpoint, dynamic assessment raises specific challenges: reliability scores depend on examiner behavior, standardization requires carefully specified teaching protocols, and traditional item response models assume that responses are independent of prior feedback, which does not hold when mediation occurs. These are real challenges, but they are challenges of method, not of substance. The fact that learning potential is harder to measure than current performance does not mean it is less important. It means that the psychometric tradition's tools are better suited to measuring static snapshots than dynamic processes. The tradition measured what it could measure easily, and then defined intelligence as what it measured. Learning potential fell outside the definition not because it is unimportant but because it is methodologically inconvenient.

What modern science says

Several lines of research support the importance of learning capacity as a dimension of excellence.

Research on cognitive plasticity shows that the brain retains the capacity for structural and functional change throughout life, though the degree of plasticity varies across brain regions, developmental stages, and individuals. Individual differences in cognitive plasticity, measured by how much people improve with training, predict outcomes above and beyond baseline performance. Some individuals show large training gains; others show minimal improvement from the same training. These differences in learning rate are at least partly independent of static IQ scores.

Research on learning curves demonstrates that the shape of individual learning trajectories, how quickly someone improves, whether they reach plateaus, how they respond to feedback, carries information about cognitive functioning that a single-point assessment cannot capture. Modeling individual learning curves, tracking intra-individual change across training sessions and assessing inter-individual differences in that change, offers finer-grained knowledge of cognitive capacity than static test scores.

The Flynn effect (the secular rise in IQ scores across the twentieth century) is itself evidence that learning capacity, at the population level, responds to environmental change. The effect is largest on

fluid intelligence measures (Raven's Progressive Matrices), which are supposedly the most "culture-fair" and the least dependent on learning. The fact that the most abstract reasoning scores have risen the most dramatically suggests that even *Gf* is partly a product of learning environments, not a fixed biological endowment.

Frontier questions

The stability-plasticity dilemma, well known in both neuroscience and AI research, poses a fundamental question: how can a learning system acquire new knowledge without catastrophically forgetting what it already knows? The human brain solves this through complementary memory systems (the hippocampus for rapid encoding, the neocortex for slow consolidation), developmental sensitive periods that progressively reduce plasticity in early-developing systems, and sleep-dependent memory consolidation.

AI systems struggle with this problem. Recent research published in *Nature* has demonstrated that standard deep learning methods systematically lose their ability to learn with extended training on new data, a phenomenon called "loss of plasticity." Plasticity is maintained only by algorithms that continually inject diversity into the network. This finding highlights that sustained learning requires mechanisms beyond simple optimization, a point that resonates with the cross-civilizational traditions' emphasis on the importance of guided instruction, mentorship, and structured developmental environments.

Computational phenotyping of learning rates, using AI-based modeling to characterize individual differences in how people learn, is an emerging frontier. If learning rate, flexibility, and trajectory shape can be reliably measured through computational models, this would provide the methodological tools that dynamic assessment currently lacks: scalable, standardized measures of learning potential that complement rather than replace static IQ scores.

Summary for the dimensional framework

Learning and plasticity is a dimension of human excellence that was valued by virtually every tradition in the survey, that has strong theoretical support from Vygotsky and the dynamic assessment tradition, and that is increasingly supported by neuroscience and computational research. It is currently undermeasured relative to its importance, largely for methodological rather than substantive reasons.

A framework for human excellence that includes only static performance and excludes learning potential is like a photograph that captures a runner at a single instant but tells you nothing about their

acceleration, their endurance, or their capacity to improve. The static snapshot is real and informative. But it is not the whole picture, and for many purposes (education, development, personnel selection for learning-intensive roles), the capacity to learn may matter more than current performance.

Chapter 22: Knowledge and Memory

Chapter 20 covered reasoning: the capacity to solve novel problems. Chapter 21 covered learning: the capacity to acquire new abilities. This chapter covers what you already know: the organized body of knowledge, skills, and expertise that a person has accumulated over a lifetime, and the memory systems that make it available for use. Of all the dimensions in this framework, this may be the one where the psychometric tradition's value hierarchy departs most dramatically from the near-universal consensus of human civilizations.

What the traditions tracked

Every tradition surveyed in Movement I placed a high premium on the retention, organization, and transmission of culturally essential knowledge. In many traditions, this was the primary form of intellectual excellence, valued above novel reasoning.

The Confucian tradition centered education on the mastery of canonical texts: the *Analects*, the *Book of Changes*, the *Spring and Autumn Annals*, the *Book of Rites*, and the *Book of Songs*. The Imperial Examination (*keju*) tested candidates primarily on their deep knowledge of these classics and their ability to apply classical learning to questions of governance and ethics. A candidate who could reason brilliantly but had not mastered the textual corpus would fail. For over thirteen centuries, the world's largest meritocratic examination system treated retained, organized, applicable knowledge as the central qualification for governing.

The Māori *whare wānanga* (Chapter 12) required students to memorize vast bodies of *whakapapa* (genealogy), *karakia* (prayers), *waiata* (songs), and oral histories, using genealogical frameworks as mnemonic architecture. The *tohunga* (knowledge experts) held the accumulated knowledge of the tribe in their memories and were responsible for transmitting it accurately to the next generation. These were not casual recollections but precise, verbatim preservation of culturally essential material across generations, without writing. The cognitive demands were comparable to memorizing the equivalent of several books, complete with structural relationships, cross-references, and contextual significance.

Pacific navigators (Chapter 12) maintained in memory an even more remarkable body of knowledge: the Hawaiian star compass, a mental construct dividing the horizon into thirty-two houses; the rising and setting points of hundreds of celestial bodies across seasons; ocean swell patterns, wave refraction signatures, and cloud formations associated with specific islands; bird flight patterns and marine life indicators. This knowledge was held in mind simultaneously during voyages lasting weeks, in-

tegrated in real time with continuous environmental observation. The navigator's memory was not a passive store of facts. It was an active, multimodal system for orienting and guiding action under life-or-death conditions.

The Islamic *ḥāfiẓ* (one who has memorized the entire Qur'an) was and remains a recognized figure of intellectual and spiritual achievement. Hadith scholarship required memorizing not only texts but the chains of transmission (*sanad*) through which each text had been passed down, so that the reliability of the knowledge could be assessed through the trustworthiness of every link in the chain. The *ijazah* system (Chapter 9) was built entirely around the authorized transmission of specific bodies of knowledge, treating the accurate preservation and passing-on of learning as the highest intellectual responsibility.

The Tibetan Geshe curriculum (Chapter 8) required memorization of five major Indian Buddhist texts and their commentaries. Students began each day with memorization sessions starting around 5:30 a.m., committing key passages to memory before engaging in debate. The ability to recall and deploy textual material instantly, in the heat of dialectical exchange, was a prerequisite for philosophical reasoning, not a substitute for it. Memory and reasoning were integrated, not separated.

The Yoruba Ifá tradition (Chapter 10) preserved a vast corpus of oral literature organized into 256 *odù*, each containing multiple verses. The *babaláwo* was trained to memorize and deploy this material in divination and counsel. The cognitive achievement involved not only retaining the verses but knowing when and how to apply the right verse to a particular situation, integrating memorization with interpretive judgment.

Crystallized intelligence in psychometrics: the *Gf/Gc* distinction

In 1943, Raymond Cattell proposed what became the most influential structural distinction in intelligence research: the separation of general intelligence into fluid intelligence (*Gf*) and crystallized intelligence (*Gc*). *Gf* is the capacity to reason and solve novel problems independently of prior knowledge. *Gc* is the ability to understand and utilize culturally valued knowledge acquired through experience and education.

Cattell's investment theory proposes that *Gc* is the product of *Gf* invested over time: individuals use their fluid intelligence in learning experiences, gradually building their store of crystallized knowledge. Higher *Gf* enables faster knowledge acquisition, so people with high *Gf* tend to accumulate more *Gc* over time. But the correlation is far from perfect. A highly educated person with modest *Gf* may have very high *Gc*; a poorly educated person with high *Gf* may have low *Gc*. Education, opportunity, motivation, and what Cattell called the "common learning investment" (time, interest,

and memory) all shape the development of *Gc* independently of *Gf*.

The two follow strikingly different developmental trajectories, and this is one of the most robust findings in cognitive aging research. *Gf* peaks in the mid-twenties and declines gradually thereafter, correlating with the efficiency of prefrontal-parietal networks and working memory capacity. *Gc* continues to increase well into middle age and remains stable or even grows into the sixties and seventies in healthy adults. This divergence explains why older professionals often outperform younger colleagues on domain-specific tasks despite lower fluid processing speed: what they have lost in raw computational power they have gained, many times over, in accumulated knowledge and expertise. This is the "classic aging pattern" identified by Horn and Cattell (1967).

Within the Cattell-Horn-Carroll (CHC) model, both *Gf* and *Gc* are broad abilities at Stratum II, beneath a general *g* factor at Stratum III. Both contribute to the Full Scale IQ. But the psychometric tradition has consistently treated *Gf* as more fundamental: Raven's Progressive Matrices, a nearly pure *Gf* measure, is widely regarded as the best single measure of *g*, and the Fluid Reasoning factor loads near 1.0 on *g* in most factor analyses.

Current measurement: what the tests actually do

Gc is measured by several types of instruments, each capturing a different aspect of accumulated knowledge.

On the WAIS (Wechsler Adult Intelligence Scale), the Verbal Comprehension Index includes three core subtests that are primary indicators of *Gc*. The **Vocabulary** subtest asks the examinee to define words of increasing difficulty, presented both orally and visually; responses are scored 0, 1, or 2 for the quality and precision of the definition. It is one of the strongest single predictors of overall IQ, with correlations of approximately $r = .75$ with Full Scale IQ. The **Information** subtest asks general knowledge questions spanning geography, history, science, and culture. Research has found it to be one of the most stable measures across the lifespan, correlating positively with gray matter volume even after controlling for age and education, making it a useful proxy for cognitive reserve. The **Similarities** subtest, which asks how two concepts are alike, straddles the *Gc/Gf* boundary: it draws on accumulated knowledge but also requires abstract verbal reasoning.

The Wechsler Individual Achievement Test (WIAT) measures acquired knowledge in specific academic domains: reading comprehension, mathematical problem-solving, written expression, and listening comprehension. Unlike the WAIS, which measures cognitive *capacity*, the WIAT measures academic *attainment*, the knowledge and skills a person has actually acquired through education.

At the international level, two large-scale assessments provide comparative data on knowledge and

skills across populations.

The **Programme for International Student Assessment (PISA)**, administered by the OECD, tests approximately 600,000 fifteen-year-olds across more than 90 countries in reading, mathematics, and science. PISA does not test rote memorization of facts; it measures the ability to apply knowledge and skills to real-world problems. Its results reveal substantial cross-national variation that correlates with educational investment, teaching quality, and socioeconomic conditions. PISA has challenged deeply embedded educational practices (such as early tracking of students into vocational or academic pathways) and has become a significant driver of educational policy worldwide.

The **Programme for the International Assessment of Adult Competencies (PIAAC)**, also coordinated by the OECD, assesses adults aged 16-65 across more than 30 countries in literacy, numeracy, and adaptive problem-solving. PIAAC results show that adult skill levels vary substantially both between and within countries, that educational attainment does not fully predict adult competency (some countries with high attainment rates have modest skill levels), and that skills can decline in adulthood without continued use. Between 2017 and 2023, average scores for U.S. adults decreased in both literacy and numeracy, suggesting that knowledge maintenance requires ongoing engagement. Notably, Japan and Finland consistently outperform other countries on PIAAC measures, despite very different educational philosophies.

These international assessments demonstrate several things relevant to this chapter. First, knowledge and skill levels vary enormously across populations, driven largely by educational opportunity, not by innate cognitive differences. Second, knowledge is not a permanent acquisition: it requires maintenance and use. Third, the relationship between educational attainment (years of schooling) and actual competence is imperfect, confirming that Gc depends on the quality, not just the quantity, of educational investment.

The value hierarchy inversion

The psychometric tradition has consistently treated Gf as more fundamental than Gc . Cattell's investment theory frames Gc as a *product* of Gf invested over time, implying that Gf is the cause and Gc the effect. Raven's Progressive Matrices, a nearly pure Gf measure, is widely regarded as the best single measure of g .

The cross-civilizational evidence challenges this hierarchy. Most of the traditions surveyed in Movement I valued accumulated knowledge, expertise, and the ability to deploy learned material wisely *more highly* than novel problem-solving. The Confucian sage was wise because of what they had learned and integrated over a lifetime. The Māori *tohunga* was revered for the knowledge they held

in trust for the community. The Islamic *ḥāfiẓ* was honored for an extraordinary feat of memorization and devoted transmission. The Yoruba *babaláwo* was valued for command of the Ifá corpus and the ability to apply it with judgment.

These traditions would find the psychometric hierarchy baffling. To rank novel problem-solving above accumulated wisdom is to privilege the cognitive mode of the young and inexperienced over the cognitive mode of the old and wise. It is to treat the capacity to figure things out from scratch as more important than the capacity to draw on a lifetime of accumulated understanding. The developmental trajectories confirm what the traditions intuited: *Gf* peaks at 25 (the age of mathematical breakthroughs and startup founders), but *Gc* peaks decades later (the age of statesmanship, judicial wisdom, clinical expertise, and scholarly synthesis). For the longest portion of adult life, the dimension that the psychometric tradition subordinates is the one that matters most.

Memory: not a subordinate function

In the CHC model, short-term memory (*Gsm*) and long-term retrieval (*Glr*) are broad abilities, but they are not typically identified with "intelligence" itself. Memory is what feeds the reasoning engine; it is not the main event.

The cross-civilizational evidence contradicts this subordination. In every oral tradition surveyed, memory is a primary cognitive achievement. The Māori student who memorized the tribal genealogies, the Pacific navigator who held the star compass in mind across weeks of open-ocean voyaging, the Tibetan monk who committed five major texts and their commentaries to memory, the Islamic hadith scholar who memorized thousands of traditions with their chains of transmission: all exercised memory at a level of sophistication and precision that dwarfs anything measured by a digit span test.

The distinction between rote memorization and meaningful retention is important. What these traditions valued was not the ability to repeat a random string of digits. It was the capacity to organize, retain, and retrieve vast bodies of meaningful, structured, culturally significant knowledge over years and decades. This is memory embedded in understanding, memory that serves judgment, memory that enables the wise application of accumulated learning. It is closer to what cognitive psychologists call "expert memory" (the chess master's recall of board positions, the physician's recognition of symptom patterns) than to the short-term memory measured by standard psychometric subtests.

Vocabulary scores, which are among the strongest single predictors of IQ, are themselves measures of memory: they reflect the breadth and depth of word knowledge accumulated over a lifetime. The average educated adult knows approximately 20,000 base words and around 42,000 multiword

meanings. Vocabulary continues to grow well into middle age and does not decline significantly in healthy aging. This is crystallized memory in action: the accumulated residue of a lifetime of reading, conversation, and engagement with the world.

Frontier questions

AI has forced a sharp reconsideration of what "knowledge" means. Large language models contain representations of vast bodies of human knowledge and can retrieve and deploy this information with impressive fluency. But whether they "know" things in the way humans do is a matter of active debate. A human expert's knowledge is embedded in personal experience, emotional associations, practical contexts, and the capacity for judgment about when and how to apply it. Whether AI "knowledge" is the same kind of thing, or merely a statistical shadow of it, is one of the central questions in the philosophy of artificial intelligence. If knowledge in a neural network turns out to be fundamentally different from knowledge in a human mind, this would confirm that the traditions were tracking something that cannot be reduced to information storage and retrieval: they were tracking understanding, not merely data.

The neuroscience of memory consolidation, particularly the complementary roles of the hippocampus (rapid encoding of new experiences) and the neocortex (slow integration into long-term knowledge), provides a biological foundation for the *Gf/Gc* distinction. The hippocampal system supports the kind of rapid learning that *Gf* demands; the neocortical system supports the kind of accumulated, integrated knowledge that *Gc* represents. These are genuinely distinct neural systems with different developmental trajectories, supporting the psychometric finding that *Gf* and *Gc* are distinguishable factors with different lifespans.

Summary for the dimensional framework

Knowledge and memory is a dimension of human excellence that was valued by every tradition in the survey and that corresponds to the well-established psychometric construct of crystallized intelligence (*Gc*). It is measured by vocabulary and information subtests on the WAIS, by achievement tests like the WIAT, and at the international level by PISA (for students) and PIAAC (for adults). These instruments reveal that knowledge and skill levels vary enormously across populations, are strongly shaped by educational opportunity, require maintenance through continued use, and continue to grow well into middle age.

The psychometric tradition's subordination of *Gc* to *Gf*, while defensible on statistical grounds, inverts the value hierarchy of most human traditions. A framework for human excellence that treats

knowledge as merely the product of reasoning, or memory as merely a supporting function, misses the insight that virtually every tradition shares: that what you know, and how wisely you can deploy it, matters at least as much as how quickly you can figure out something new.

Chapter 23: Practical Judgment

If the previous three chapters covered what the psychometric tradition measures well (reasoning, learning capacity, and accumulated knowledge), this chapter enters the territory where the tradition is weakest and the cross-civilizational evidence is strongest. Practical judgment, the capacity to act well in messy, real-world situations where the rules do not fully determine what to do, is arguably the most universally valued dimension of human excellence in the entire survey. It is also one of the hardest to measure.

What the traditions tracked

Practical judgment occupies the center of more traditions' accounts of excellence than any other single dimension.

Aristotle's *phronēsis* (Chapter 4) is the master concept here: practical wisdom, the capacity to perceive the relevant features of a particular situation, weigh competing considerations, and act in ways that are appropriate and beneficial. *Phronēsis* cannot be reduced to the application of rules because no rule can anticipate every situation, and the wise person's judgment is precisely the capacity to see what this situation requires even when no rule is available. Aristotle noted that young people can be mathematicians but not practically wise, because *phronēsis* requires the kind of experience that only accumulates over time.

Binet, the inventor of the intelligence test (Chapter 16), defined intelligence in terms that are strikingly close to *phronēsis*: "judgment, otherwise called good sense, practical sense, initiative, the faculty of adapting oneself to circumstances." David Wechsler, who created the most widely used adult intelligence test, defined intelligence as "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment." Both definitions center practical effectiveness and contextual adaptation, but the tests these men created measured abstract reasoning, vocabulary, and processing speed instead. The dimension they named as central proved too difficult to operationalize.

The Akan concept of *nyansa* (Chapter 10) is wisdom understood as the ability to make good practical use of what one knows: to speak appropriately, judge soundly, and contribute to communal harmony. The deployment of proverbs in council, selecting the right condensed principle for this particular situation and applying it with timing and social sensitivity, is a form of practical judgment that parallels *phronēsis* in structure even though it operates through a completely different cultural medium.

Confucian cultivated judgment (Chapter 6) similarly centers the capacity to act well in the specific relational and ethical circumstances one faces. The Imperial Examination tested candidates not only on textual knowledge but on their ability to apply classical learning to questions of governance: a practical task that required judgment about how abstract principles should inform concrete policy.

Sternberg's practical intelligence (Chapter 19) is the modern Western psychological construct most directly aimed at capturing this dimension. Developed with Richard Wagner, the concept centers on tacit knowledge: what one needs to know to succeed in a particular environment that is not explicitly taught and often not even verbalized. Sternberg defines practical intelligence as the ability to adapt to, shape, and select environments effectively. His tacit knowledge inventories present realistic scenarios with no single correct answer and ask respondents to rate the quality of various courses of action. Research has found that tacit knowledge scores predict managerial performance even after controlling for IQ, and that they correlate only modestly with IQ scores (approximately $r = .20$), suggesting that practical intelligence captures something real that standard IQ tests miss.

However, the status of practical intelligence as a construct independent from general cognitive ability remains contested. Gottfredson (2003) argued that tacit knowledge inventories overlap substantially with domain-specific job knowledge and that the evidence for practical intelligence as truly separable from g is weaker than Sternberg claims. This debate is important because it tests whether the dimension the traditions valued most can be captured by the psychometric tradition's methods at all, or whether practical judgment is a kind of competence that resists the standardization, decontextualization, and single-score scoring that psychometric measurement requires.

Current measurement: Situational Judgment Tests

The most widely used attempt to measure practical judgment in applied settings is the Situational Judgment Test (SJT). SJTs present realistic scenarios, typically drawn from workplace contexts, and ask respondents to select or rank the most and least effective responses from a set of options. They are classified as low-fidelity simulations: they do not place the respondent in an actual situation but ask them to exercise judgment about what would be effective.

The most comprehensive meta-analysis of SJT validity, conducted by McDaniel, Hartman, Whetzel, and Grubb (2007), analyzed 118 validity coefficients and found an overall criterion-related validity of $r = .26$ for predicting job performance. This is below cognitive ability tests ($r = .51$), structured interviews ($r = .51$), and integrity tests ($r = .41$), but above most individual personality measures. SJTs demonstrate persistent incremental validity above cognitive ability and personality measures: they predict something about job performance that neither IQ nor personality tests capture.

SJTs are a measurement method rather than a measure of a single construct. An SJT score reflects a blend of cognitive ability (understanding the scenario), personality (preferred interpersonal style), job knowledge (knowing what works in practice), and practical judgment (evaluating trade-offs among imperfect options). This construct heterogeneity is both the strength and the limitation of SJTs. It means they capture something multifaceted about real-world competence. But it also means that a single SJT score is diagnostically ambiguous: a low score could reflect deficits in cognition, personality, experience, or values.

Two features of SJTs are significant for this book's argument. First, SJTs with knowledge-based instructions ("What is the best response?") correlate more strongly with cognitive ability ($r = .35$), while SJTs with behavioral tendency instructions ("What would you most likely do?") correlate more with personality traits ($r = .19$). This shows that the way judgment is assessed changes what it measures, a point consistent with the cross-civilizational evidence that practical judgment is an integrated capacity rather than a single isolable trait. Second, SJTs show substantially lower adverse impact than cognitive ability tests: Black-White standardized mean differences on SJTs are approximately $d = 0.38$, compared to approximately $d = 1.0$ for cognitive ability tests. This means that a measurement approach closer to what the traditions valued also produces fairer outcomes across demographic groups.

Beyond SJTs, clinical judgment, legal reasoning, and military decision-making all involve forms of practical judgment that are assessed through case-based methods, supervisor evaluations, and performance in realistic simulations rather than through standardized tests. Medical school admissions increasingly use SJTs and multiple mini-interviews to assess candidates' judgment and interpersonal skills alongside academic ability. These developments represent an institutional recognition that practical judgment matters and that standard cognitive tests do not capture it.

Why practical judgment resists standardization

Kant's analysis of judgment (Chapter 15) identified the philosophical reason: judgment is the capacity to subsume a particular case under a general rule, and this capacity cannot itself be taught by rules, because any rule for applying rules would itself require judgment to apply. This creates an irreducible dependence on experience, perception, and the kind of situated understanding that no fixed procedure can replace.

The cross-civilizational evidence confirms Kant's insight. Every tradition that valued practical judgment treated it as inseparable from experience. Aristotle said young people cannot be practically wise. The Akan recognized wisdom through decades of observed conduct. The Confucian sage achieved judgment through a lifetime of cultivation. The *omolúàbí* was someone whose entire for-

mation, from childhood to elder status, had been oriented toward sound judgment and virtuous conduct.

Standardized testing, by design, eliminates context: it presents the same problem to every test-taker under controlled conditions. But practical judgment is constituted by context: it is the capacity to perceive what this particular situation requires, given this particular set of constraints, in this particular relational and environmental setting. Strip away the context and you strip away much of what judgment consists of. This is not a technical limitation waiting to be overcome by better test design. It is a structural feature of the dimension itself.

Frontier questions

Naturalistic neuroscience, studying cognition in real-world environments rather than in laboratory settings, is beginning to investigate how the brain supports situated judgment. Brain imaging in ecological contexts (decision-making during social interactions, navigation in complex environments, moral judgment under uncertainty) reveals neural processes that differ from those engaged by abstract reasoning tasks in scanners. This suggests that practical judgment may involve brain networks and processes that are genuinely distinct from those measured by standard cognitive tests.

AI alignment presents a striking contemporary parallel to the practical judgment problem. The challenge of aligning AI systems with human values is, at its core, a judgment problem: how to specify what "acting well" means in situations that cannot be fully anticipated by rules. The difficulty AI researchers face in encoding practical wisdom into formal systems mirrors the difficulty psychometricians face in measuring it. Both problems suggest that judgment is a form of competence that exceeds any formal specification, a conclusion that Aristotle, Kant, and the Akan sage tradition would all endorse.

Summary for the dimensional framework

Practical judgment is the dimension where the gap between what the traditions valued and what psychometrics measures is largest. It is central to Aristotle's *phronēsis*, Binet's own definition of intelligence, Wechsler's definition, Sternberg's practical intelligence, Akan *nyansa*, Confucian cultivated judgment, and the Yoruba *omolúàbí*. It is what most people mean when they call someone "wise" in everyday conversation.

SJTs and tacit knowledge inventories represent genuine attempts to measure this dimension, and they demonstrate incremental validity above IQ. But their validity is modest ($r = .26$ for job performance), their construct heterogeneity makes them diagnostically ambiguous, and they remain peripheral to

the psychometric mainstream. The dimension that virtually every human tradition placed at the center of excellence remains at the margins of modern measurement.

This is not because practical judgment is unimportant. It is because practical judgment, by its nature, resists the decontextualization, standardization, and single-score scoring that psychometric measurement requires. A framework for human excellence must find a way to assess this dimension that respects its contextual, experiential, and integrative character rather than trying to force it into a format designed for abstract reasoning.

Chapter 24: Emotional Attunement

This chapter enters territory where the cross-civilizational evidence is strong, the modern measurement landscape is confused, and the underlying philosophical question is profound: Is the capacity to perceive, understand, and manage emotions a form of intelligence, a personality trait, or something that resists both categories? The answer depends on which tradition you ask, and the fact that the modern field of emotional intelligence has not settled on a single answer, after three decades of research, is itself evidence that the psychometric narrowing created a problem it cannot easily solve.

What the traditions tracked

Many of the traditions surveyed in Movement I treated emotional perception, regulation, and expressiveness not as separate from intellect but as part of what it means to know and act well. The modern separation of cognition from emotion, which required the later invention of "emotional intelligence" as a distinct construct, was not a separation these traditions made.

The Chinese concept of *xin* (心, heart-mind) (Chapter 6) is the foundational case. *Xin* is a single faculty that integrates what English separates into cognition and emotion. The Confucian tradition does not ask whether the heart-mind is cognitive or affective; it is both, simultaneously and inseparably. The Mencian "four sprouts" (compassion, shame, deference, judgment of right and wrong) are emotional responses that are also moral cognitions: they are felt and known in the same act. When later neo-Confucians debated how to cultivate *xin*, they were debating how to develop a faculty that thinks and feels as one. The modern Western invention of "emotional intelligence" as a supplement to "cognitive intelligence" would have struck Mencius as solving a problem that should never have been created.

The Buddhist tradition (Chapter 8) makes the strongest philosophical case for the inseparability of wisdom and compassion. *Prajñā* (wisdom) without *karuṇā* (compassion) is not yet wisdom; *karuṇā* without *prajñā* is not yet compassion. Speaking of them as separate things to be developed independently is, in this framework, a pedagogical convenience rather than a description of reality. If the deepest form of understanding is structurally bound to how one relates to others emotionally, then defining intelligence without reference to emotional orientation is not merely incomplete. It is a mistake about the nature of knowing itself.

The Yoruba distinction between *ọgbón-ori* (intellectual wisdom, wisdom of the head) and *ọgbón-inú* (emotional or interior wisdom, wisdom of the belly) (Chapter 10) is the most direct pre-modern parallel to the modern emotional intelligence construct. The Ifá tradition teaches that the person

who fails to make use of their *ogbón-inú* becomes a fool, regardless of their intellectual knowledge. This distinction, made within an oral philosophical tradition centuries before the modern invention of EI, independently identifies the same phenomenon: there is a form of wisdom that is felt rather than reasoned, that operates through emotional sensitivity and interior knowing, and that is essential for sound judgment.

The Stoic tradition presents the counter-case within the Western philosophical heritage. For the Stoics, the wise person achieves *apatheia*: not the absence of feeling, but the rational mastery of emotion. Emotions, in this framework, are judgments that can be correct or incorrect, and the excellent person is one whose emotional responses are rationally governed. This is not the denial of emotion's importance but rather an insistence that emotion should be subordinated to rational self-command. It is a precursor to the modern executive function literature (Chapter 26) rather than to the emotional intelligence literature.

Hume's sentimentalism (Chapter 15) represents the Western philosophical tradition's strongest internal argument for the centrality of emotion: "Reason is, and ought only to be, the slave of the passions." Moral judgment depends on feeling, not on rational demonstration. Hume's position was largely sidelined by the mainstream that produced psychometrics, but it re-emerged in the 1990s when Salovey and Mayer proposed emotional intelligence as a scientific construct.

The invention of emotional intelligence

The term "emotional intelligence" appeared sporadically in academic literature from the 1960s, but the construct as we now understand it was formally introduced by Peter Salovey and John Mayer in 1990. They defined it as "the ability to monitor one's own and others' feelings and emotions, to discriminate among them, and to use this information to guide one's thinking and actions." Daniel Goleman's bestselling *Emotional Intelligence* (1995) brought the concept to a mass audience, though Goleman's popular treatment was considerably less rigorous than the Salovey-Mayer scientific framework.

For this book's argument, the invention of EI is itself evidence of the amputation described in Chapter 18. The Yoruba *ogbón* tradition did not need a separate construct for emotional wisdom; it was already part of wisdom. The Chinese *xin* tradition did not need to supplement cognitive intelligence with emotional intelligence; the heart-mind was always both. It was the psychometric tradition's decision to define intelligence in purely cognitive terms that forced the emotional dimension out, creating a vacuum that the EI construct was invented to fill. Emotional intelligence is a phantom limb of the amputated dimension.

The three-model confusion

One of the most significant problems in EI research is that the term "emotional intelligence" is used to refer to at least three different constructs, measured in fundamentally different ways, with different relationships to existing personality and cognitive ability measures.

Ability EI (Salovey and Mayer's model) treats emotional intelligence as a genuine cognitive ability: the capacity to perceive, use, understand, and manage emotions. It is measured by performance tests, most importantly the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), a 141-item assessment that presents actual emotional problems to solve. The MSCEIT asks respondents to identify emotions in faces and abstract art (perception), describe which feelings would help or hinder specific thinking tasks (facilitation), analyze how emotions combine and change (understanding), and evaluate the effectiveness of strategies for managing emotions in scenarios (management). Responses are scored against expert or consensus norms, analogous to how IQ tests score reasoning responses against correct answers. The recently released MSCEIT 2 updated the theory, broadened the skills assessed, and renamed "facilitating thought" as "connecting emotions." Ability EI correlates modestly with IQ (approximately $r = .20-.30$) and shows incremental validity above IQ and personality for predicting some outcomes, including relationship quality and workplace behavior.

Trait EI (Petrides's model) treats emotional intelligence as a constellation of emotional self-perceptions located at the lower levels of personality. It is measured by self-report questionnaires, primarily the Trait Emotional Intelligence Questionnaire (TEIQue), which assesses 15 facets organized under four factors: wellbeing, self-control, emotionality, and sociability. Trait EI is unrelated to cognitive ability (it does not correlate with IQ tests or Raven's matrices) and instead maps onto established personality dimensions, correlating moderately with extraversion, agreeableness, conscientiousness, and (negatively) neuroticism. Petrides argues that because emotional experience is inherently subjective, self-report is the appropriate measurement method, and that the construct should be understood as a personality trait rather than a cognitive ability.

Mixed models (Goleman's model, Bar-On's EQ-i) combine elements of both ability and trait approaches, mixing perceived emotional skills with personality traits, motivational factors, and social competencies. Bar-On's Emotional Quotient Inventory (EQ-i 2.0) includes intrapersonal skills, interpersonal skills, adaptability, stress management, and general mood. These models are widely used in corporate training and leadership development, but psychometricians have criticized them for lacking discriminant validity from existing personality measures and for using self-report to assess what are claimed to be abilities.

The three-model confusion matters for this book because it reveals that the modern field has not

achieved consensus on what emotional intelligence *is*. Is it a cognitive ability to process emotional information (Salovey and Mayer)? A personality trait reflecting emotional self-perceptions (Petrides)? Or a grab-bag of personality characteristics, social skills, and motivational factors lumped together under an appealing label (Goleman, Bar-On)? The fact that research using one model's measures produces different results from research using another's means that "EI predicts X" is almost meaningless without specifying which model and which measure.

From the cross-civilizational perspective, this confusion is diagnostic. The traditions did not separate emotional competence into ability versus trait because they did not separate emotion from cognition in the first place. The *xin* is not an ability to process emotions or a personality disposition toward emotional self-awareness. It is an integrated faculty that thinks and feels as one. The three-model confusion in modern EI research is, in part, a consequence of trying to re-integrate a dimension that should never have been separated from cognition, using the tools of a tradition (psychometrics) that was built on the assumption that cognition and emotion are distinct domains.

Current measurement beyond EI scales

Beyond the major EI instruments, the Levels of Emotional Awareness Scale (LEAS) measures how precisely and complexly individuals describe their own and others' emotional experiences, using a developmental model that parallels cognitive developmental stages. Research on interoception (the perception of internal bodily signals such as heartbeat, breathing, and gut feelings) has opened a frontier in understanding the bodily basis of emotional processing. People who are more accurate at detecting their own heartbeat tend to experience emotions more intensely and show greater emotional awareness, suggesting that emotional intelligence has a somatic foundation that neither ability nor trait models adequately capture.

Frontier questions

Predictive processing models of affect propose that emotions are not reactions to events but predictions about the body's internal state and its relationship to the environment. If emotions are predictions, then emotional intelligence may involve the accuracy and flexibility of these predictions: the emotionally intelligent person is one whose affective predictions are well-calibrated, who updates them appropriately in response to new information, and who uses them effectively to guide action. This framework connects emotional processing to the broader predictive architecture of the brain, potentially unifying emotional and cognitive intelligence within a single computational framework. Body-based approaches to emotion, building on Antonio Damasio's somatic marker hypothesis and

more recent interoceptive inference models, suggest that emotional processing is fundamentally embodied: it arises from the brain's monitoring and regulation of the body's physiological state. If this is correct, then the traditions that treated knowing as inseparable from bodily feeling (the Chinese *xin*, Japanese embodied practice, the Yoruba *ogbón-inú* as wisdom "of the belly") were tracking a genuine feature of human cognition that the mind-body dualism of the Western tradition obscured.

Summary for the dimensional framework

Emotional attunement is a dimension of human excellence that was central to multiple traditions in the survey, that has been the subject of intense research interest since the 1990s, and that is currently measured by several instruments that operationalize fundamentally different constructs. The three-model confusion (ability, trait, mixed) reflects the difficulty of re-integrating a dimension that the psychometric tradition amputated.

The cross-civilizational evidence suggests that the traditions were right: emotional attunement is not a separate add-on to intelligence but an integral dimension of excellent functioning. The Chinese *xin*, the Buddhist wisdom-compassion pairing, and the Yoruba *ogbón-inú* all point to the same insight: cognition and emotion are not separate systems that happen to interact but aspects of a single integrated faculty. A framework for human excellence that treats emotional attunement as a separate dimension from cognitive ability (as this book must, for analytical purposes) should do so with the explicit acknowledgment that the separation is an analytical convenience, not a description of how the mind actually works.

Chapter 25: Social and Relational Intelligence

Chapter 24 addressed how individuals perceive and manage their own emotions and those of others. This chapter extends outward: the capacity to understand other people, their intentions, feelings, norms, and the dynamics of social interaction. This is a domain of competence that some traditions treat as the very essence of human excellence and others barely mention in their formal definitions of intelligence.

What the traditions tracked

The traditions surveyed in Movement I diverge sharply on how much weight they give to social and relational competence.

The Ubuntu/Hunhu/Botho tradition (Chapter 11) places relational understanding at the absolute center: *umuntu ngumuntu ngabantu* (a person is a person through other persons). In this framework, cognitive competence extracted from relational context is not merely incomplete; it is conceptually incoherent. The unit of analysis is not the individual mind but the relationship, and the measure of excellence is not individual performance but the quality of one's participation in the community. To be wise, in this tradition, is to understand others deeply, to maintain harmonious relationships, and to contribute to collective flourishing. Social intelligence is not a supplement to intelligence. It is what intelligence means.

Confucian role-attunement (Chapter 6) similarly embeds excellence in a relational matrix. The *junzi* (exemplary person) is defined by the quality of their conduct within the five fundamental relationships (ruler-minister, parent-child, husband-wife, elder-younger, friend-friend). To be excellent is to understand what each relationship requires and to fulfill those requirements with *li* (ritual propriety) and *ren* (humane-heartedness). Social perception, knowing what the situation demands, reading the other person's position and needs, responding with appropriate timing and tone, is a central intellectual virtue, not a soft skill peripheral to cognition.

The Inuit tradition (Chapter 12) valued a cluster of capacities centered on collaborative competence: the ability to work well with others in small groups under harsh environmental conditions, to anticipate others' needs, to manage conflict without disruption, and to coordinate complex tasks (hunting, travel, camp construction) through implicit understanding rather than explicit instruction. The Inuit concept of *isuma* (reason, sense, wisdom) includes social maturity and the ability to maintain equanimity in group settings.

Thorndike's concept of social intelligence (1920) is the earliest formal Western attempt to identify

this dimension. He defined it as "the ability to understand and manage men and women, boys and girls, to act wisely in human relations." This formulation directly parallels the Confucian emphasis on relational wisdom and the Ubuntu emphasis on understanding others. But Thorndike's concept never achieved the measurement precision that would have made it a permanent part of the psychometric mainstream. It was overshadowed by the focus on abstract cognitive ability, and the social dimension was largely sidelined for decades.

The Greek tradition, interestingly, gave social intelligence a less central role. Aristotle discussed *phronēsis* (practical wisdom) as essential for political life, and his account of friendship in the *Nicomachean Ethics* involves sophisticated social understanding. But the philosophical tradition that flowed from Greece through the Enlightenment tended to define intellectual excellence in terms of individual rational capacity rather than relational skill. This is one of the most significant divergences between the Greek-descended tradition and the traditions of sub-Saharan Africa, East Asia, and the circumpolar North.

Modern social cognition research

Modern psychology has studied social cognition extensively, identifying several component processes: theory of mind (the ability to attribute mental states to others), emotion recognition (the ability to identify emotional states from facial expressions, vocal tone, and body language), empathy (the capacity to share or understand others' emotional experiences), and social perception (the ability to read social situations, understand norms, and predict social outcomes).

These components are distinguishable but correlated. A person can be good at reading facial expressions but poor at understanding complex social dynamics. A person can understand others' mental states (cognitive empathy) without sharing their feelings (affective empathy). The decomposition of "social intelligence" into measurable components has been a significant achievement of modern psychology, but it has also revealed how far the field is from possessing a single, validated measure of the broad social intelligence that traditions like Ubuntu or Confucianism valued.

Current measurement: a landscape of limitations

Several instruments attempt to measure aspects of social cognition, but the measurement landscape is considerably less developed than for cognitive ability.

The **Reading the Mind in the Eyes Test (RMET)**, developed by Baron-Cohen and colleagues (2001), is the most widely used measure of social cognition in research. It presents 36 photographs of the eye region of faces and asks respondents to select which of four mental state descriptors best

describes what the person is thinking or feeling. The RMET has been cited over 2,000 times and used to study gender differences, cultural variation, genetic influences, and clinical conditions including autism spectrum disorder and schizophrenia.

However, the RMET has drawn increasing psychometric criticism. A 2023 study by Higgins, Ross, Langdon, and Polito found that the RMET shows poor psychometric properties in a large, demographically representative U.S. sample: nearly a quarter of the items did not meet the original validation criteria, and the factor structure showed inadequate model fit. There is also fundamental uncertainty about what the RMET measures. It was designed to assess theory of mind, but research has shown that it correlates more strongly with emotion recognition than with alternative theory of mind measures. A PMC study by Oakley and colleagues found that when the effects of alexithymia (difficulty identifying and describing emotions) were controlled for, the difference between autistic and non-autistic participants on the RMET disappeared, suggesting the test may measure emotion recognition rather than theory of mind. Furthermore, RMET performance is influenced by vocabulary knowledge, social class, and cultural familiarity with the mental state descriptors, raising questions about whether low scores reflect deficits in social cognition or merely unfamiliarity with the vocabulary of emotion terms. A survey of 52 social cognition experts from 20 countries found that ten experts urged discontinuation of the RMET, mostly due to validity concerns.

The **Social Shapes Test** and related tasks use animated geometric shapes to assess whether respondents spontaneously attribute mental states (intentions, goals, emotions) to moving objects. This approach bypasses some of the vocabulary and cultural bias problems of the RMET, but it remains primarily a research tool rather than a standardized clinical instrument.

360-degree feedback instruments, widely used in organizational settings, collect ratings of an individual's social competence from supervisors, peers, and subordinates. These provide ecologically valid information about how a person functions in actual social contexts, but they are susceptible to rater bias, halo effects, and organizational politics. They measure reputation for social competence rather than social competence itself.

The **Movie for the Assessment of Social Cognition (MASC)** presents film clips of social interactions and asks respondents to infer the mental states of the characters. It has better ecological validity than still photographs and has been found to discriminate between clinical and control groups more effectively than the RMET for some conditions. However, it is time-intensive and not yet widely standardized.

The overall picture is that social intelligence is acknowledged as important by both the cross-civilizational traditions and modern psychology, but the measurement tools available are significantly less reliable, less valid, and less standardized than those available for cognitive ability. This measurement

gap is not because social intelligence is less real or less important. It is because social cognition is inherently more context-dependent, culturally variable, and multidimensional than abstract reasoning, making it harder to capture with the standardized, decontextualized, single-score instruments that the psychometric tradition prefers.

Frontier questions

Second-person neuroscience, a research program pioneered by Leonhard Schilbach and colleagues, argues that social cognition has been studied primarily from a spectator's perspective: observing others passively, reading photographs, watching videos. But real social understanding is interactive: it emerges in the back-and-forth of live engagement with another person. Brain imaging studies of live social interaction reveal neural processes (including inter-brain synchrony, where the neural oscillations of interacting partners become temporally coupled) that are absent when people merely observe social stimuli. If social intelligence is fundamentally interactive rather than observational, then the entire measurement paradigm (presenting static stimuli and asking for judgments) may systematically miss the core of what social competence involves.

This finding has direct implications for the traditions' accounts. The Ubuntu understanding of personhood as constituted through relationship, the Confucian emphasis on role-based attunement that emerges only in interaction, and the Inuit model of collaborative competence that develops through shared activity all describe social intelligence as something that exists in the interaction, not in the individual. Second-person neuroscience provides empirical support for this philosophical insight: the brain processes involved in genuine social engagement are different from those involved in observing social situations from the outside.

Summary for the dimensional framework

Social and relational intelligence is a dimension that several traditions treated as the core of human excellence (Ubuntu, Confucianism, Inuit collaborative competence) and that modern psychology has investigated extensively under labels including social intelligence, theory of mind, social cognition, and empathy. The measurement tools available are weaker than those for cognitive ability, with the most widely used instrument (the RMET) drawing serious psychometric criticism. The emerging field of second-person neuroscience suggests that the measurement paradigm itself may be flawed: social intelligence may be fundamentally interactive, existing in the relationship rather than in the individual, exactly as the Ubuntu and Confucian traditions claimed.

A framework for human excellence that excludes social and relational competence excludes what

several of the world's major intellectual traditions placed at the center of their accounts of the excellent person. The measurement challenges are real, but they reflect the difficulty of the dimension, not its unimportance.

Chapter 26: Self-Regulation and Executive Control

The ability to direct, inhibit, and monitor one's own cognition, to stay focused when distracted, to resist impulses when restraint is required, to hold information in mind while working with it, and to switch flexibly between tasks or mental sets, is now recognized as a major determinant of real-world functioning. It overlaps with but is not identical to intelligence as conventionally measured, and it has ancient roots in traditions that prized self-governance as the foundation of all other excellences.

What the traditions tracked

Self-regulation appears in the cross-civilizational record under many names, but the underlying idea is remarkably consistent: the excellent person is one who governs their own mental life rather than being governed by impulse, habit, or circumstance.

The Stoic tradition provides the most explicit Western philosophical account. Stoic *apatheia* (not the absence of feeling, but mastery over the passions) and *prosoche* (attention, vigilant self-monitoring) describe a systematic program of cognitive self-governance. The Stoic sage does not suppress emotion; they examine each impression (*phantasia*) before assenting to it, testing whether it accurately represents reality before allowing it to guide action. This is, in modern terms, a theory of inhibitory control and metacognitive monitoring: the capacity to pause the automatic response, evaluate its appropriateness, and act only on the basis of considered judgment.

Confucian self-cultivation (*xiuyang*) (Chapter 6) is a lifelong program of directed attention, impulse management, and behavioral refinement. The *junzi* regulates speech, conduct, and emotional expression according to *li* (ritual propriety), which requires constant self-monitoring and adjustment. Confucius's famous statement that at seventy he could "follow his heart's desire without transgressing what is right" describes the developmental endpoint of executive control: a state in which appropriate self-regulation has become so deeply internalized that it operates without conscious effort. In the language of modern dual-process theory, this is the transition from effortful System 2 regulation to automated System 1 compliance with learned standards.

Buddhist *sati* (mindfulness) and *samādhi* (concentration) (Chapter 8) are perhaps the most systematically developed contemplative technologies for training attentional control in any tradition. Vipassanā meditation requires sustained attention to present-moment experience while inhibiting the habitual tendency to elaborate, judge, or narrate. Samatha meditation develops the capacity to maintain unwavering focus on a single object for extended periods. These are, in modern terms, training regimens for sustained attention, selective attention, and inhibitory control, the three core components

of executive function.

The scholastic tradition's treatment of *prudentia* (Chapter 5) treated self-governance as a precondition for wisdom: the prudent person is one who directs their own cognitive resources wisely, allocating attention, managing desire, and maintaining focus on the right goals. Aquinas's analysis of prudence as requiring *memoria* (memory of past experience), *intelligentia* (understanding of present circumstances), *docilitas* (openness to instruction), *sollertia* (quick-wittedness), *ratio* (reasoning), *providentia* (foresight), *circumspectio* (attentiveness to circumstance), and *cautio* (caution about risks) reads, remarkably, like a medieval theory of executive function.

What these traditions share is the recognition that self-regulation is not peripheral to excellence but foundational. Without the capacity to govern one's own attention, impulses, and cognitive processes, no other cognitive capacity, however powerful, can be deployed effectively. A brilliant reasoner who cannot sustain attention, a knowledgeable expert who cannot inhibit impulsive responses, a socially perceptive person who cannot regulate their own emotional reactions: all are diminished by failures of executive control.

Modern executive function research

Modern psychology has decomposed executive function into three core components, following the influential work of Miyake and colleagues (2000): **inhibition** (the ability to suppress prepotent responses), **working memory updating** (the ability to hold and manipulate information in mind), and **cognitive flexibility** (the ability to switch between tasks or mental sets). These three components are correlated but distinguishable, and they collectively predict a wide range of real-world outcomes including academic achievement, job performance, health behaviors, and relationship quality, often above and beyond what IQ predicts.

The relationship between executive function and intelligence is complex. Working memory, one of the three core components, correlates substantially with fluid intelligence (*Gf*): individual differences in working memory capacity account for a large proportion of the variance in *Gf*. Some researchers have argued that working memory *is* the core of fluid intelligence. But inhibition and cognitive flexibility show weaker relationships with IQ, suggesting that executive function is a broader construct that partially overlaps with but extends beyond what intelligence tests measure.

Current measurement: the divergence problem

Executive function is assessed through two fundamentally different approaches that produce disturbingly different results.

Performance-based tests present controlled tasks in laboratory or clinical settings. The **Stroop Test** measures inhibitory control by asking respondents to name the ink color of a word rather than reading the word itself (for example, saying "red" when the word "BLUE" is printed in red ink). The **Trail Making Test (TMT)** measures cognitive flexibility: Part A requires connecting numbered circles in sequence, while Part B requires alternating between numbers and letters (1-A-2-B-3-C...). **Working memory tasks** (digit span backward, n-back tasks) measure the capacity to hold and manipulate information in mind. These tests are standardized, reliable, and widely used in clinical neuropsychology.

Rating scales collect reports from the individual or from informants (parents, teachers, supervisors) about executive functioning in everyday life. The **Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2)** is the most widely used, covering inhibition, shift, emotional control, initiation, working memory, planning, and organization of materials across home and school settings.

The critical finding for this book's argument is that these two approaches often diverge. The Toplak et al. (2013) meta-analysis demonstrated that most performance-based executive function tests measure different constructs than rating scales or direct evaluations of executive function in daily life. Standard executive function tests (Trail Making, Stroop, Wisconsin Card Sorting) accounted for only 18-20% of the variance in everyday executive ability as measured by rating scales. Performance-based tests and rating scales show modest correlations, indicating partial overlap but also substantial divergence.

This divergence has led some researchers to conclude that executive function tests, despite their widespread clinical use, may not accurately assess real-world executive functioning. A person can perform well on the Stroop Test in a quiet clinic room (where the examiner provides the task structure, eliminates distractions, and supplies motivation) and still struggle with inhibitory control in the chaotic, self-directed, emotionally charged environment of daily life. Conversely, a person who performs poorly on timed tests may function perfectly well in their accustomed environment, where they have developed compensatory strategies and environmental supports.

This problem echoes the broader theme of this book. The psychometric tradition's instruments are designed for standardized, decontextualized, controlled conditions. But self-regulation, like practical judgment and social intelligence, is a capacity that is exercised in context, under real-world conditions, where the individual must supply their own task structure, manage their own motivation, and navigate competing demands. Strip away the context and you change what you are measuring.

Frontier questions

Computational models of cognitive control, particularly the expected value of control (EVC) framework developed by Shenhav and colleagues, propose that the brain allocates cognitive control based on a cost-benefit analysis: the amount of control deployed depends on the expected reward for effortful engagement minus the effort cost. This framework treats self-regulation not as a fixed capacity but as a resource allocation decision, and it predicts that individuals will show different levels of executive control in different contexts depending on their motivational state and the available rewards. If correct, this means that executive function is not a stable trait to be measured once and reported as a single score. It is a dynamic process that varies with context, motivation, and the perceived stakes of the situation.

Precision-weighting accounts, emerging from the predictive processing framework, propose that cognitive control involves adjusting the precision (reliability weighting) of different information sources: upweighting task-relevant signals and downweighting distractors. This computational account unifies attention, inhibition, and flexibility under a single mechanism (precision adjustment) and connects executive function to the broader predictive architecture of the brain.

Summary for the dimensional framework

Self-regulation and executive control is a dimension of human excellence that was central to Stoic self-command, Confucian self-cultivation, Buddhist contemplative training, and scholastic prudence. Modern research confirms that executive function predicts real-world outcomes above and beyond IQ. But the measurement of executive function reveals a familiar problem: performance-based tests show modest correspondence with real-world ratings, suggesting that the standardized, decontextualized measurement paradigm captures only part of what self-regulation involves.

The traditions recognized something that the measurement tools are still struggling to capture: self-regulation is not a fixed quantity to be assessed once in a controlled setting. It is a practiced capacity that develops over years of disciplined cultivation, that operates differently in different contexts, and that integrates cognitive, emotional, and motivational processes that the psychometric tradition tends to measure separately.

Chapter 27: Creativity and Generativity

The ability to produce novel, useful, or surprising ideas, artifacts, or solutions is valued in many traditions but rarely treated as the core of "intelligence." This makes creativity one of the most interesting dimensions in the framework: nearly everyone recognizes it when they see it, most traditions honored it, but few placed it at the center of their account of human excellence, and modern research has established it as a partly distinct construct that overlaps with but is not reducible to general cognitive ability.

What the traditions tracked

Creativity appears in the cross-civilizational record under a wider range of institutional and conceptual forms than almost any other dimension.

Greek *technē* (skilled making) and *poiēsis* (bringing into being) recognized the capacity to produce things that did not previously exist. Aristotle treated *technē* as one of the five intellectual virtues, distinct from *epistēmē* (scientific knowledge) and *phronēsis* (practical wisdom). The craftsman, the architect, the poet, and the physician all exercised *technē*: a rational capacity for making things well. But *technē* was subordinated to *epistēmē* and *phronēsis* in the Aristotelian hierarchy: knowing and judging ranked above making. The Greek tradition valued creative production but did not treat it as the highest form of intellectual excellence.

The Nahua *tlamatini* (Chapter 13) presents a striking contrast. The ideal wise person in Nahua thought was described as having "a face and a heart" (*in ixtli in yōllotl*), and the cultivation of this ideal explicitly included poetic and artistic training. The *cuicapicqui* (song-composer) and the *tlacuilo* (painter-scribe) exercised forms of creative excellence that were integral to the Nahua understanding of wisdom, not peripheral to it. The flower-and-song (*in xōchitl in cuīcatl*) tradition treated poetic creation as a mode of access to truth: the highest knowledge was reached not through deductive argument but through metaphorical and artistic expression.

Islamic artistic and architectural traditions developed creative excellence within a theological framework that both honored and constrained it. The arabesque, the muqarnas vault, geometric tessellation, and calligraphic elaboration represent extraordinary creative achievement within formal rules that channeled innovation into specific domains. The prohibition on figurative representation of living beings in many Islamic contexts did not suppress creativity; it redirected it toward abstract pattern, architectural engineering, and the decorative arts, producing some of the most technically sophisticated creative work in human history. Creativity here operated within constraints, and the

constraints themselves became generative: limitation bred invention.

The Romantic tradition (late eighteenth and nineteenth century Europe) represents the strongest case for creativity as the highest form of human excellence. The Romantic genius, from Beethoven to Byron to Goethe, was celebrated precisely for the capacity to originate: to produce works of art, literature, or philosophy that had never existed before and that transformed how people saw the world. This elevation of creative originality was new in Western intellectual history. For the Greeks, the highest intellectual achievement was contemplation of eternal truths. For the Romantics, it was the production of novel ones. The Romantic re-valuation of creativity is the historical backdrop against which modern creativity research developed.

The relationship to intelligence: correlated but not identical

The relationship between creativity and intelligence has been debated since Guilford's famous 1950 presidential address to the American Psychological Association, in which he argued that creativity was a neglected dimension of intellect and proposed divergent thinking as its cognitive foundation.

The threshold hypothesis, proposed by Torrance and others, suggests that a moderate level of intelligence is necessary for creative achievement but that above a certain threshold (often estimated at around IQ 120), further increases in IQ add little to creative potential. The evidence for a strict threshold is mixed: some studies find it, others do not. What is clear is that the correlation between IQ and creativity, while positive, is modest (typically $r = .20-.30$), leaving substantial room for individuals who are highly creative but not exceptionally high in IQ, and vice versa. Creativity and intelligence share some cognitive processes (working memory, associative fluency, processing speed) but diverge on others (openness to experience, tolerance of ambiguity, willingness to violate conventions, capacity for remote associations).

Modern research distinguishes between convergent thinking (finding the single correct answer to a well-defined problem, what IQ tests primarily measure) and divergent thinking (generating many possible answers to an open-ended problem, what creativity tests primarily measure). This distinction maps loosely onto the cross-civilizational evidence: most traditions valued convergent mastery of an established body of knowledge or skill (Confucian textual learning, Islamic hadith scholarship, Tibetan memorization) more highly than divergent originality, while the Romantic and post-Romantic West came to privilege originality, novelty, and disruption.

Current measurement

The **Torrance Tests of Creative Thinking (TTCT)**, developed by E. Paul Torrance in the 1960s and building on Guilford's work, remain the most widely used creativity assessment. The TTCT includes Verbal tests (thinking creatively with words) and Figural tests (thinking creatively with pictures). The Figural tests ask respondents to complete incomplete drawings, make pictures from shapes, or generate as many different images as possible from a repeated stimulus. Responses are scored on fluency (number of responses), originality (statistical rarity of responses), elaboration (detail), resistance to premature closure, and abstractness of titles.

The TTCT has genuine strengths. A meta-analysis of its reliability found acceptable composite reliability ($\omega = .81$). Longitudinal studies initiated in 1958 showed that TTCT scores hold moderate correlations with later creative achievements in personal domains over 50 years later, a remarkable finding for any psychometric instrument. However, the TTCT has also drawn significant criticism. Some researchers argue that fluency (simply generating many responses) overwhelms the scoring, so that the test may measure verbal or figural productivity rather than genuine creativity. The correlation with actual creative performance in specific domains is contested: one study found that TTCT scores and students' actual creative performance in three separate domains were completely unrelated. The validity debate remains active, with defenders arguing that divergent thinking is one genuine component of creative capacity and critics arguing that domain-general creativity tests cannot capture what is fundamentally a domain-specific phenomenon.

PISA 2022 introduced, for the first time, a creative thinking assessment at the international level. The OECD tested fifteen-year-olds' capacity to generate and evaluate novel ideas across domains including written expression, visual expression, scientific problem-solving, and social problem-solving. This represents a significant institutional acknowledgment that creativity can and should be measured alongside reading, mathematics, and science, although results are still being analyzed and the psychometric properties of the new measures are under active investigation.

Creative Achievement Inventories (such as Carson, Peterson, and Higgins's Creative Achievement Questionnaire) take a different approach: rather than testing creative *potential* through laboratory tasks, they measure actual creative *output* by asking respondents to report their achievements across domains (visual arts, music, creative writing, dance, theater, architecture, humor, inventions, science, culinary arts). These instruments have the advantage of measuring what people have actually done rather than what they can do on a timed test, but they are susceptible to self-report bias and confound creativity with opportunity (a person cannot report achievements in domains they have never had access to).

Frontier questions

AI-generated art, music, text, and code have forced a fundamental question: is creativity a uniquely human capacity, or is it a pattern of behavior that any sufficiently complex system can produce? Current large language models and image generators produce outputs that many observers cannot distinguish from human creative work. If creativity is defined behaviorally (the production of novel, useful, or surprising outputs), then AI systems may already be creative. If creativity requires subjective experience, intentionality, or the kind of meaning-making that the Nahua *in xōchitl in cuīcatl* tradition describes, then the question remains open. The AI creativity debate is not peripheral to this book's argument: it forces the question of whether creativity is a cognitive capacity (measurable by its outputs) or a form of human engagement with the world (measurable only by its meaning to the creator and the community).

Summary for the dimensional framework

Creativity and generativity is a dimension of human excellence that was recognized by every tradition in the survey, that occupies a unique position in the framework (valued by all, central to few), and that is currently measured by instruments of modest but genuine validity. The relationship to intelligence is one of partial overlap: creativity requires some cognitive capacity but adds something that IQ does not capture, including openness, flexibility, tolerance of ambiguity, and the willingness to produce something that did not exist before.

The cross-civilizational evidence suggests that most traditions valued creative excellence but embedded it within broader frameworks: the Greek hierarchy of intellectual virtues, the Islamic channeling of creativity through formal constraints, the Nahua integration of artistic expression with wisdom. Only the Romantic tradition elevated creative originality to the summit of human achievement. A framework for human excellence that includes creativity as one dimension among twelve is more faithful to the cross-civilizational record than either the Romantic celebration of genius or the psychometric tradition's near-total neglect of creativity in standard intelligence testing.

Chapter 28: Wisdom and Uncertainty Management

This chapter arrives at the dimension that most older traditions treated as the highest form of human excellence: wisdom. Good judgment under uncertainty, long-term perspective, the ability to balance competing interests, knowing when not to act, integrating cognitive, emotional, moral, and experiential capacities into a unified orientation toward life. Modern psychology has studied wisdom for three decades and has produced genuine measurement instruments. But the honest state of the field is that wisdom remains the most difficult dimension to operationalize, the one where different measures show the weakest convergence, and the one that most directly challenges the psychometric tradition's preference for single-score, decontextualized assessment.

What the traditions tracked

Wisdom appears in more traditions' accounts of the highest human excellence than any other dimension in this framework. Its centrality is nearly universal; what varies is how traditions conceptualized it.

Aristotelian *sophia* (theoretical wisdom: contemplative understanding of eternal truths) and *phronēsis* (practical wisdom: good judgment in particular situations) represent the Greek tradition's two-part account (Chapter 4). For Aristotle, *sophia* was the higher achievement, but *phronēsis* was necessary for living well. The practically wise person perceives the morally relevant features of situations, deliberates about what is best, and acts accordingly. *Phronēsis* requires experience, cannot be taught by rules alone (a theme revisited in Chapter 23), and develops over a lifetime.

Buddhist *prajñā* (Chapter 8) is wisdom understood as penetrating insight into the nature of reality, inseparable from compassion (*karuṇā*). The three-form developmental model (*śrutamayī*, *cintāmayī*, *bhāvanāmayī*) describes wisdom as progressing from intellectual learning through reflective analysis to direct meditative realization. The highest wisdom transcends conceptual thought altogether, reaching an understanding that cannot be captured in propositions. This is a dimension of excellence that resists measurement by any instrument requiring verbal articulation.

Confucian sagacity (*shèng*, 聖) represents the highest achievable human state in the Confucian tradition: the sage is one who has perfected self-cultivation to the point where moral perception, emotional attunement, and practical judgment operate in effortless harmony (Chapter 6). Confucius himself disclaimed sage status, implying that it is an aspiration that orients the project of self-cultivation without being fully attainable.

Akan and Yoruba traditions treat wise eldership as the culmination of a lifetime of accumulated ex-

perience, social observation, and moral development (Chapter 10). The wise elder is not simply old; they are someone whose decades of living have been transmuted, through reflection and communal engagement, into the capacity to counsel, adjudicate, and guide. The deployment of proverbs in council is a concrete expression of this wisdom: selecting the right condensed principle for the specific situation at hand.

Biblical and Solomonic wisdom represents a distinct tradition in which wisdom is partly a divine gift and partly a human achievement. The Wisdom literature (Proverbs, Ecclesiastes, Job, Wisdom of Solomon) presents wisdom as involving practical prudence, moral discernment, acceptance of human limitation, and reverence for the transcendent. The Book of Job, in particular, presents wisdom as the capacity to maintain integrity in the face of suffering that defies rational explanation, a form of excellence that combines cognitive humility with moral steadfastness.

What these traditions share is a conception of wisdom as integrative: it combines cognitive capacity (understanding complex situations), emotional attunement (responding with appropriate feeling), moral orientation (acting for the good), self-regulation (maintaining perspective under pressure), and experiential depth (drawing on decades of accumulated understanding). This is precisely what makes wisdom so difficult to measure. It is not a single capacity that can be isolated and tested. It is the integration of multiple capacities in the service of good judgment about how to live.

Modern wisdom science: the Berlin paradigm and its descendants

Modern psychological research on wisdom was pioneered by Paul Baltes and his colleagues at the Max Planck Institute in Berlin. The Berlin Wisdom Paradigm, introduced in the 1990s, defined wisdom as "expertise in the fundamental pragmatics of life" and assessed it through open-ended performance measures. Participants are presented with difficult life dilemmas (such as a fifteen-year-old girl wanting to get married, or a person receiving a call from a friend announcing they intend to commit suicide) and asked to think aloud about what they would consider and recommend. Trained coders rate responses on five criteria: rich factual knowledge about life, rich procedural knowledge (strategies for dealing with life's challenges), lifespan contextualism, value relativism (understanding that people hold different values), and recognition and management of uncertainty.

The Berlin paradigm was the first attempt to measure wisdom with the rigor of a psychological assessment, and it produced important findings. Wisdom-related performance shows a quadratic relationship with fluid intelligence (a minimal level is necessary, but above that level, more IQ does not mean more wisdom). Wisdom scores show zero correlation with age in most studies, a finding that challenges both the folk assumption that age brings wisdom and the psychometric assumption that cognitive decline with age implies declining excellence. The absence of age effects suggests that

wisdom depends on what one has done with one's experience, not on how long one has lived.

However, the Berlin paradigm is labor-intensive (requiring trained coders to rate extended verbal responses), expensive to administer, and impractical for large-scale research. This led to the development of several alternative measures.

Current measurement: a landscape of partial convergence

Monika Ardelt's **Three-Dimensional Wisdom Scale (3D-WS)** is one of the most widely used self-report measures. It assesses three dimensions: cognitive (deep understanding of life, acceptance of ambiguity), reflective (ability to examine events from multiple perspectives, overcome self-centeredness), and affective (compassion and concern for others). Ardelt's inclusion of the affective dimension distinguishes her approach from the Berlin paradigm and from Grossmann's approach, both of which focus primarily on cognitive processes. Ardelt argues that wisdom without compassion is not wisdom, a position that aligns with the Buddhist *prajñā-karunā* inseparability and the Yoruba integration of emotional and intellectual wisdom.

Igor Grossmann's **Situated Wise Reasoning Scale (SWIS)** takes a different approach. Rather than asking people to rate themselves globally on wisdom-related traits, the SWIS asks respondents to recall specific situations and rate the extent to which they demonstrated intellectual humility, recognition of uncertainty and change, consideration of multiple viewpoints, use of an outsider's vantage point, and search for compromise. This situated approach is motivated by Grossmann's finding that wisdom varies considerably within individuals across situations: a person may reason wisely about a relationship conflict but unwisely about a political disagreement. A 2017 study showed that over a nine-day period, there was considerable within-person variability in wise reasoning, suggesting that wisdom is not a fixed trait but a context-dependent process. Brienza et al. (2017) found that two to five episodes need to be sampled for sufficient reliability.

The **SD-WISE** (San Diego Wisdom Scale) is a brief self-report measure developed by Dilip Jeste and colleagues, designed for use in clinical and community settings. It assesses pro-social behaviors, emotional regulation, self-reflection, acceptance of uncertainty, decisiveness, and social advising.

The honest state of the field, as Glück and colleagues have documented in a comprehensive comparative study, is that these different measures show only modest convergence. The Berlin Wisdom Paradigm, the 3D-WS, the Self-Assessed Wisdom Scale, and the Adult Self-Transcendence Inventory correlate with each other, but the correlations are far from strong enough to conclude that they are measuring the same thing. Different measures also show different relationships with age: zero correlations for the Berlin paradigm and Grossmann's measure, negative correlations for the Bremen

paradigm and 3D-WS, and an inverse U-shaped relationship for the Self-Assessed Wisdom Scale. A construct whose measures disagree on whether it increases, decreases, or remains stable with age is a construct that has not yet been adequately defined.

This is not a failure of the researchers. It is a consequence of the nature of the dimension itself. Wisdom, as the traditions understood it, integrates cognitive, emotional, moral, and experiential dimensions that the psychometric tradition has been designed to measure separately. Trying to capture wisdom with a single instrument is like trying to capture a symphony with a single microphone: you can record something, but you will inevitably lose the relationships between parts that constitute the musical meaning. The different measures capture different aspects of wisdom (cognitive, affective, reflective, situated), and their modest convergence reflects the fact that wisdom is not one thing but the integration of many things.

Wisdom as the test case for the plural framework

This chapter's argument is that wisdom may be the dimension where the plural-axes framework proposed in this book matters most. If wisdom is the integration of cognition, emotion, morality, self-regulation, and perspective-taking, then a single-axis model of intelligence (however well it measures reasoning or processing speed) cannot capture wisdom, because wisdom is constituted by the relationships among dimensions rather than by high performance on any one.

The traditions recognized this. The Buddhist insisted that wisdom without compassion is not wisdom. The Confucian sage integrated moral perception with practical judgment with emotional attunement. The Akan elder combined deep experience with social sensitivity with the capacity to deploy the right proverb at the right moment. The Solomonic tradition combined practical prudence with moral discernment with acceptance of human limitation. In every case, what made someone wise was not exceptional performance on any single axis but the integration of multiple capacities into a coherent, well-calibrated orientation toward the challenges of living.

The psychometric tradition's difficulty with wisdom is not a technical limitation waiting to be solved by a better test. It is a structural consequence of the tradition's foundational assumption that excellence can be decomposed into measurable components and that the components, measured separately, capture what matters. Wisdom is the dimension that most directly refutes this assumption, because in the case of wisdom, what matters *is* the integration, and the integration is what decomposition destroys.

Summary for the dimensional framework

Wisdom is the dimension that the traditions valued most and that modern psychology measures least well. The Berlin Wisdom Paradigm, the 3D-WS, and the Situated Wise Reasoning Scale are genuine and serious measurement instruments, but they show only modest convergence, disagree on the relationship between wisdom and age, and have not achieved the level of standardization that would allow wisdom to be assessed with the same confidence as IQ.

This is the right chapter to state the book's core claim about measurement in its strongest form: some dimensions of human excellence, and wisdom above all, are constituted by the integration of multiple capacities rather than by high performance on any single capacity. A measurement framework adequate to these dimensions must find ways to assess integration, not just decomposition. The final chapters of this book will propose how such a framework might work.

Chapter 29: Moral Discernment and Character

This chapter addresses the dimension of excellence that the psychometric tradition most deliberately excluded: moral character. Not just knowing what is right, but being reliably disposed to do it. Many of the traditions surveyed in Movement I made moral reliability central to their definition of excellence. Modern psychometrics stripped it out. But it keeps returning: in SJTs, in personality inventories, in character strengths research, and in virtue-based frameworks that recapitulate, often without acknowledgment, what the traditions always claimed.

What the traditions tracked

The convergence on this dimension is one of the most striking findings of the entire book. Chapter 19 identified five independently developed ideals of the excellent person that integrate cognitive and moral excellence: the Aristotelian *phronimos*, the Confucian *junzi*, the Islamic *adīb*, the Yoruba *omolúàbí*, and the Nahua *tlamatini*. In every case, intellectual excellence without moral character was not merely incomplete. It was disqualifying.

Aristotelian virtue ethics (Chapter 4) provides the most philosophically developed Western account. For Aristotle, the virtues are not rules to follow but stable dispositions of character acquired through practice: one becomes courageous by doing courageous things, temperate by practicing temperance. The virtues are unified by *phronēsis* (practical wisdom), which discerns what virtue requires in each particular situation. A person of high cognitive ability who lacks virtue is not simply excellent-minus-one-dimension; they are dangerous, because their intelligence serves ends that are not good. Confucian character cultivation (Chapter 6) treats moral development as the central project of education. The *junzi* is defined by *ren* (humane-heartedness), *yi* (righteousness), *li* (ritual propriety), and *zhi* (wisdom), and the cultivation of these virtues is the purpose of the entire educational system, from childhood instruction through the Imperial Examination. To separate cognitive assessment from moral assessment, as the psychometric tradition does, would be incomprehensible within this framework: the point of developing the mind is to become a good person.

The Yoruba *omolúàbí* (Chapter 10) is a person of good character: honest, respectful, diligent, generous, and wise. The Ifá corpus teaches that character (*ìwà*) is the paramount quality: "Gentle character is what enables the rope of life to stay unbroken." A person who is clever but lacks *ìwà* is not an *omolúàbí* and will not be trusted with responsibility regardless of their cognitive abilities.

Buddhist ethical training (Chapter 8) makes *śīla* (ethical conduct) the foundation of the path: before meditation can develop concentration (*samādhi*), and before concentration can produce insight

(*prajñā*), the practitioner must establish a foundation of ethical discipline. The five precepts (abstaining from killing, stealing, sexual misconduct, false speech, and intoxicants) are not separate from intellectual development. They are its prerequisite.

Christian *prudentia* (Chapter 5), as developed by Aquinas, integrates moral and intellectual virtue: the prudent person is one who perceives correctly, deliberates wisely, and acts rightly. Prudence governs all other virtues because without it, courage becomes recklessness, temperance becomes rigidity, and justice becomes fanaticism.

The psychometric exclusion and its consequences

The psychometric tradition's exclusion of moral character from intelligence was deliberate, not accidental. Gordon Allport's influential claim in the 1930s that "character is merely personality evaluated, and personality is character devalued" effectively severed the connection between moral excellence and psychological measurement. The tradition that followed measured cognitive ability and personality traits while treating moral character as either a subset of personality (conscientiousness, agreeableness) or as a domain-specific set of behaviors not relevant to intelligence.

This exclusion created a vacuum that multiple research programs have attempted to fill.

Current measurement: personality, character, and the return of virtue

The **Big Five personality model** captures some aspects of moral character through two of its five factors. **Conscientiousness** (being organized, dependable, self-disciplined, and achievement-oriented) predicts job performance, educational attainment, health behaviors, and longevity. **Agreeableness** (being cooperative, trusting, empathic, and compliant) predicts relationship quality, prosocial behavior, and reduced antisocial conduct. These two factors, taken together, capture much of what the traditions meant by reliability and social virtue. But the Big Five was not designed as a moral framework; it emerged from factor analyses of personality descriptors in the English language, and its moral content is incidental rather than intentional.

Peterson and Seligman's **Values in Action (VIA) Classification of Character Strengths** (2004) represents the most ambitious modern attempt to create a scientific taxonomy of moral excellence. They reviewed the world's major philosophical, religious, and cultural traditions (including Confucianism, Taoism, Buddhism, Hinduism, Athenian philosophy, Christianity, Islam, and Judaism) to identify virtues valued across cultures. They identified six core virtues (wisdom, courage, humanity, justice, temperance, and transcendence) and 24 character strengths as the psychological processes that define them. The VIA Inventory of Strengths has been administered to over 13 million people

worldwide.

The VIA represents a genuine attempt to do what this book argues the psychometric tradition should have done all along: take moral excellence seriously as a measurable dimension of human functioning. However, the VIA has drawn criticism on several fronts. Factor analyses consistently fail to reproduce the six-virtue structure that Peterson and Seligman proposed. Instead, the 24 strengths tend to cluster into four or five factors that resemble the Big Five personality dimensions rather than the six virtues derived from philosophical and religious traditions. The mean percentage of variance in character strengths explained by the Big Five domains is 33%, leaving substantial but not overwhelming distinctiveness. The VIA is also a self-report instrument, vulnerable to social desirability bias, and its items are all positively keyed, which may inflate apparent agreement.

SJTs (Chapter 23) implicitly assess moral character in many workplace contexts. When an SJT presents a scenario involving an ethical dilemma (should you report a colleague's misconduct? Should you prioritize a client's interests over your organization's?) and asks what you would do or what the best response is, it is assessing moral judgment and, to some degree, moral disposition. The incremental validity of SJTs above cognitive ability and personality suggests that they capture something about moral-practical judgment that neither IQ nor Big Five scores fully account for.

The central tension: measurement versus formation

The deepest tension in this chapter is between the psychometric tradition's goal (measuring traits) and the virtue traditions' goal (forming character).

The psychometric tradition wants to know: How much of this trait does this person currently possess? The answer should be a number, comparable across individuals, stable over short periods, and predictive of future behavior. This is the logic of measurement: capture the current state and use it to predict.

The virtue traditions want something different: to develop a person's character over time through practice, habituation, mentorship, and communal formation. The question is not "how much virtue do you have?" but "are you becoming more virtuous?" The Aristotelian tradition emphasized that virtue is acquired through practice, not born. The Confucian tradition treated education as a decades-long project of moral formation. The Buddhist path is a progressive training in ethical conduct, concentration, and wisdom. The Yoruba *omolúàbí* ideal is something one grows into through a lifetime of communal engagement.

This tension is not merely philosophical. It has practical consequences for assessment. If moral character is a stable trait, then it can be measured once and used to predict future behavior (as the

Big Five does with conscientiousness). If moral character is a developmental achievement, shaped by practice, context, and formation over time, then a single-point assessment misses what matters most: the trajectory, the effort, the direction of growth. The traditions treated character as a developmental achievement. The psychometric tradition treats it as a trait. These are fundamentally different frameworks, and the choice between them determines what measurement can and cannot reveal.

Summary for the dimensional framework

Moral discernment and character is the dimension where the traditions' convergence is strongest and the psychometric tradition's exclusion is most conspicuous. Five independently developed ideals of the excellent person (Greek, Confucian, Islamic, Yoruba, Nahua) all integrate moral and cognitive excellence. The VIA Classification represents a serious modern effort to bridge this gap, though its factor structure does not replicate its theoretical virtue categories. The Big Five captures some moral content incidentally through conscientiousness and agreeableness, and SJTs implicitly assess moral judgment in applied settings.

The central tension between measurement (assessing a current trait) and formation (developing character over time) is not resolvable within the psychometric framework alone. A plural framework for human excellence must acknowledge that moral character is both measurable (as a current disposition) and developmental (as a lifelong project of formation), and that the traditions' emphasis on the latter is not a pre-scientific confusion but a genuine insight about the nature of the dimension itself.

Chapter 30: Embodied and Ecological Skill

Several of the traditions surveyed in this book define excellence partly through skilled attunement to environments, bodies, places, and ecological systems. This is not a metaphor for "real" intelligence located somewhere inside the head. It is a form of intelligence that the psychometric tradition largely ignores and that the emerging 4E cognition framework in cognitive science has begun to vindicate.

What the traditions tracked

The traditions that valued embodied and ecological skill most highly were those that survived in direct, sustained engagement with demanding physical environments.

Pacific wayfinding (Chapter 12) is the most dramatic case. Polynesian navigators crossed thousands of miles of open ocean without instruments, integrating knowledge of star positions, ocean swell patterns, wave refraction, cloud formations, bird flight, and marine phosphorescence into a continuous, real-time navigational judgment. This was not abstract reasoning applied to environmental data. It was embodied perception: the navigator lay in the hull of the canoe, feeling the wave patterns through their body, detecting subtle shifts in swell direction that indicated proximity to land. The Hawaiian star compass divided the horizon into thirty-two houses, each associated with the rising or setting point of a celestial body, and this entire framework was held in memory and applied through perceptual attunement rather than calculation. The skill was validated through the most demanding possible test: arriving alive at the intended destination after weeks at sea.

Inuit land, sea, and weather knowledge (Chapter 12) represents a different form of ecological intelligence: the capacity to read environmental signs (snow texture, ice color, wind direction, animal behavior, cloud shape) as an integrated perceptual field that guides survival decisions. An experienced Inuit hunter does not consult a mental checklist of environmental indicators; they perceive the landscape as a meaningful whole, in which danger, opportunity, and navigation are directly disclosed through skilled perception. This knowledge develops over decades of practice and mentorship and is validated through demonstrated competence, not through any form of standardized testing.

Māori ecological and genealogical memory (Chapter 12) integrates knowledge of landscape, waterways, species behavior, seasonal patterns, and ancestral history into a unified framework. *Whakapapa* (genealogy) does not merely record human descent; it maps relationships between people, land, rivers, mountains, and the natural world, treating ecological knowledge and historical knowledge as aspects of a single relational system.

Japanese embodied praxis (Chapter 13) takes a different approach. The *dō* disciplines (budō, chadō,

kadō, shodō) treat the development of physical skill as a cognitive and spiritual achievement. In martial arts, tea ceremony, flower arrangement, and calligraphy, the body is not merely the instrument of a mind that directs it. The body is the site of understanding. The *shu-ha-ri* progression (protect the form, break the form, transcend the form) describes a developmental trajectory in which knowledge that begins as explicit instruction becomes embodied, then becomes spontaneous, and finally transcends the distinction between knowing and doing. At the highest levels, the practitioner acts with a fluency that cannot be articulated in propositions, a phenomenon that Dreyfus and Dreyfus later identified as the hallmark of genuine expertise.

Modern cognitive science: the 4E framework

The 4E cognition framework, associated with researchers including Shaun Gallagher, Andy Clark, Alva Noë, and Francisco Varela, argues that cognition is not confined to the brain but is embodied (shaped by the body's structure and sensorimotor capacities), embedded (situated in and dependent on environmental context), enacted (constituted through active engagement with the world), and extended (distributed across brain, body, and environmental resources like tools, technologies, and cultural artifacts).

This framework directly challenges the assumption that underlies the psychometric tradition: that cognition is an internal computational process that can be measured by extracting a person from their environment, placing them in a controlled setting, and presenting abstract problems. If cognition is embodied, embedded, enacted, and extended, then the controlled testing environment does not merely simplify the measurement context. It removes the very conditions under which the capacities in question operate. Testing a Pacific navigator's intelligence with Raven's Progressive Matrices is not like testing a fish's swimming ability in a tank. It is like testing a fish's swimming ability on land.

The ecological psychology tradition, initiated by J.J. Gibson, provides the perceptual theory that supports this view. Gibson argued that perception is not the internal processing of sensory data but the direct detection of affordances: possibilities for action that the environment offers to a perceiving organism. A skilled perceiver does not see raw visual data and then infer meaning; they see opportunities, obstacles, paths, and dangers directly. The Pacific navigator perceives the ocean as navigable space; the Inuit hunter perceives the ice as a field of risks and resources. These are not inferences drawn from perceptual input. They are perceptions, shaped by decades of embodied practice.

Current measurement: almost nonexistent in standardized form

There is no standardized psychometric instrument for embodied or ecological intelligence. None. The dimension that several traditions treated as the most important cognitive achievement for survival, community welfare, and cultural transmission has no IQ test, no standardized scale, no norm-referenced assessment.

This is not because the dimension is unmeasurable. It is measured constantly within the traditions themselves, through apprenticeship, demonstrated performance, and peer or elder evaluation. The Pacific navigator's competence is validated by successful voyages. The Inuit hunter's skill is assessed by elders who have watched them in the field over years. The Japanese martial artist's mastery is certified through demonstration before qualified judges. The Māori *tohunga* earns recognition through decades of supervised practice and accurate performance.

These are rigorous assessments. They are reliable (experts agree on who is competent), valid (they predict future performance), and consequential (failure has real-world costs, up to and including death). They satisfy the basic requirements of good measurement. What they are not is standardized, decontextualized, or reducible to a single score. They require expert observation of performance in context over extended periods, exactly the kind of assessment that the psychometric tradition was designed to replace with something more efficient.

The ecological validity movement in neuropsychology (discussed in Chapters 25 and 26) represents a partial recognition within the psychometric tradition that testing under controlled conditions may not capture real-world functioning. But this movement has focused primarily on bringing testing closer to real-world conditions (through virtual reality, naturalistic tasks, and ecologically valid test batteries) rather than on recognizing that some dimensions of excellence can only be assessed in the actual environments where they are exercised.

Frontier questions

Naturalistic neuroscience, studying brain function during real-world activities rather than laboratory tasks, is beginning to document the neural processes involved in skilled environmental engagement. Mobile EEG and fNIRS allow brain imaging during navigation, craftsmanship, and skilled movement, revealing neural dynamics that are absent in seated laboratory tasks. These studies suggest that skilled environmental engagement involves brain-body-environment coupling that is qualitatively different from the neural processes engaged by abstract cognitive tasks.

Interoception research (the study of how the brain monitors internal bodily states) provides a bridge between embodied cognition and the emotional attunement discussed in Chapter 24. Skilled prac-

tioners in many traditions report that their expertise involves "feeling" the right action rather than calculating it: the martial artist who perceives the opponent's intention through a bodily felt sense, the navigator who detects the change in swell pattern through the motion of the hull against the body. If interoceptive sensitivity is a component of skilled environmental engagement, then embodied intelligence has a measurable physiological substrate, even if standardized tests have not yet been developed to capture it.

Summary for the dimensional framework

Embodied and ecological skill is the dimension where the gap between the traditions and the psychometric tradition is widest not in terms of valuation (as with practical judgment or moral character) but in terms of recognition. The psychometric tradition does not merely undervalue this dimension; it has no instruments for it at all. The traditions that valued it most (Pacific, Inuit, Māori, Japanese embodied disciplines) assessed it rigorously through apprenticeship and demonstrated performance, and these assessments were reliable, valid, and consequential.

The 4E cognition framework provides the theoretical basis for taking this dimension seriously: if cognition is embodied, embedded, enacted, and extended, then a framework for human excellence that measures only disembodied, decontextualized, internal cognitive processes is not merely incomplete. It is measuring a fundamentally different thing from what the traditions tracked. A plural framework must find ways to recognize embodied and ecological skill as a genuine dimension of excellence, even if standardized measurement is not available and may never be, because the dimension's nature resists the decontextualization that standardized measurement requires.

Chapter 31: Collective and Distributed Intelligence

The previous eleven chapters have examined dimensions of excellence as properties of individuals. This chapter challenges the assumption that intelligence is an individual property at all. Some traditions and some modern research programs treat intelligence as something that emerges from groups, relationships, and distributed systems, not merely from individual minds working side by side.

What the traditions tracked

Several traditions in the survey resisted the extraction of the individual from the collective that the psychometric tradition treats as methodologically necessary.

The Inuit concept of *Piliriqatigiingniq* (working together for a common purpose) (Chapter 12) treats collaborative competence as a core social value and cognitive achievement. In the demanding conditions of Arctic life, no individual, however skilled, could survive alone. Hunting, travel, shelter construction, and child-rearing all required seamless coordination among group members, often under conditions where explicit verbal communication was impossible (during a hunt in silence, during travel in high winds). The "intelligence" that kept the group alive was not the aggregate of individual cognitive abilities. It was the capacity of the group to function as a coordinated unit, anticipating each other's actions, distributing tasks according to shifting circumstances, and maintaining cohesion under extreme stress.

Ubuntu communal personhood (Chapter 11) makes the most radical philosophical claim: that cognition and personhood are relationally constituted. *Umuntu ngumuntu ngabantu* (a person is a person through other persons) is not merely a social norm. It is an ontological claim about where intelligence resides. If personhood is constituted through relationship, then the unit of cognitive analysis is not the individual mind but the relational network. The "intelligence" of a community is not the sum of its members' IQ scores. It is a property of the relational system itself, emerging from the quality of connections, the distribution of knowledge, and the patterns of mutual responsiveness.

Confucian relational ethics (Chapter 6) similarly embeds cognitive excellence in a relational matrix. The *junzi* is excellent not in isolation but in relationship: their wisdom manifests in how they conduct the five fundamental relationships. The quality of the relationship, not the cognitive capacity of either individual alone, is what Confucian ethics tracks and evaluates.

The Western philosophical tradition has been less attentive to collective intelligence, but there are important exceptions. Aristotle's account of political wisdom in the *Politics* argues that the collective judgment of many citizens may exceed the judgment of any single expert: the many, even if individ-

ually mediocre, may collectively possess wisdom that no individual has alone. This is arguably the first statement of the "wisdom of crowds" hypothesis.

Modern collective intelligence research

The most important modern empirical contribution is the work of Anita Williams Woolley, Thomas Malone, and colleagues at MIT's Center for Collective Intelligence. In a landmark 2010 study published in *Science*, they examined whether a "collective intelligence" factor (*c*) exists for groups, analogous to the general intelligence factor (*g*) for individuals.

In two studies with 699 people working in groups of two to five, they found converging evidence for a general collective intelligence factor that explains a group's performance on a wide variety of tasks. Groups were given diverse challenges including visual puzzle-solving, collective brainstorming, moral decision-making, and resource bargaining. A single factor predicted over 40% of the variance in performance across all tasks.

The most striking finding was what predicted the *c* factor. It was **not** strongly correlated with the average or maximum individual intelligence of group members. Instead, it was correlated with three factors: the average social sensitivity of group members (measured by the Reading the Mind in the Eyes Test), the equality of conversational turn-taking (groups where a few people dominated performed worse), and the proportion of women in the group (mediated by their higher average social sensitivity). Group cohesion, satisfaction, and motivation were not significant predictors.

A subsequent PNAS study (2021) analyzing data from 22 studies with 5,279 individuals in 1,356 groups confirmed the robustness of the *c* factor and found that group collaboration process was more important in predicting collective intelligence than the skill of individual members.

However, the findings have been contested. A replication study by Bates and Gupta (2017) with 312 participants found that individual IQ accounted for approximately 80% of group-IQ differences, and that the effects of social sensitivity, gender proportion, and turn-taking did not replicate. They concluded that collective intelligence may predominantly reflect individual cognitive ability rather than a genuinely emergent group property.

This debate is important for this book because it tests whether intelligence can be a genuinely collective phenomenon or whether it reduces to the aggregate of individual capacities. The evidence is genuinely mixed: some studies support emergence (group performance exceeding what individual abilities predict), while others find that individual IQ dominates. The resolution may depend on the type of task, the size and composition of the group, and the nature of the collaboration required.

Distributed cognition

The theoretical framework of distributed cognition, developed by Edwin Hutchins and others, provides a complementary perspective. Hutchins's landmark study of naval navigation showed that the cognitive process of determining a ship's position is not located in any single person's head. It is distributed across multiple people (bearing takers, plotters, supervisors), instruments (compasses, charts, parallel rulers), and procedures (communication protocols, calculation routines). No individual in the navigation team performs the complete cognitive task. The "intelligence" of the navigation system is a property of the system, not of any individual component.

This framework connects directly to the Pacific wayfinding tradition (Chapter 12 and Chapter 30), where navigational intelligence is distributed across the navigator, the canoe, the ocean, and the star compass. It connects to the Inuit hunting group, where the "intelligence" of the hunt is distributed across the hunters, the dogs, the landscape, and the prey. And it connects to Ubuntu philosophy, where the "intelligence" of the community is distributed across its relational network.

Distributed cognition also provides the theoretical basis for understanding modern team science, surgical teams, air traffic control, and other high-stakes collaborative enterprises where cognitive performance is a system-level property that cannot be reduced to the abilities of any individual team member.

Current measurement

The measurement of collective intelligence remains at an early stage. Woolley and Malone developed an online battery of collective intelligence tests, modeled on the methodology of their original study: groups perform multiple diverse tasks, and a c factor is extracted from the correlations among task performances. This approach has genuine psychometric rigor (it mirrors the factor-analytic methodology that produced the individual g factor), but it is labor-intensive, requires assembling groups, and has not yet been standardized for widespread use.

Group performance assessment in organizational settings typically uses outcome-based measures (team productivity, project success, error rates) rather than psychometric tests of collective intelligence. The 360-degree feedback approach (Chapter 25) captures some aspects of collaborative competence, but it assesses individuals' contributions to group functioning rather than the group's emergent cognitive capacity.

The measurement gap here reflects a deeper conceptual challenge: the psychometric tradition was built to measure individuals. Its entire apparatus (individual testing, individual scores, individual norms) assumes that the unit of analysis is the single person. Measuring collective intelligence re-

quires a different apparatus: one that tests groups, produces group-level scores, and norms group performance against other groups. This apparatus is being built, but it is decades behind the individual-testing tradition.

Frontier questions

Multi-agent modeling and swarm intelligence research in AI demonstrate that collective cognitive capacities can emerge from the interaction of simple agents following local rules, without any individual agent possessing the collective capacity. Ant colonies optimize foraging routes, bee swarms select nest sites, and bird flocks navigate obstacles, all through distributed processes that no individual agent directs. These biological examples provide proof of concept that intelligence can be genuinely collective, not merely aggregated.

Human-AI teaming represents perhaps the most consequential frontier. As AI systems become more capable, the relevant unit of cognitive performance is increasingly not the individual human or the individual AI system but the human-AI team. The "intelligence" of a physician using diagnostic AI, a lawyer using legal research tools, or a scientist using computational modeling is a property of the human-tool system, not of either component alone. This connects directly to the extended cognition thesis (Chapter 30): if cognitive tools are part of the cognitive system, then measuring individual intelligence without reference to the tools the individual habitually uses provides an incomplete picture.

Summary for the dimensional framework

Collective and distributed intelligence challenges the foundational assumption of the psychometric tradition: that intelligence is a property of individuals. The Ubuntu tradition, the Inuit concept of *Piliriqatigiingniq*, and modern collective intelligence research all suggest that some forms of cognitive performance are genuinely emergent, arising from the interaction of group members rather than from the sum of individual abilities.

The Woolley and Malone c-factor research provides the most rigorous empirical evidence for this claim, though the replication debate indicates that the relationship between individual and collective intelligence is more complex than the original findings suggested. Distributed cognition theory provides the conceptual framework for understanding how cognitive processes can be spread across people, tools, and environments.

A framework for human excellence that measures only individual capacities, however thoroughly, will miss the dimension that Ubuntu, Inuit collaborative culture, and modern team science all identify

as essential: the capacity of groups to think, decide, and act together in ways that exceed what any individual member could achieve alone.

Chapter 32: Recognition, Selection, and Measurement: A Comparative History

Movement II has mapped twelve dimensions of human excellence across traditions and modern research. This final chapter of the movement steps back from the dimensions themselves to examine the methods by which different traditions have recognized, validated, and selected for excellence. Modern standardized testing is one point on a much longer spectrum. This chapter reconstructs that spectrum, argues that every tradition on it is conducting genuine measurement, and assesses what standardization gained and what it cost.

The spectrum of measurement traditions

Humanity has always needed to identify who is excellent and who is not. Every complex society must allocate roles, responsibilities, and resources, and doing so requires some means of recognizing competence. The methods vary enormously, but they share a common structure: a domain of excellence is defined (however implicitly), a procedure for assessing it is established, a standard of adequate performance is applied, and consequences follow from the assessment (selection, advancement, exclusion, honor).

What follows is a comparative overview of seven families of measurement that the traditions employed.

1. Formal examination

The oldest, largest, and most influential examination system in human history was the Chinese Imperial Examination (*keju*), which operated in various forms from 605 CE to 1905 CE (Chapter 6). At its peak, the *keju* assessed millions of candidates through a multi-tiered system of written examinations testing knowledge of the Confucian classics, literary composition (the "eight-legged essay"), and the ability to apply classical learning to practical questions of governance. The system was designed to select officials on the basis of merit rather than birth, and it shaped Chinese society, education, and intellectual life for thirteen centuries.

The *keju* was, in many respects, a precursor of modern standardized testing. It used written examinations administered under controlled conditions (candidates were isolated in individual cells, sometimes for days). It employed anonymous grading (papers were recopied by clerks to prevent identification by handwriting). It applied uniform scoring criteria across all candidates. And it was consequential: success or failure on the examination determined a person's social position, career prospects, and family fortunes.

Scholastic disputation in the medieval European university (Chapter 5) was a different form of formal examination. Candidates for degrees were required to defend theses publicly against challengers, demonstrating not only knowledge of the material but the ability to reason under pressure, respond to objections, and articulate positions with precision. The examination was oral, public, and adversarial, and the quality of performance was judged by faculty members who had observed the candidate over years of study.

Modern standardized testing (Chapters 16-17), from Binet's original intelligence test through the WAIS, SAT, GRE, and PISA, inherited from the *keju* the principle of controlled administration, uniform scoring, and consequential selection, while abandoning the *keju*'s emphasis on deep textual knowledge and moral cultivation in favor of abstract reasoning, processing speed, and domain-general cognitive ability.

2. Debate and dialectic

Several traditions developed formal debate as a method of assessing intellectual excellence.

Indian *śāstrārtha* (scholarly debate) was a rigorous institutional practice in which scholars from different philosophical schools engaged in structured argumentation before qualified judges (Chapter 7). The Nyāya tradition developed a detailed taxonomy of valid and invalid argumentative moves, and the outcome of formal debates could determine the institutional fortunes of entire philosophical schools. The Buddhist tradition adopted and refined these debating practices, and they became central to Tibetan monastic education (Chapter 8), where the Geshe degree required years of formal debate training. Tibetan debate is a physically expressive, publicly conducted assessment in which the challenger claps, stamps, and gestures while pressing logical questions, and the defender must respond correctly or concede.

Greek dialectic (Chapter 4) was less formally structured but served a similar function: Socratic questioning, Platonic dialogue, and Aristotelian disputation were all methods for testing the quality of a person's reasoning through live intellectual exchange.

What these debate traditions share is an emphasis on reasoning under pressure, in real time, in public, before qualified judges. They assess not only whether a person knows the right answer but whether they can defend it against sophisticated objections. This is a form of measurement that captures something that written examinations miss: the dynamic, interactive quality of intellectual competence.

3. Teacher and elder authorization

Several traditions measured excellence through the judgment of a recognized authority who had observed the candidate over an extended period.

The Islamic *ijazah* (Chapter 9) is an authorization granted by a teacher to a student, certifying that the student has mastered a specific body of knowledge and is qualified to transmit it to others. The *ijazah* is personalized (granted by a specific teacher to a specific student), content-specific (authorizing transmission of particular texts or chains of knowledge), and based on extended observation (the teacher has worked with the student over months or years). The system treats the teacher's judgment, informed by prolonged direct acquaintance with the student's abilities, as the gold standard of assessment.

The Māori *whare wānanga* (Chapter 12) similarly relied on the judgment of senior *tohunga* who had trained the student over years and could assess not only what the student knew but how they had integrated that knowledge into their character, conduct, and communal engagement.

European craft apprenticeship assessed competence through the "masterpiece": a demonstration piece produced by the apprentice under the supervision of master craftsmen, who judged whether the work met the standards of the guild. The apprentice had been observed by the master over years of training, and the masterpiece was the culmination of a long developmental process.

What these authorization traditions share is an emphasis on expert judgment based on extended observation. They are the opposite of anonymous, standardized testing: they are personal, contextual, and informed by the kind of knowledge that only prolonged relationship can produce.

4. Demonstrated performance

Some traditions assessed excellence through demonstrated competence in the actual domain of practice, with no separation between the assessment and the real-world activity.

Pacific wayfinding (Chapters 12 and 30) validated navigational expertise through the most demanding possible test: did the navigator bring the canoe safely to the intended destination? Success was binary, consequential, and directly observable. Inuit hunting competence (Chapter 12) was assessed by elders who had observed the hunter in the field over years. Clinical competence in medicine is assessed through a combination of examination, supervised performance, and peer evaluation.

What demonstrated-performance traditions share is the refusal to separate assessment from the activity being assessed. They do not use proxy measures but instead observe performance directly. This eliminates the ecological validity problem but introduces others: it is labor-intensive, difficult to standardize, and impossible to scale to large populations.

5. Self-report and ratings

Modern psychology has developed self-report instruments for dimensions that are difficult to assess through performance tests. Personality inventories (Chapter 29), trait EI measures (Chapter 24), wisdom self-report scales (Chapter 28), and informant-completed rating scales (the BRIEF2 for executive function, Chapter 26; 360-degree feedback for social competence, Chapter 25) all rely on verbal descriptions of behavior rather than observation of performance. These instruments are efficient, scalable, and well-normed, but they are vulnerable to social desirability bias and the gap between self-description and actual behavior.

6. Behavioral tasks

Modern cognitive science has developed laboratory tasks that present controlled stimuli and measure behavioral responses. The Stroop Test and Trail Making Test (Chapter 26), the MSCEIT (Chapter 24), the RMET (Chapter 25), the TTCT (Chapter 27), and SJTs (Chapter 23) all measure actual performance rather than self-perception. But as the dimension chapters documented, they often show modest ecological validity, uncertain construct validity, and limited cultural generalizability.

7. Computational modeling

The newest entry on the measurement spectrum includes AI-scored responses, computational phenotyping of learning rates (Chapter 21), and machine learning approaches to assessing cognitive profiles. Grossmann's group has suggested that accurate wisdom assessment might ultimately require monitoring behavior continuously for weeks and analyzing it using machine learning (Chapter 28). Computational modeling promises scalability and the capacity to capture complex patterns, but it raises concerns about transparency, cultural bias, and the risk of automating the very narrowing this book documents.

What all these traditions share

Despite their enormous differences in method, all seven families share a common structure. They all define a domain of excellence. They all establish a procedure for assessment. They all apply a standard of adequate performance. And they all attach consequences to the outcome.

In this fundamental sense, the Māori elder evaluating a student's recitation of *whakapapa*, the Tibetan geshe examiner judging a debate, the Islamic scholar granting an *ijazah*, the Inuit elder assessing a young hunter, and the psychometrician administering a WAIS are all doing the same thing: mea-

asuring human excellence against a standard and certifying the result. The methods differ, but the purpose is constant. They are all genuine forms of measurement in the sense that matters: recognition, validation, and proof of competence.

What standardization gained

Modern standardized testing achieved three things that the earlier traditions could not.

Comparability. Standardized tests produce scores on a common scale, making it possible to compare individuals assessed in different places, at different times, by different examiners. The *keju* achieved some degree of comparability through its uniform examination and anonymous grading, but most pre-modern assessment traditions produced judgments that were not comparable across assessors or contexts. Standardized testing made large-scale comparison possible for the first time.

Fairness (of a specific kind). By applying the same instrument under the same conditions to all candidates, standardized testing eliminates some sources of bias: the examiner's personal relationship with the candidate, the candidate's social status, the subjective impressions of the assessor. The *keju* was explicitly designed to counteract aristocratic privilege, and modern standardized testing inherits this egalitarian impulse. The fairness is real but limited: standardized tests eliminate bias from the assessment procedure while potentially introducing bias through the content of the test itself (cultural loading, linguistic assumptions, implicit value hierarchies).

Scale. Standardized testing can assess millions of people efficiently. No apprenticeship system, no elder-judgment tradition, and no debate-based examination can operate at the scale that modern societies require for educational placement, employment selection, and credential certification.

What standardization cost

The gains of standardization came at significant cost, documented throughout Movement II.

Context. Standardized testing eliminates the environmental, relational, and situational context in which excellence is normally exercised. This is not merely a practical limitation; for dimensions like practical judgment (Chapter 23), social intelligence (Chapter 25), executive function (Chapter 26), embodied skill (Chapter 30), and collective intelligence (Chapter 31), removing the context changes what is being measured.

Meaning. The assessment traditions that preceded standardized testing were embedded in systems of meaning: the *keju* assessed mastery of texts encoding a civilization's moral and philosophical commitments; the *ijazah* certified participation in a living chain of knowledge transmission; the

Geshe degree recognized decades of intellectual and spiritual formation. Standardized testing measures performance on tasks deliberately stripped of cultural meaning to achieve "culture-fairness." The gain in cross-cultural comparability comes at the cost of the moral, spiritual, and cultural significance that made the assessment meaningful to the traditions that conducted it.

Moral thickness. The traditions that valued moral character as integral to excellence (Chapter 29) assessed it as part of the same process that assessed cognitive capacity. The Confucian examination evaluated moral reasoning alongside textual knowledge. The *ijazah* was granted partly on the basis of the student's character. The *omolúàbí* ideal was recognized through decades of observed conduct. Standardized testing, by design, separates cognitive measurement from moral evaluation. This separation created the vacuum this book has documented: the dimension of excellence on which the traditions converged most strongly was excluded from the measurement system that replaced them.

Summary for Movement II

Movement II has mapped twelve dimensions of human excellence across traditions and modern measurement:

1. Reasoning and Abstraction (well measured)
2. Learning and Plasticity (undermeasured)
3. Knowledge and Memory (well measured but undervalued relative to reasoning)
4. Practical Judgment (modestly measured; context-dependent)
5. Emotional Attunement (three competing models; construct confusion)
6. Social and Relational Intelligence (weakly measured; psychometric criticism of leading tools)
7. Self-Regulation and Executive Control (measured but ecologically questionable)
8. Creativity and Generativity (modestly measured; validity debated)
9. Wisdom and Uncertainty Management (early-stage measurement; low convergence)
10. Moral Discernment and Character (excluded then partially reintroduced)
11. Embodied and Ecological Skill (no standardized measurement)
12. Collective and Distributed Intelligence (emerging measurement; replication controversy)

The pattern is clear. The dimensions that the psychometric tradition measures best (reasoning, knowledge) are the ones that most traditions valued as one component among many. The dimensions that most traditions placed at the center (practical judgment, wisdom, moral character, social and relational competence) are the ones that the psychometric tradition measures least well or not at all.

This is not because the traditions were confused about what matters. It is because the psychometric tradition optimized for a specific set of methodological virtues (standardization, comparability, scale) that are structurally incompatible with the dimensions of excellence that most traditions valued most highly. The measurement system that replaced the older traditions is genuinely powerful within its

domain. But its domain is narrower than its users typically assume, and the dimensions it excludes are not marginal additions to a complete framework. They are central to what most human civilizations meant by excellence.

Movement III will propose a replacement framework.

MOVEMENT III: THE REPLACEMENT

Why "intelligence" is a false unity, and what should replace it

Chapter 33: Why There Is No Single Natural Kind Here

This chapter makes the book's central philosophical argument. It is not a polemic against IQ testing, not a claim that cognitive abilities are unreal, and not a dismissal of the psychometric tradition's genuine achievements. It is something more precise: the argument that "intelligence," as a single natural kind, a unified capacity that explains human cognitive excellence, does not exist. The phenomena the word points at are real. The umbrella category that groups them is not.

The consciousness analogy

Consider consciousness. No one doubts that conscious experience is real: the redness of red, the painfulness of pain, the felt quality of being aware. But "consciousness" as a single unified phenomenon has proven extraordinarily difficult to define, measure, or explain. Philosophers of mind have distinguished phenomenal consciousness (what it is like to have an experience), access consciousness (the availability of information for reasoning and action), self-consciousness (awareness of oneself as a subject), and monitoring consciousness (the capacity to track one's own cognitive processes). These phenomena are related but distinguishable: a person can be access-conscious of information without being phenomenally conscious of it (as in blindsight), and can be phenomenally conscious without being self-conscious (as in absorbed flow states).

The lesson is not that consciousness is unreal. The lesson is that a single word can group genuinely different phenomena under one umbrella, and that the umbrella category can obstruct rather than advance understanding. "Consciousness research" has been hampered for decades by the assumption that there is one thing here to explain, when there may be several related but distinct things.

"Intelligence" faces the same problem. The word groups together reasoning ability, processing speed, working memory capacity, accumulated knowledge, practical judgment, emotional perception, social competence, self-regulation, creativity, wisdom, moral character, embodied skill, and collective cognitive performance. Movements I and II of this book have shown that these are genuinely different phenomena: they have different developmental trajectories (Gf peaks at 25, Gc peaks decades later), different neural substrates, different relationships to age, different cultural valuations, and different measurement properties. Grouping them under one word creates the illusion that measuring some of them (reasoning, processing speed) is measuring all of them, and that high performance on some implies high performance on the rest.

The psychometric evidence from within

The most important evidence against the unity of intelligence comes not from critics of the psychometric tradition but from within it.

The Cattell-Horn-Carroll (CHC) model, now the dominant framework in psychometric research, identifies multiple broad abilities at Stratum II: fluid intelligence (Gf), crystallized intelligence (Gc), short-term memory (Gsm), long-term retrieval (Glr), visual processing (Gv), processing speed (Gs), and others. These are distinguishable factors with different developmental trajectories, different neural correlates, and different relationships to real-world outcomes. The CHC model treats g (at Stratum III) as a higher-order factor that emerges from the correlations among the Stratum II abilities. But the status of g as a causal entity versus a statistical summary has been debated since Spearman first proposed it in 1904.

Van der Maas and colleagues at the University of Amsterdam proposed the mutualism model (2006), which offers a fundamentally different explanation of the positive manifold (the finding that cognitive tests correlate positively with each other). In the mutualism model, the positive manifold emerges not from a single underlying g factor but from reciprocal beneficial interactions between cognitive processes during development: improvements in one process support improvements in others, creating positive correlations without any single common cause. Computer simulations confirmed that mutualism produces the same positive manifold, hierarchical factor structure, increasing heritability with age, and other empirical signatures that g-factor models predict.

This means that the positive manifold, the statistical foundation on which the entire g-factor edifice rests, is genuinely ambiguous. It can be explained by a single common cause (g-factor theory) or by reciprocal developmental interactions (mutualism) or by shared environmental inputs (nutrition, education, stimulation that benefit all cognitive processes simultaneously). As Savi and colleagues (2019) argued, psychometric g should be interpreted as a formative variable (an index of the general quality of a cognitive system, like the Dow Jones Industrial Average for an economy) rather than as a reflective variable (a single causal entity that produces all cognitive performance). The debate between these interpretations remains unresolved: Gignac (2014, 2016) argued that empirical tests favor g-factor theory, while van der Maas and Kan (2016) argued that the mutualism model accounts for the same data with equal statistical precision.

Contemporary neuroscience largely agrees that although g is statistically unitary, the physiology behind it is multifactorial: a nontrivial amount of evidence renders it highly unlikely that g can be reduced to a single biological entity.

The cross-cultural evidence

The cross-civilizational survey in Movement I provides a different kind of evidence: not statistical but conceptual and historical.

Five civilizations independently developed ideals of the excellent person that integrate cognitive and moral excellence: the Aristotelian *phronimos*, the Confucian *junzi*, the Islamic *adīb*, the Yoruba *omolúàbí*, and the Nahua *tlamatini*. In every case, intellectual excellence without moral character was not merely incomplete but disqualifying. These traditions did not lack words for cognitive ability; they chose not to separate it from character because they understood these as aspects of a single integrated achievement.

Three civilizations independently developed formal logical systems (Greek deductive, Indian Nyāya inductive-deductive, Chinese Mohist analogical) without knowledge of each other's work. This confirms that formal reasoning is a genuine human capacity. But the three systems use different architectures, and modern IQ tests measure only the Greek type.

The Chinese *xin* (heart-mind) refuses the cognition/emotion separation. The Buddhist *prajñā-karuṇā* pairing insists that wisdom and compassion are structurally inseparable. The Ubuntu tradition challenges whether the individual extracted from relational context is a meaningful unit of analysis. These are not pre-scientific confusions. They are alternative conceptualizations that track real features of human cognition that the Western tradition's analytical separation obscured.

If intelligence were one natural kind, one would expect convergence on what it is, the way different cultures converge on the existence of fire, water, or death. Instead, what the traditions tracked overlaps substantially but bundles differently. The overlap confirms that the phenomena are real. The different bundling confirms that the category is constructed, not discovered.

The measurement evidence from Movement II

Movement II provided a systematic inventory across twelve dimensions. The pattern is devastating for the claim that "intelligence" is a single well-understood construct.

The three-model confusion in emotional intelligence (Chapter 24) means researchers using the word "EI" may be measuring an ability, a personality trait, or a grab-bag of social competencies. The ecological validity gap in executive function (Chapter 26) means laboratory tasks predict only 18-20% of real-world self-regulation. The RMET's psychometric collapse (Chapter 25) means the most widely used measure of social cognition may be measuring vocabulary rather than social understanding. The modest convergence of wisdom measures (Chapter 28) means different instruments disagree on whether wisdom increases with age.

And there is one dimension (embodied-ecological skill, Chapter 30) for which no standardized measurement exists at all, despite its centrality to several major traditions.

If intelligence were one natural kind, its measurement would converge. Instead, it fragments into twelve partly distinct dimensions with different measurement properties, different cultural valuations, different developmental trajectories, and different relationships to real-world functioning. The fragmentation is not a failure of measurement. It is evidence about the nature of the thing being measured.

The strongest objection: "But g is real and predicts things"

The most sophisticated defense of the single-construct view argues that g, the general factor, predicts important outcomes: educational attainment, job performance, health behaviors, longevity. If g predicts these things, then whatever it is, it matters.

This objection deserves a precise answer.

First, predictive power does not prove ontological unity. Body mass index (BMI) predicts cardiovascular disease, diabetes, and mortality with moderate accuracy. But BMI is not a natural kind: it is a ratio of weight to height that compresses body composition, fitness, bone density, and metabolic health into one number. Its predictive power comes from the fact that its components correlate with outcomes, not from the existence of a single underlying "body condition." Gross domestic product (GDP) predicts national prosperity, but a country with high GDP from oil extraction and a country with high GDP from technology innovation are not the same kind of economy.

g may be like BMI or GDP: a useful summary statistic that predicts outcomes because its components matter, not because it captures a single underlying reality. The positive manifold ensures that any battery of cognitive tests will produce a first principal component, just as any collection of national economic statistics will. The existence of this component does not prove that there is one thing being measured.

Second, and more important: g predicts outcomes within the domain of what the psychometric tradition measures. It predicts educational attainment (assessed by methods resembling IQ tests), job performance (partly predicted by the same cognitive abilities IQ tests measure), and health behaviors (influenced by socioeconomic factors that correlate with IQ). But g does not predict the dimensions the tradition excluded. A first principal component of reasoning, processing speed, working memory, and vocabulary tells you very little about practical judgment (SJTs show incremental validity above IQ), moral character (the VIA captures variance unexplained by cognitive ability), embodied skill (no standardized measurement exists), or collective intelligence (the c-factor is not strongly correlated

with individual IQ).

The strongest form of the argument: *g* is real as a statistical regularity within the domain of cognitive abilities that the psychometric tradition chose to measure. It is not real as a comprehensive account of human cognitive excellence. The difference between these two claims is the difference between "this index summarizes a specific set of measurements" and "this index captures the essence of what it means to be cognitively excellent." The first is defensible. The second is not.

What follows

If intelligence is not a single natural kind, then the task is not to find the one true definition of intelligence but to replace the single-axis model with a framework adequate to the plurality. The traditions were not confused when they tracked multiple dimensions, bundled them differently, and refused to reduce them to one. They were describing what is actually there: a plural field of capacities that develop together, interact with each other, and contribute differently to different forms of human flourishing.

The next chapter proposes such a framework.

Chapter 34: A Plural Framework for Human Excellence

Chapter 33 argued that “intelligence” is not a single natural kind. This chapter proposes what should replace it: not a new single score, not a renamed version of IQ, but a dimensional framework consisting of twelve partly independent axes of human excellence that different traditions have tracked, different measurement systems assess, and different contexts weight differently. The framework is the book’s constructive contribution, and it must meet three tests: it must be grounded in the evidence assembled in Movements I and II, it must be honest about what can and cannot currently be measured, and it must be useful for the practical domains (education, institutions, AI alignment) that Movement III’s remaining chapters will address.

The twelve dimensions

Each dimension is presented here with a compressed summary of what Movement II established: what the traditions tracked, what modern measurement captures, and what remains unresolved.

1. Cognitive power. The capacity to reason abstractly, detect patterns, and solve novel problems. This corresponds to fluid intelligence (*Gf*) in the CHC model and is what IQ tests most directly measure. Chapter 20 established that three civilizations independently developed formal logical systems (Greek deductive, Indian Nyāya inductive-deductive, Chinese Mohist analogical), confirming that formal reasoning is a genuine human capacity. But the three systems use different architectures, and modern IQ tests measure only the Greek type. Even within the single best-measured dimension, measurement is narrower than the phenomenon. *Current measurement:* strong (Raven’s Progressive Matrices, WAIS Matrix Reasoning, Figure Weights). *Limitation:* captures only one type of formal reasoning.

2. Learning capacity. The ability to acquire new knowledge and skills, to respond to instruction, and to transfer learning to novel contexts. Chapter 21 established that Vygotsky’s Zone of Proximal Development and Feuerstein’s dynamic assessment provide the theoretical and practical foundation for measuring this dimension, and that dynamic assessment has been found to be more accurate than static IQ tests for reflecting the learning potential of minority and learning-disabled children. *Current measurement:* underdeveloped. Dynamic assessment (test-teach-retest) is methodologically sound but harder to standardize, more expensive, and less scalable than static testing. The dimension is undermeasured not because it is unimportant but because it is methodologically inconvenient.

3. Retained knowledge. The organized body of knowledge, expertise, and skills accumulated over a lifetime. This corresponds to crystallized intelligence (*Gc*) in the CHC model. Chapter 22 doc-

umented the *Gf/Gc* value inversion: most traditions valued accumulated knowledge and wisdom above novel reasoning, while psychometrics reverses this hierarchy by treating *Gf* as more fundamental. The developmental trajectories confirm the traditions' intuition: *Gc* peaks decades after *Gf*, and for the longest portion of adult life, accumulated knowledge matters more than raw processing speed. *Current measurement*: strong (WAIS Vocabulary and Information subtests, PISA, PIAAC). *Limitation*: the vocabulary and general knowledge assessed by standardized tests represent only a fraction of what the traditions meant by mastered knowledge (Confucian textual learning, Islamic hadith scholarship, Māori whakapapa, Yoruba Ifá corpus).

4. Practical judgment. The capacity to act well in messy, real-world situations where rules do not fully determine what to do. Chapter 23 established that this is the dimension where the gap between what the traditions valued and what psychometrics measures is largest. Aristotle's *phronēsis*, Binet's own definition of intelligence, Wechsler's definition, Sternberg's practical intelligence, Akan *nyansa*, and the Yoruba *omolúàbí* all center this capacity. SJTs show a criterion-related validity of $r = .26$ for predicting job performance, with incremental validity above IQ and personality, and lower adverse impact ($d = 0.38$ vs. $d = 1.0$ for cognitive tests). Kant's analysis explains why judgment resists standardization structurally: judgment is the capacity to apply rules, and this capacity cannot itself be taught by rules. *Current measurement*: modest (SJTs, tacit knowledge inventories, clinical judgment assessment). *Limitation*: context-dependent by nature; stripping context changes what is measured.

5. Emotional attunement. The capacity to perceive, understand, and use emotional information in oneself and others. Chapter 24 documented the three-model confusion: ability EI (Salovey and Mayer, measured by the MSCEIT), trait EI (Petrides, measured by the TEIQue), and mixed models (Goleman, Bar-On) are not the same construct, do not use the same measurement methods, and do not produce the same results. The Chinese concept of *xin* (heart-mind) and the Buddhist inseparability of wisdom and compassion suggest that the modern separation of cognition from emotion was a methodological choice, not a discovery about the structure of the mind. *Current measurement*: contested (MSCEIT for ability EI, TEIQue for trait EI, EQ-i 2.0 for mixed EI, LEAS for emotional awareness). *Limitation*: three competing models measuring different constructs under the same name.

6. Social-relational competence. The capacity to understand other people, navigate social dynamics, and sustain productive relationships. Chapter 25 established that the most widely used measure of social cognition, the Reading the Mind in the Eyes Test (RMET), has drawn serious psychometric criticism: poor factor structure, uncertain construct validity, and influence from vocabulary and culture rather than social perception per se. Second-person neuroscience confirms what the Ubuntu tradition claimed philosophically: social intelligence may be fundamentally interactive, existing in

the relationship rather than in the individual. *Current measurement:* weak (RMET under criticism, Social Shapes Test experimental, 360-degree feedback captures reputation rather than competence). *Limitation:* the measurement paradigm (presenting static stimuli to isolated individuals) may miss the core of the phenomenon.

7. Self-regulation. The capacity to direct, inhibit, and monitor one's own cognition and behavior. Chapter 26 documented the ecological validity gap: standard executive function tests (Stroop, Trail Making, Wisconsin Card Sorting) account for only 18-20% of the variance in everyday executive ability. The traditions that valued self-regulation (Stoic *prosoche*, Confucian *xiuyang*, Buddhist *sati/samādhi*, scholastic *prudentia*) all treated it as a practiced capacity that develops over decades, not a fixed quantity assessable in a single session. *Current measurement:* moderate (performance tasks for lab-based assessment, BRIEF2 for real-world ratings). *Limitation:* the two approaches measure different things and show only modest convergence.

8. Creative generativity. The capacity to produce novel, useful, or surprising ideas, artifacts, or solutions. Chapter 27 established that creativity occupies a unique position in the framework: valued by all traditions, central to few. The Torrance Tests of Creative Thinking show modest longitudinal validity but are contested as measures of genuine creativity. Islamic creativity-within-constraints demonstrates that originality can flourish under formal rules, challenging the Romantic assumption that creativity requires freedom from constraint. *Current measurement:* modest (TTCT, PISA 2022 Creative Thinking, Creative Achievement Questionnaire). *Limitation:* divergent thinking tests may measure fluency rather than genuine creativity; domain-general tests may miss domain-specific creative performance.

9. Wisdom. Reflective, perspective-taking, uncertainty-aware judgment with ethical depth. Chapter 28 established that this is the dimension most traditions placed at the summit and modern psychology measures least well. The Berlin Wisdom Paradigm, Ardel's Three-Dimensional Wisdom Scale, and Grossmann's Situated Wise Reasoning Scale all measure something related to wisdom, but their modest convergence means they may not be measuring the same construct. Different measures disagree on whether wisdom increases with age. The key philosophical finding: wisdom is constituted by the integration of cognitive, emotional, moral, and experiential capacities, and integration is what decomposition destroys. *Current measurement:* early-stage (BWP, 3D-WS, SWIS, SD-WISE). *Limitation:* low convergence among measures; the dimension may resist decomposition by nature.

10. Moral reliability. The disposition to act on what one knows to be right. Chapter 29 documented the psychometric tradition's deliberate exclusion of moral character (Allport's 1930s move) and the subsequent attempts to reintroduce it through the Big Five (conscientiousness, agreeableness), the VIA Classification of Character Strengths, and SJTs. The central tension: psychometrics asks "how

much of this trait do you currently possess?” while the virtue traditions ask “are you becoming more virtuous?” *Current measurement:* partial (Big Five captures some moral content incidentally, VIA assesses 24 character strengths but its factor structure doesn’t replicate its theoretical virtue categories). *Limitation:* the measurement-versus-formation tension; single-point assessment captures a snapshot but not a trajectory.

11. Embodied-ecological skill. Knowing shown in practice, place, and bodily attunement. Chapter 30 established that this is the only dimension with no standardized measurement at all. Pacific wayfinding, Inuit ecological knowledge, and Japanese embodied praxis (*dō* disciplines) all represent forms of excellence that are assessed rigorously through apprenticeship and demonstrated performance but that have zero representation in the psychometric toolkit. The 4E cognition framework (embodied, embedded, enacted, extended) provides the theoretical vindication. *Current measurement:* none in standardized form; validated through apprenticeship and demonstrated performance. *Limitation:* the dimension may resist standardization by nature, because removing the body and environment removes the medium in which the capacity operates.

12. Collective capability. The capacity to contribute to and sustain group intelligence. Chapter 31 documented the Woolley/Malone c-factor research (collective intelligence not strongly correlated with individual IQ, correlated with social sensitivity and conversational equality) and the Bates/Gupta replication failure (individual IQ accounts for approximately 80% of group performance). The dimension is genuinely contested but conceptually important: the Ubuntu tradition, the Inuit concept of *Piliriqatigiingniq*, and modern team science all treat group cognitive performance as irreducible to individual ability. *Current measurement:* emerging (group task batteries, c-factor methodology). *Limitation:* replication controversy; the psychometric tradition’s entire apparatus assumes individual testing.

Why these twelve and not eight or twenty

The twelve dimensions are not arbitrary. Each meets four criteria.

First, each is **empirically partly separable**. The correlations among them are positive (the positive manifold extends beyond what IQ tests measure) but far from perfect. Ability EI correlates only modestly with IQ ($r = .20-.30$). SJTs show incremental validity above both IQ and personality. Executive function ratings predict real-world outcomes that executive function tests miss. Creativity correlates with IQ at approximately $r = .20-.30$. These are not independent axes, but they are distinguishable ones.

Second, each has been **independently identified by multiple traditions**. No dimension in this

framework is a Western invention projected onto other cultures. Cognitive power was tracked by Greek, Indian, and Chinese logicians. Retained knowledge was central to Confucian, Islamic, Māori, and Yoruba traditions. Practical judgment was identified by Aristotle, Binet, the Akan, and the Confucians. Emotional attunement was integrated into the Chinese *xin*, Buddhist *prajñā-karuṇā*, and Yoruba *ogbón-inú*. The cross-civilizational convergence on these dimensions is the strongest evidence that they track real features of human cognition rather than culturally local categories.

Third, each can **in principle be assessed**, even if standardized instruments do not yet exist for all of them. The traditions that valued embodied skill and collective capability assessed them rigorously through demonstrated performance and expert observation. These are genuine assessments; they simply do not take the form of standardized, decontextualized, single-score tests. A framework that excluded dimensions merely because they resist one particular form of assessment would be confusing the limits of a method with the limits of reality.

Fourth, **no single dimension reduces cleanly to any other**. Practical judgment and wisdom overlap substantially, but they are distinguishable (judgment is about action in specific situations; wisdom is about reflective understanding of life's larger patterns). Emotional attunement and social-relational competence share variance (social sensitivity predicts both), but a person can be emotionally perceptive and socially clumsy, or emotionally unaware and socially skilled through learned routines. Self-regulation and moral reliability both involve governing one's behavior, but self-regulation is about cognitive control while moral reliability is about ethical disposition.

If the criteria were applied more strictly (demanding perfect separability), the list would shrink. If applied more loosely (allowing any capacity that any tradition valued), it would expand. Twelve represents the level of resolution at which the dimensions are empirically distinguishable, cross-culturally attested, and practically useful without becoming so fine-grained that the framework loses coherence.

What holds them together

The twelve dimensions are not held together by a single underlying trait (as g-factor theory holds for cognitive abilities) but by a shared functional context. They are all capacities that enable agents to perceive, learn, judge, act, relate, and adapt within social and ecological worlds. A human being who excels on all twelve is not exhibiting twelve separate excellences. They are functioning well as a whole person in a complex world.

The traditions recognized this unity of function even while distinguishing the component capacities. The Confucian *junzi* is excellent not because they score high on twelve separate scales but because

their cognitive power, knowledge, judgment, emotional attunement, self-regulation, and moral character operate in harmony. The Buddhist practitioner who has completed the Geshe curriculum has developed reasoning, knowledge, memory, self-regulation, and wisdom in an integrated training program, not in twelve separate modules. The *omolúàbí* is recognized not through a profile of subscores but through the integrated quality of their conduct over a lifetime.

The framework this book proposes is analytical: it distinguishes twelve dimensions for the purpose of understanding, measuring, and cultivating them. But the goal of human excellence, as the traditions understood it, is integration: the harmonious functioning of all these capacities in the service of a well-lived life. The framework is a tool for analysis. Integration is the goal of formation.

Convergence without shared metaphysics

Different traditions weight the twelve dimensions differently. The Confucian tradition weights retained knowledge, moral reliability, and self-regulation heavily. The Greek philosophical tradition weights cognitive power, practical judgment, and wisdom. The Ubuntu tradition weights social-relational competence and collective capability. The Pacific navigational tradition weights embodied-ecological skill. The Buddhist tradition weights wisdom, self-regulation, and emotional attunement. None of these weightings is wrong. They reflect different ecological, social, and cultural contexts in which different dimensions matter more or less. A framework for human excellence should accommodate these different weightings without declaring any of them the correct one. What the framework provides is a common language: a set of dimensions that are recognizable across civilizations, even when civilizations disagree about their relative importance.

This is the convergence-without-shared-metaphysics principle introduced in Chapter 3. The dimensions converge (every tradition tracked most of them). The metaphysics diverge (different traditions embedded them in different cosmological, theological, and philosophical frameworks). The convergence is evidence that the dimensions are real. The divergence is evidence that the weighting is contextual.

A framework adequate to human excellence must be plural enough to accommodate different weightings, rigorous enough to demand evidence for each dimension, and honest enough to acknowledge where measurement is strong, where it is weak, and where it does not yet exist.

Chapter 35: What Should Be Measured Separately

Chapter 34 proposed twelve dimensions. This chapter asks a harder question: how should they be assessed in practice? Some dimensions should never be collapsed into a single score. Others cluster naturally and may be assessed together. The right stance is not twelve separate tests administered to every person on every occasion, but profile-based assessment: recognizing that a person (or an AI system, or a team) has a shape across these dimensions, not a position on a single axis.

The case for profiles over scores

The psychometric tradition already knows that profiles are more informative than single scores. The WAIS produces four index scores (Verbal Comprehension, Perceptual Reasoning, Working Memory, Processing Speed) that map onto different CHC broad abilities. A clinician examining a WAIS report pays as much attention to the pattern of index scores as to the Full Scale IQ. A 60-year-old with a Verbal Comprehension Index of 130 and a Processing Speed Index of 95 has a very different cognitive profile from a 25-year-old with both indices at 112, even if their Full Scale IQs are similar. The profile reveals what the single score conceals.

But the Full Scale IQ persists because institutions demand a single number. College admissions, employment screening, disability determinations, and gifted program placements all want one score that ranks candidates on one axis. The single number is not more informative than the profile; it is more administratively convenient. The convenience comes at a real cost: it obscures the between-dimension variation that is often the most practically important information.

The evidence from Movement II demonstrates that the cost extends far beyond the WAIS profile. SJTs show incremental validity above IQ for predicting job performance (Chapter 23): they capture information about practical judgment that cognitive ability tests miss. Ability EI shows incremental validity above IQ and personality for predicting relationship quality and certain workplace outcomes (Chapter 24). Executive function ratings predict real-world self-regulatory behavior that executive function tests do not (Chapter 26). In every case, the additional dimension carries information that a single cognitive score destroys. Each time a researcher demonstrates incremental validity, they are proving that the single-score model is incomplete.

Which dimensions are most separable

The twelve dimensions are not equally separable. Some can be reliably distinguished from each other; others overlap substantially. Understanding the structure of separability and overlap is neces-

sary for designing practical assessment systems.

Cognitive power and emotional attunement are among the most clearly separable dimensions. Ability EI, as measured by the MSCEIT, correlates only modestly with IQ (approximately $r = .20-.30$). This means that knowing a person's IQ tells you relatively little about their emotional perception, understanding, and management, and vice versa. A person can be an exceptional abstract reasoner and an emotional novice, or can have profound emotional intelligence with modest reasoning ability. The traditions that integrated these (Chinese *xin*, Buddhist *prajñā-karuṇā*) were not denying their separability; they were insisting that both are needed for excellence.

Cognitive power and moral reliability are separable by both evidence and history. Allport's deliberate exclusion of moral character from the domain of personality measurement (and the broader psychometric tradition's exclusion from intelligence measurement) was a methodological choice, not a scientific discovery. The VIA Classification captures variance in character strengths that is only partially explained by the Big Five personality dimensions (mean of 33% explained variance across the 24 strengths), confirming that moral character is partly distinct from both cognitive ability and personality as conventionally measured. The traditions' insistence that a brilliant person without character is dangerous rather than excellent reflects a genuine separability: cognitive power does not guarantee moral reliability.

Fluid intelligence and crystallized intelligence are one of the most robust separabilities in all of psychometrics. The Cattell-Horn distinction has been replicated countless times, and the different developmental trajectories (*Gf* peaking in the mid-twenties, *Gc* continuing to grow into middle age and beyond) provide biological confirmation. The two factors share variance (Cattell's investment theory explains the correlation: higher *Gf* enables faster *Gc* accumulation), but they are empirically, developmentally, and neurologically distinguishable. This separability is the foundation for the book's argument that the psychometric tradition's privileging of *Gf* over *Gc* is a value choice, not a scientific necessity.

Individual performance and collective capability are the most conceptually radical separability in the framework. If the Woolley/Malone c-factor research is correct (and the replication controversy means this is genuinely uncertain), then group cognitive performance is at least partially distinct from the aggregate of individual cognitive abilities. A group of individually brilliant people who cannot coordinate, who dominate each other's conversational contributions, and who lack social sensitivity may perform worse than a group of moderately able people who collaborate well. Whether this separability is robust remains an open empirical question, but the Ubuntu and Inuit traditions treated it as obvious: collective competence is not reducible to individual ability.

Which dimensions cluster

Some dimensions share enough conceptual and empirical overlap that they tend to move together, even if they are not identical.

Practical judgment and wisdom are the most natural cluster. Both involve judgment under uncertainty. Both require accumulated experience. Both resist standardization for the same structural reason (Kant's insight about judgment). Both integrate cognitive, emotional, and moral capacities. The distinction between them is primarily one of scope: practical judgment addresses specific situations (should I take this job? How should I handle this conflict?), while wisdom addresses life's larger patterns (what matters most? How should one respond to suffering? What does it mean to live well?). In practice, the same person who exercises good practical judgment in daily decisions is often the person who displays wisdom in reflecting on their life. A framework for assessment might treat these as closely related dimensions that can be assessed partly through the same methods (situated reasoning tasks, advice-giving paradigms, case-based evaluation).

Self-regulation and moral reliability both involve governing one's own behavior, but they diverge in content. Self-regulation is about cognitive control: staying focused, inhibiting impulses, switching flexibly. Moral reliability is about ethical disposition: being honest, keeping commitments, acting on principles. A person can have excellent self-regulation (they are disciplined, organized, and focused) and poor moral reliability (they use their discipline to pursue selfish ends). Conversely, a person can have strong moral commitments and poor self-regulation (they intend to do right but are impulsive and inconsistent). The Confucian tradition's insistence on integrating self-cultivation (*xiuyang*) with moral development (*ren, yi*) reflects an awareness that these are related but distinguishable: developing one without the other is incomplete.

Emotional attunement and social-relational competence share variance through the common mediator of social sensitivity (which predicts both emotional perception and effective social interaction). But they are partially separable: a person can be skilled at perceiving emotions in others (high emotional attunement) without being effective at navigating complex social dynamics (low social-relational competence), as when someone accurately reads a room but cannot influence it. The measurement instruments confirm partial separability: the MSCEIT (emotional ability) and social cognition measures show modest convergence, suggesting related but distinct constructs.

"Good enough" measurement for resistant dimensions

Several dimensions resist standardized measurement for reasons documented in Movement II. For these dimensions, the question is not "can we build a perfect test?" but "what does good enough

measurement look like?”

For wisdom: Grossmann’s Situated Wise Reasoning Scale (SWIS) represents the most promising current approach. It asks respondents to recall specific situations and rate the extent to which they demonstrated intellectual humility, recognition of uncertainty, consideration of multiple perspectives, and search for compromise. The situated approach captures the within-person variability that makes wisdom hard to assess globally (a person may reason wisely about relationships but unwisely about politics). Brienza et al. (2017) found that two to five episodes need to be sampled for sufficient reliability. This is more labor-intensive than a single-session test, but it provides genuinely useful information about a dimension that most traditions considered the highest form of excellence.

For embodied-ecological skill: Demonstrated performance assessed by qualified practitioners is the only currently available method, and it may be the only method that the dimension’s nature permits. The key to making this ”good enough” for broader use is specifying explicit assessment criteria, training evaluators for inter-rater reliability, and documenting the assessment process so that it can be scrutinized and improved. Medical residency evaluation, musical performance assessment, and martial arts grading all provide existing models for rigorous expert-evaluated demonstration.

For collective capability: Group task batteries modeled on the Woolley/Malone methodology (multiple diverse tasks administered to a group, with a collective intelligence factor extracted from the correlations among task performances) represent the most psychometrically rigorous approach. The replication concerns require that any such battery be validated across multiple samples and contexts before being used for consequential decisions. Team assessment in organizational settings (using a combination of group performance tasks, peer evaluations, and process observation) provides a less rigorous but more immediately practical alternative.

The general principle for ”good enough” measurement is threefold: reliable enough to inform decisions (the measurement must produce consistent results across occasions and evaluators), valid enough to improve on current alternatives (it must capture information that existing methods miss), and honest enough to acknowledge what it does not capture (every measurement has limitations, and the limitations should be stated, not hidden).

The shape of a person

The framework this book proposes does not rank people from best to worst on a single axis. It describes the shape of a person’s capacities across twelve dimensions. One person may have exceptional cognitive power and retained knowledge, moderate practical judgment and self-regulation, and limited social-relational competence and collective capability. Another may have moderate cognitive

power but exceptional wisdom, moral reliability, and embodied-ecological skill. Neither is "more intelligent" than the other. They have different profiles, suited to different contexts, roles, and forms of contribution.

This is not a retreat from rigor. It is an advance in precision. A single IQ score tells you that someone is "above average." A twelve-dimensional profile tells you *in what ways* they are above average, where they are average, and where they may need development or support. The profile is more useful for every practical purpose (education, hiring, team composition, self-development) than the single score, because it preserves the information that the single score compresses away.

The cost of this precision is complexity. A twelve-dimensional profile is harder to compute, harder to communicate, and harder to use for rank-ordering. It cannot be reduced to a headline number. It requires that the user think about what dimensions matter for a particular purpose, rather than defaulting to a single all-purpose ranking. This complexity is not a design flaw. It is a reflection of reality: human excellence is complex, and any representation that makes it simple is making it false.

Chapter 36: Civilizationally Plural Assessment

If the twelve dimensions proposed in Chapter 34 are real, and if different traditions legitimately weight them differently (as Chapter 34's convergence-without-shared-metaphysics principle holds), then assessment should be plural. Not in the sense of "anything goes," not in the sense that all assessments are equally valid for all purposes, but in the precise sense that what counts as excellence depends partly on context, community, and purpose, and that honest assessment must be transparent about which dimensions it measures and why.

This is not a theoretical proposal waiting to be tested. It is already happening.

Plural assessment without relativism

The twelve dimensions are not relative. Reasoning is a real capacity everywhere. Practical judgment matters in every human community. Moral character is valued across civilizations. The cross-civilizational survey in Movement I established this: the dimensions converge even when the metaphysical frameworks diverge.

What is relative is the weighting: how much each dimension matters for a particular purpose. A research university selecting graduate students in theoretical physics legitimately weights cognitive power and creative generativity heavily. A hospital selecting surgeons legitimately weights embodied-ecological skill, self-regulation, and practical judgment. A community selecting elders legitimately weights wisdom, moral reliability, and social-relational competence. A tech company assembling a product team legitimately weights collective capability, creative generativity, and cognitive power. None of these weightings is wrong. They reflect genuine differences in what the role requires.

The problem with the single-axis model is not that it weights cognitive power too heavily for some purposes. The problem is that it presents one weighting as the universal standard and then applies it across purposes for which it was not designed. Using an IQ-adjacent metric to select surgeons, community leaders, and team members is like using a speedometer to evaluate a vehicle's overall quality: speed is relevant, but it is not the only thing that matters, and optimizing for speed alone may sacrifice safety, fuel efficiency, and cargo capacity.

Plural assessment does not mean relativism because it retains two shared commitments across all assessment traditions.

First, **no assessment system should systematically exclude populations on the basis of characteristics irrelevant to the capacity being assessed.** The Chinese *keju* was designed to counteract

aristocratic privilege: it selected on the basis of textual mastery and judgment regardless of birth. The Islamic *ijazah* was granted on the basis of knowledge mastery regardless of social origin. The Māori *whare wānanga* selected students for aptitude and trained them regardless of rank. Modern standardized testing was itself invented partly to counteract patronage and favoritism in educational and employment selection. The commitment to assessing relevant capacity rather than irrelevant social characteristics is cross-civilizational, not culture-specific.

Second, **every assessment system should be transparent about what it measures and what it does not.** An IQ test that presents itself as measuring "intelligence" in general, when it actually measures abstract reasoning, processing speed, and vocabulary, is not transparent. An SJT that presents itself as measuring "judgment" without specifying what kind of judgment and in what domain is not transparent. A character assessment that presents itself as measuring "virtue" without specifying which virtues and by whose standards is not transparent. Honest assessment states which dimensions it targets, which methods it uses, and which dimensions it excludes.

Indigenous-led assessment in practice

The most important current examples of civilizationally plural assessment come from Indigenous communities that are developing their own assessment frameworks, grounded in their own epistemologies and values.

Kaupapa Māori assessment in Aotearoa New Zealand represents one of the most fully developed models. Kaupapa Māori theory, articulated by Graham Smith and others, holds that Māori values, knowledge, language, and ways of being are valid and legitimate, and that assessment should be derived from a Māori epistemological perspective. The national resource *Te Whatu Pōkeka: Kaupapa Māori Assessment for Learning* (Ministry of Education, 2009) provides a professional assessment framework based on Māori values, philosophies, and cultural contexts.

Kaupapa Māori assessment does not view the child in isolation but recognizes that the mokopuna (child) emerges from rich traditions and is linked strongly with *whānau* (family), *hapū* (sub-tribe), and *iwi* (tribe). Assessment begins with recognition of what children bring: their strengths, history, traditions, and *whakapapa* (genealogy). The child is understood as possessing *mana* (potential and spiritual power), *mauri* (living essence), and *wairua* (spiritual self), and assessment should enhance and nurture these rather than reduce the child to a deficit score.

In the terms of this book's framework, Kaupapa Māori assessment explicitly assesses dimensions that standardized Western testing ignores: relational competence (*whanaungatanga*), ecological and place-based knowledge (*tangata whenuatanga*), cultural identity and spiritual wellbeing, and the

contribution of learning to collective flourishing. These are not "soft" additions to "real" assessment. They are assessments of the dimensions that Māori culture places at the center of human excellence: social-relational competence, retained cultural knowledge, embodied-ecological skill, and collective capability.

The rigor of Kaupapa Māori assessment is different in kind from psychometric rigor but no less demanding. It requires deep knowledge of *te ao Māori* (the Māori world), sustained observation of children across contexts, collaboration with *whānau*, and documentation that makes visible learning valued within Māori epistemology. It is assessed by practitioners trained in Māori cultural knowledge and pedagogical practice, not by automated scoring algorithms.

Inuit educational frameworks in Canada and Greenland provide another model. The concept of *Piliriqatigiingniq* (working together for a common purpose) is treated as a core educational outcome, not as a supplementary social skill. Inuit curriculum frameworks assess children's development of collaborative competence, environmental attunement, and the capacity to contribute to group functioning under demanding conditions. These frameworks treat collective capability and embodied-ecological skill as primary dimensions of excellence, alongside the cognitive and literacy skills that standardized Western curricula emphasize.

First Nations Australian educational assessment is developing frameworks that recognize land-based knowledge, story-based learning, and community contribution as genuine forms of educational achievement. These frameworks challenge the assumption that educational assessment must take the form of individual, written, timed tests of abstract cognitive ability.

What these examples demonstrate

These Indigenous-led assessment projects are not "alternative" assessments in the sense of being softer, less rigorous, or less consequential than standardized testing. They are assessments of **different dimensions**, grounded in different value weightings, and they are at least as rigorous as standardized testing within their domains. A Kaupapa Māori practitioner's assessment of a child's *mana* development, relational competence, and cultural knowledge involves years of training, sustained observation, and expert judgment, the same qualities that make demonstrated-performance assessment rigorous (Chapter 30) and that the apprenticeship and teacher-authorization traditions relied on for centuries (Chapter 32).

What these frameworks lack is standardization, cross-population comparability, and scale: the three things that modern standardized testing provides and that the older traditions could not. What they provide is contextual richness, dimensional breadth, and cultural meaningfulness: the three things

that standardization sacrificed (Chapter 32).

A plural assessment system does not require choosing between these approaches. It requires being honest about what each approach provides and what it costs. Standardized testing provides comparability and scale at the cost of context, meaning, and dimensional breadth. Kaupapa Māori assessment provides cultural meaning, dimensional breadth, and relational richness at the cost of cross-population comparability. For different purposes, different assessment approaches are appropriate. The error is not in using standardized tests. It is in treating standardized tests as the only form of legitimate assessment and then applying them across all purposes, populations, and dimensions.

Connecting to the measurement spectrum

Chapter 32 documented seven families of measurement across the history of human assessment. Plural assessment draws on several of these families, selecting the approach most appropriate to the dimension being assessed.

Dynamic assessment (Chapter 21) is particularly aligned with the developmental emphasis of many traditions. The Confucian, Buddhist, and apprenticeship traditions all treated human development as the central object of educational concern: the question is not "what can you do now?" but "what are you becoming?" Dynamic assessment's test-teach-retest methodology captures exactly this: it measures learning potential rather than current performance, and it has been found to be more accurate than static IQ tests for reflecting the abilities of students whose prior educational experiences have been limited.

SJTs (Chapter 23), with their lower adverse impact ($d = 0.38$ vs. $d = 1.0$ for cognitive ability tests), represent a measurement approach that comes closer to what the traditions valued (practical judgment) while producing fairer outcomes across demographic groups. This is not a coincidence: an assessment method that measures a dimension more closely aligned with what most traditions considered central to excellence also happens to produce less disparate outcomes. The traditions' broader conception of excellence and modern assessment's narrower conception produce different patterns of inclusion and exclusion.

Demonstrated performance (Chapter 30) is the gold standard for embodied-ecological skill, and it is the method that the Pacific, Inuit, Japanese, and craft-apprenticeship traditions all used. Its weakness (lack of standardization and scale) is real, but for the dimension it assesses, no standardized alternative exists or may ever exist.

What plural assessment requires

Plural assessment is not a call for the abolition of standardized testing. Standardized tests serve real purposes: they provide comparability across populations, they counteract some forms of bias (favoritism, patronage, social-class discrimination), and they enable assessment at scales that no other method can match.

Plural assessment is a call for three things.

First, **honesty about what is being measured**. Every assessment should state which dimensions of excellence it targets and which it excludes. An IQ test that reports a "general intelligence" score should acknowledge that it measures abstract reasoning, processing speed, and vocabulary, and that it does not measure practical judgment, moral character, emotional attunement, embodied skill, or collective capability.

Second, **appropriateness of assessment to purpose**. The dimensions assessed should match the capacities that the role, program, or context actually requires. Selecting surgeons on abstract reasoning alone, when the role requires embodied skill, self-regulation, and practical judgment, is a mismatch between assessment and purpose. Selecting community leaders on cognitive ability alone, when the role requires wisdom, moral reliability, and social-relational competence, is a mismatch.

Third, **recognition that multiple assessment traditions can coexist**. Standardized testing, dynamic assessment, demonstrated performance, expert observation, and community-based evaluation are all genuine forms of measurement. They assess different dimensions, they have different strengths and limitations, and they are appropriate for different purposes. A mature assessment ecosystem would include all of them, deployed according to what needs to be measured and why.

Chapter 37: Human-AI Comparison Without Category Mistakes

The rise of AI makes the need for a plural framework urgent. Current AI benchmarking recapitulates the IQ mistake in real time: collapsing many capacities into composite scores, treating high performance on measured dimensions as evidence of general excellence, and ignoring the dimensions that are not measured. As of April 2026, the benchmarking landscape itself is producing the evidence for why this approach fails.

The benchmarking landscape in 2026

The AI evaluation ecosystem has grown rapidly but remains structurally committed to single-score reporting. The major benchmarks each test a narrow capacity:

MMLU (Massive Multitask Language Understanding) tests knowledge retrieval across 57 subjects through multiple-choice questions. Frontier models now score above 90%, and the benchmark is widely considered saturated. MMLU-Pro, an enhanced version with ten answer options and harder questions, has extended the runway, but the format remains the same: multiple-choice knowledge retrieval producing a single accuracy number.

GPQA Diamond tests graduate-level scientific reasoning in biology, physics, and chemistry. Frontier models have surged from 39% (November 2023) to above 90% (early 2026), approaching the level of domain experts. But GPQA tests reasoning within scientific domains using text-based questions. It does not test whether the system can apply scientific understanding to novel practical situations, exercise judgment about what questions to ask, or recognize the limits of its own knowledge.

ARC-AGI-3, released in March 2026, represents the most significant challenge to the single-score paradigm. Unlike its predecessors, ARC-AGI-3 is fully interactive: it presents hundreds of novel game-like environments with no instructions, no rules, and no stated goals. The AI agent must explore, discover how the environment works, infer what success looks like, and carry learning forward across increasingly difficult levels. Humans score 100%. The best frontier AI (Gemini 3.1 Pro) scores 0.37%. Every other frontier model scores below 1%.

This result is devastating for the claim that frontier AI systems are approaching general intelligence. The same models that score above 90% on MMLU and GPQA score below 1% on a benchmark that tests the capacity to learn from interaction in novel environments, exactly the capacity that Vygotsky's Zone of Proximal Development (Chapter 21) and the traditions' emphasis on adaptive learning identified as central to intelligence. ARC-AGI-3's designer, François Chollet, explicitly frames the benchmark in terms of the fluid/crystallized distinction: ARC-AGI-3 targets fluid intelligence (the

ability to reason through novel problems and adapt to new situations) rather than crystallized intelligence (retrieving memorized facts and applying trained heuristics). On crystallized measures, AI dominates. On fluid measures that require genuine adaptive learning, the gap between humans and AI remains 99.63 percentage points wide.

SimpleBench, created by AI Explained, tests common-sense reasoning using "trick" questions that are trivially easy for humans but systematically mislead LLMs. A human baseline of 83.7% outperforms all tested models. The BrainBench study (2026) found that even the best model (Claude Opus 4.6 with extended thinking) achieves only 80.3% on questions designed to be trivially easy for any adult human, and exhibits a 6-16 percentage-point gap between accuracy and consistency, revealing stochastic rather than reliable reasoning. The Opper.ai "Car Wash Test" found that the majority of frontier models recommended walking to return a rental car that was just across the street, failing to recognize that the car itself needed to be returned.

Humanity's Last Exam (HLE), developed by the Center for AI Safety, contains 2,500 of the hardest questions across mathematics, humanities, and natural sciences, contributed by nearly 1,000 experts. Frontier models remain far from completing it, confirming that deep expert-level reasoning across domains remains a genuine challenge.

Artificial Analysis Intelligence Index (v4.0) represents the most sophisticated attempt at multi-dimensional AI evaluation, incorporating ten separate evaluations across different domains. The HELM framework similarly evaluates models across multiple metrics: accuracy, calibration, robustness, fairness, bias, toxicity, and efficiency, producing a multi-dimensional profile rather than a single score. These are structurally aligned with the dimensional approach this book proposes.

What the benchmarking landscape reveals

The 2026 benchmarking landscape demonstrates three things relevant to this book's argument.

First, **benchmark saturation follows a predictable pattern that mirrors IQ test saturation**. As models improve on a benchmark, the benchmark loses its ability to discriminate between models, and harder benchmarks must be created. MMLU was hard in 2021 and saturated by 2024. GPQA Diamond was hard in 2023 and is approaching saturation in 2026. Each new benchmark raises the ceiling without changing the methodology. This is the same dynamic that produced the WAIS-IV from the WAIS-III, and the WAIS-V from the WAIS-IV: harder items, same format, same underlying assumption that more of the same kind of performance means more intelligence.

Second, **the gap between crystallized and fluid performance is enormous**. AI systems that score above 90% on knowledge retrieval and trained reasoning patterns score below 1% on tasks requiring

genuine adaptive learning in novel environments (ARC-AGI-3) and fail trivially easy common-sense reasoning questions that any human adult could answer (SimpleBench, BrainBench). This is the Gf/Gc distinction (Chapter 22) made dramatically visible: AI has achieved extraordinary Gc-like performance (vast knowledge, trained pattern application) while remaining profoundly limited on Gf-like capacities (novel reasoning, adaptive learning, common-sense inference). The traditions' insistence that accumulated knowledge is not the same as the capacity to figure out new things is confirmed by every new benchmark that tests the latter.

Third, **multi-dimensional evaluation is already emerging but is not yet the default.** Artificial Analysis, HELM, and LM Council's dashboard all provide multi-dimensional assessments. But the public discourse and the media coverage still collapse these into single narratives: "GPT-5 is the smartest AI ever." The dimensional information is available; it is the institutional and communicative habits that lag behind.

The twelve-dimensional profile of an LLM in 2026

Applying the framework from Chapter 34 to a current frontier LLM produces a profile that the single-score benchmarks conceal:

Cognitive power: High on certain reasoning types (mathematical, logical, pattern-based with chain-of-thought), but brittle. ARC-AGI-3's sub-1% scores demonstrate that the reasoning does not generalize to interactive, exploratory, goal-inferred environments. SimpleBench and BrainBench demonstrate that common-sense reasoning, trivially easy for humans, remains systematically unreliable.

Learning capacity: Severely limited. LLMs do not learn from individual interactions without fine-tuning. ARC-AGI-3 is explicitly designed to test this: can the system learn from experience inside each environment? The answer, as of March 2026, is almost entirely no.

Retained knowledge: Very high. This is the dimension where LLMs are strongest, and it corresponds to the saturated benchmarks (MMLU, GPQA in part). The traditions would recognize this as impressive Gc-like achievement while noting that it is the dimension they valued within a broader framework, not as a standalone measure of excellence.

Practical judgment: Absent in any robust sense. The Car Wash Test result (majority of models recommend walking to return a rental car across the street) is a precise illustration: the models lack the common-sense practical judgment that any human adult exercises automatically.

Emotional attunement, social-relational competence, wisdom, moral reliability, embodied-ecological skill, collective capability: Absent or not applicable, as analyzed in the previous version of this chapter and for the same structural reasons. No benchmark improvement on MMLU or GPQA

will change these absences, because they are absences of different dimensions, not lower scores on the same one.

The category mistake, updated

The category mistake is now more visible than it was even a year ago. A system that scores 94% on GPQA Diamond and 0.37% on ARC-AGI-3, that passes graduate-level physics exams and fails to recognize that a rental car must be driven to be returned, is not "almost generally intelligent." It is extraordinary on one cluster of dimensions (knowledge retrieval, trained pattern reasoning) and absent on another (adaptive learning, common-sense judgment, interactive exploration). Calling it "superintelligent" is like calling an encyclopedia a wise elder: the knowledge is vast, but the judgment, adaptability, and practical sense that the traditions placed at the center of excellence are not there.

The dimensional framework prevents this conflation by requiring separate reporting. A responsible evaluation would state: "This system scores at the 95th percentile on knowledge retrieval, above 90% on domain-specific reasoning under trained conditions, below 1% on novel interactive adaptive reasoning, and is not assessable on practical judgment, common-sense inference, emotional attunement, moral reliability, embodied skill, or collective capability." This is more complex than a single leaderboard ranking. It is also more honest, more useful for deployment decisions, and less likely to produce the catastrophic overconfidence that comes from treating crystallized performance as general intelligence.

Human-AI teaming revisited

Chapter 31 established that the relevant unit of cognitive performance is increasingly the human-AI system. The 2026 benchmarking data sharpens this point. AI systems contribute extraordinary knowledge retrieval and trained pattern reasoning. Humans contribute adaptive learning (100% on ARC-AGI-3 vs. sub-1% for AI), common-sense practical judgment (83.7% on SimpleBench vs. sub-50% for most models), moral responsibility, social-relational competence, and embodied understanding.

The complementary profile is not a temporary state waiting to be resolved by the next model release. The ARC-AGI series has tracked progress for seven years, and each new version reveals a new structural gap. The gap is not closing monotonically; it is being revealed in new domains as measurement improves. This pattern, strikingly similar to the pattern this book documented for human intelligence measurement, suggests that the dimensional structure is genuine: there are real,

partly separable capacities here, not a single axis along which AI is steadily advancing.

Chapter 38: Education After Intelligence

If intelligence is not one thing, then education should not be organized around developing or selecting for one thing. This chapter examines how the single-axis model shaped twentieth-century schooling, proposes what education looks like under a plural framework, and is honest about the trade-offs involved.

How IQ-centered thinking shaped schooling

The influence of the psychometric tradition on education has been pervasive, even where it is not explicitly acknowledged. Several features of twentieth-century schooling reflect the assumption that intelligence is a single, measurable, largely fixed capacity.

Tracking by "ability." School systems across the industrialized world sort students into different educational pathways based on assessed ability, typically using tests that are heavily loaded on abstract reasoning and verbal skill. Students identified as "high ability" are placed in accelerated tracks; students identified as "low ability" are placed in remedial or vocational tracks. The PISA findings documented in Chapter 22 showed that early tracking into vocational or academic pathways is one of the educational practices most strongly associated with educational inequality: it locks in advantage and disadvantage at a point when students' capacities are still developing. The assumption behind tracking is that ability is stable enough to justify sorting children at age 10 or 12 into pathways that will shape the rest of their lives. Dynamic assessment research (Chapter 21) challenges this assumption: what children can learn with proper support often exceeds what they can demonstrate on a static test of current performance.

Gifted programs selecting for abstract reasoning. Programs for gifted and talented students typically select on the basis of IQ or IQ-adjacent tests: Raven's Progressive Matrices, the WISC, or similar instruments that emphasize abstract reasoning, processing speed, and working memory. A child with exceptional practical judgment, extraordinary social-relational competence, remarkable creative generativity, or profound emotional attunement is unlikely to be identified as "gifted" unless they also score high on abstract reasoning. The dimensional framework reveals that gifted programs, as typically designed, select for one or two dimensions while systematically ignoring ten others.

Curricula prioritizing Gf over Gc, judgment, character, and relational competence. Modern curricula, particularly in mathematics and science education, increasingly emphasize "higher-order thinking" and "problem-solving," which in practice means fluid reasoning applied to abstract problems. This is not because practical judgment, emotional attunement, self-regulation, or collaborative

competence are unimportant. It is because abstract reasoning is easier to test at scale, easier to grade objectively, and easier to compare across schools and districts. The curriculum follows the assessment, and the assessment follows the psychometric tradition's emphasis on Gf.

The result: educational systems that identify, develop, and reward a specific profile of excellence (high cognitive power, high processing speed, moderate retained knowledge) while neglecting the dimensions that most traditions placed at the center of education's purpose (practical judgment, moral character, social-relational competence, self-regulation, and wisdom). Children who are strong on the neglected dimensions and moderate on the measured dimensions are classified as "average" or "below average," while children who are strong on abstract reasoning but weak on judgment, character, and relational skill are classified as "gifted." The single-axis model does not merely fail to recognize the full range of excellence. It systematically misallocates educational resources and opportunity.

Education under a plural framework

A plural framework would not abandon the teaching and assessment of cognitive skills. Reasoning, knowledge, and problem-solving are genuine dimensions of excellence that deserve rigorous development. What changes is the scope of what education attempts to cultivate and the methods it uses to assess progress.

Dimensional profiles for placement, not single scores. Instead of sorting students into tracks on the basis of a single ability score, a plural framework would generate profiles that describe each student's shape across multiple dimensions. A student with high cognitive power, moderate self-regulation, and low collaborative competence needs a different educational approach from a student with moderate cognitive power, high creative generativity, and exceptional social-relational competence. Profiles do not eliminate the need for differentiated instruction; they make differentiation more precise and more fair, because they prevent the collapse of a complex cognitive landscape into a single ranking.

Explicit cultivation of judgment, character, and relational skill. The traditions documented in Movement I provide existence proofs that practical judgment, moral character, self-regulation, and relational competence can be systematically cultivated, not merely selected for.

The Confucian educational tradition (Chapter 6) treated moral development as the central purpose of education. Students studied canonical texts not merely to acquire information but to develop the judgment needed to govern wisely and the character needed to govern justly. The textual curriculum was the vehicle for character formation, not a substitute for it.

Buddhist contemplative training (Chapter 8) developed systematic methods for cultivating self-regulation (*sati*, mindfulness; *samādhi*, concentration) that modern research has confirmed produce measurable improvements in attentional control, emotional regulation, and stress resilience. Mindfulness-based programs adapted for school settings have shown positive effects on students' self-regulation and wellbeing, though the evidence base is still developing.

The apprenticeship traditions (Chapter 32) cultivated practical judgment and embodied skill through extended supervised practice: years of working alongside an expert, receiving feedback on performance, and gradually taking on more complex and consequential tasks. Modern professional education in medicine, law, and engineering retains elements of this model (clinical rotations, clerkships, internships), but primary and secondary education has largely abandoned it in favor of classroom-based instruction and standardized testing.

These are not nostalgic appeals to a pre-modern golden age. They are evidence that the dimensions this book identifies as undervalued by the psychometric tradition can be, and have been, systematically developed through intentional educational practice. The question is not whether judgment, character, and relational competence can be taught. The question is whether modern educational systems are willing to invest the time, resources, and institutional commitment required.

Dynamic assessment as a complement to static testing. Chapter 21 established that dynamic assessment (test-teach-retest) is more accurate than static IQ tests for reflecting the learning potential of students whose prior educational experiences have been limited. This finding has direct implications for educational equity. Students from disadvantaged backgrounds often score lower on static tests not because they lack the capacity to learn but because they have had fewer opportunities to develop the skills the test measures. Dynamic assessment distinguishes between lack of opportunity and lack of ability by measuring how much the student learns when given appropriate instruction. A plural framework would use dynamic assessment alongside static testing, especially at educational transition points (entry to school, placement decisions, identification for enrichment programs) where the stakes of misclassification are highest.

Collaborative competence as a curricular goal. Chapter 31 documented the c-factor research showing that group performance depends on social sensitivity, conversational equality, and collaborative process, not merely on the aggregate intelligence of group members. If collaborative competence is a genuine dimension of excellence (and both the research and the traditions suggest it is), then schools should explicitly develop it: through structured group work with assessed outcomes, through peer teaching and mentoring, and through projects that require coordination, negotiation, and shared responsibility. This is different from simply putting students in groups and hoping they learn to collaborate. It requires intentional instruction in collaborative skills, explicit assessment of

collaborative competence, and educational structures that value group achievement alongside individual achievement.

Honest trade-offs

A plural framework for education would be better in important ways. It would be more accurate (recognizing the full range of human excellence rather than a narrow slice), more fair (reducing the systematic disadvantage imposed by measuring only the dimensions that one civilization prioritized), and more aligned with the traditions' understanding of what education is for (forming excellent persons, not merely maximizing test scores).

But it would also be harder, more expensive, and less scalable than the current system.

Plural assessment is more labor-intensive than single-score ranking. Generating a twelve-dimensional profile requires multiple forms of assessment: performance tests, dynamic assessment, demonstrated performance, expert observation, and self/peer ratings. Each of these requires trained assessors, time, and resources that single-score standardized testing does not. Schools operating under severe resource constraints may be unable to implement plural assessment, and mandating it without providing the resources to do it well would produce worse outcomes than the current system.

Some dimensions may not be assessable within the institutional structure of mass schooling. Embodied-ecological skill (Chapter 30) requires physical environments and extended practice that most schools cannot provide. Wisdom (Chapter 28) develops over decades and may not be meaningfully assessable in children and adolescents. Collective capability (Chapter 31) requires group assessment methods that are still in early development. A plural framework must be honest about which dimensions it can realistically assess in school settings and which require different institutional contexts (apprenticeships, community-based learning, professional training) for their development and assessment.

Skills require maintenance. The PIAAC findings (Chapter 22) showed that adult literacy and numeracy scores declined in the United States between 2017 and 2023, suggesting that knowledge and skills atrophy without continued use. A plural framework that cultivates twelve dimensions must reckon with the resource costs of maintaining all twelve, or must acknowledge that some dimensions will be developed in school and others will be developed later, in different institutional settings.

The political economy of assessment matters. Single-score ranking is politically convenient because it produces a clear hierarchy: this school is better than that one, this student ranks above that one. Profiles are harder to rank. A system that reports "this student has high cognitive power, moderate self-regulation, and low collaborative competence" does not produce a single number that can

be compared across students, schools, or districts. Politicians, administrators, and parents who want simple answers will resist the complexity that honest assessment requires.

Summary

Education after intelligence would not abandon the teaching of cognitive skills. It would situate cognitive development within a broader framework that also cultivates practical judgment, emotional attunement, self-regulation, moral character, creative generativity, collaborative competence, and (where possible) embodied skill and wisdom. The traditions provide existence proofs that these dimensions can be systematically developed. Dynamic assessment provides a fairer complement to static testing. And the dimensional profile provides more useful information than a single score for every educational purpose: placement, instruction, enrichment, and self-understanding.

The trade-offs are real. Plural assessment is harder, more expensive, and less scalable. Some dimensions resist school-based assessment. The political convenience of single-score ranking will create resistance to change. But the current system's costs are also real: it systematically misidentifies excellence, misallocates opportunity, and narrows the conception of human potential to a fraction of what humanity has actually valued. A framework that is more honest about the full range of human excellence, even if harder to implement, is worth pursuing.

Chapter 39: Institutions After Intelligence

The single-axis model of intelligence has shaped not only education (Chapter 38) but also how organizations hire, promote, and select leaders. This chapter examines how IQ-adjacent thinking currently dominates institutional selection, presents existing practices that already move in the direction of a plural framework, and argues for dimensional profiles in organizational contexts.

How IQ-adjacent thinking shapes institutional selection

The most influential finding in personnel selection research is the Schmidt and Hunter (1998) meta-analysis, which established that cognitive ability tests are the single best predictor of job performance across occupations, with a criterion-related validity of approximately $r = .51$. This finding has been enormously influential. It provides the scientific basis for organizations that use cognitive ability tests as the primary or sole selection instrument, and it is cited in virtually every industrial-organizational psychology textbook.

But the finding is more limited than it appears.

First, $r = .51$ means that cognitive ability accounts for approximately 26% of the variance in job performance. This is substantial, but it means that 74% of the variance in job performance is explained by other factors. Whatever those other factors are (practical judgment, social-relational competence, self-regulation, moral reliability, domain-specific knowledge, motivation, collaborative capability), they collectively matter nearly three times as much as cognitive ability alone. The single-axis model's dominance in hiring practice is not proportional to its explanatory power.

Second, cognitive ability tests produce the largest adverse impact of any major selection method. The standardized mean difference between Black and White test-takers on cognitive ability tests is approximately $d = 1.0$, meaning that the average Black test-taker scores roughly one standard deviation below the average White test-taker. This gap has been extensively studied, is partly attributable to socioeconomic and educational inequalities, and has significant legal and ethical implications. Organizations that rely primarily on cognitive ability tests for selection systematically exclude a disproportionate share of minority candidates, regardless of whether those candidates possess the practical judgment, social competence, moral reliability, and collaborative capability that the role actually requires.

Third, cognitive ability tests measure what the psychometric tradition chose to measure (abstract reasoning, processing speed, vocabulary) and do not measure the dimensions that the traditions in this book identified as central to leadership, governance, and institutional effectiveness. A person

who scores at the 95th percentile on a cognitive ability test may be an exceptional abstract reasoner who lacks the practical judgment to navigate organizational politics, the social-relational competence to build and maintain a team, the moral reliability to be trusted with consequential decisions, or the wisdom to know when not to act. The psychometric tradition's tools cannot distinguish this person from one who has the same cognitive ability plus all of these additional capacities. Both get the same score.

Partial correctives already in practice

Several existing selection methods already move in the direction of a plural framework, though they are typically used as supplements to cognitive ability testing rather than as replacements.

Situational Judgment Tests (Chapter 23) present realistic workplace scenarios and ask candidates to evaluate courses of action. SJTs have a criterion-related validity of $r = .26$ for predicting job performance, which is lower than cognitive ability tests but with two important advantages. First, SJTs show incremental validity above cognitive ability and personality: they predict aspects of job performance that neither IQ nor Big Five scores capture. Second, SJTs produce substantially lower adverse impact ($d = 0.38$ vs. $d = 1.0$ for cognitive ability tests), meaning that a selection system that includes SJTs alongside cognitive tests will produce a more diverse candidate pool without sacrificing predictive validity.

The finding that SJTs both measure a different dimension (practical judgment) and produce less disparate outcomes is not coincidental. It reflects the fact that practical judgment, unlike abstract reasoning, draws on a broader range of cognitive, social, and experiential capacities that are more evenly distributed across demographic groups than performance on timed tests of abstract pattern completion. A measurement approach that captures more of what the traditions valued also produces fairer outcomes. This pattern has appeared repeatedly throughout the book and deserves emphasis: broadening the conception of excellence and broadening the inclusion of populations are not competing goals. They are aspects of the same correction.

Assessment centers use multiple methods (simulations, group exercises, in-basket tasks, structured interviews, role plays, and presentations) to build a multidimensional profile of each candidate. A well-designed assessment center evaluates leadership potential, interpersonal effectiveness, decision-making under pressure, communication skill, and strategic thinking, producing scores on multiple dimensions rather than a single composite. Assessment centers have good criterion-related validity for predicting managerial performance, and their multidimensional structure aligns naturally with the plural framework proposed in this book.

The limitation of assessment centers is practical: they are expensive, time-consuming, and require trained assessors. They are used primarily for senior-level selection (executive recruitment, military officer selection, public-sector leadership programs) and are rarely scalable to high-volume hiring. But their existence demonstrates that the conceptual and methodological resources for dimensional assessment already exist in organizational psychology. What is lacking is not the know-how but the institutional willingness to apply it more broadly.

Team composition research (Chapter 31) suggests that who is on the team matters as much as or more than how individually able the team members are. The Woolley/Malone c-factor research found that group performance was predicted by the average social sensitivity of group members and the equality of conversational turn-taking, not by the average or maximum IQ of group members. Even if the replication controversy means this finding should be treated with caution, the practical implication is significant: an organization that selects exclusively for individual cognitive ability may be assembling teams that are individually brilliant but collectively dysfunctional.

The most effective teams, according to this research, are not composed of the people with the highest IQ scores. They are composed of people who listen to each other, who take turns contributing, and who are sensitive to each other's cognitive and emotional states. These are dimensions of social-relational competence and collaborative capability that cognitive ability tests do not measure and that the single-axis model systematically ignores.

Dimensional profiles for organizational selection

The plural framework implies that different roles require different dimensional profiles, and that selection should be tailored to the profile the role demands.

A **research scientist** needs high cognitive power, strong retained knowledge in their domain, creative generativity, and self-regulation (the discipline to pursue long-term research programs through setbacks and dead ends). Practical judgment, social-relational competence, and collaborative capability are less critical for individual research but become important as the scientist takes on mentoring, leadership, or team-based projects.

A **manager** needs practical judgment (making good decisions in ambiguous situations), social-relational competence (understanding and motivating team members), self-regulation (maintaining composure under pressure), and collaborative capability (building and sustaining effective teams). Cognitive power is necessary but not sufficient; a manager who is brilliant but cannot read people, manage conflict, or delegate effectively will fail in ways that their intelligence cannot compensate for.

A **senior leader** needs all of the above plus wisdom (long-term perspective, tolerance of uncertainty, capacity to balance competing interests), moral reliability (being trustworthy with consequential decisions and institutional resources), and the capacity to maintain collective capability across an entire organization. The current practice of promoting the most technically able individual contributor into leadership, and then being surprised when they struggle, is a predictable consequence of single-axis selection. The dimensional framework predicts exactly this failure: a person selected for cognitive power and technical expertise will not necessarily possess the practical judgment, social competence, and wisdom that leadership requires, because these are partly separable dimensions, not automatic consequences of high IQ.

What changes

Under a plural framework, organizational selection would change in three ways.

First, **job analysis would specify dimensional profiles**, not just "ability requirements." Instead of "requires high cognitive ability," a job specification would state which dimensions the role demands at what levels: "requires high practical judgment, moderate cognitive power, high social-relational competence, and high self-regulation." This is already done in some assessment center designs; the plural framework would make it standard practice.

Second, **selection batteries would be designed to match the required profile**. A role requiring practical judgment would include SJTs. A role requiring collaborative capability would include group exercises. A role requiring self-regulation would include assessments of executive function in realistic contexts. Cognitive ability tests would remain part of the battery for roles that require cognitive power, but they would not be the sole or dominant instrument for roles where other dimensions matter more.

Third, **team composition would be explicitly managed**. Instead of assembling teams by selecting the highest-scoring individuals on cognitive ability and hoping they work well together, organizations would consider the dimensional profiles of team members and the complementary strengths needed for the team's tasks. The c-factor research suggests that teams with diverse cognitive styles, high social sensitivity, and equitable participation patterns outperform teams of individually brilliant but socially insensitive members. The dimensional framework provides the vocabulary for making these composition decisions explicit rather than leaving them to intuition.

Summary

The current institutional selection model is dominated by cognitive ability testing because cognitive ability tests are the single best predictor of job performance. But "best single predictor" is not the same as "sufficient predictor." Cognitive ability accounts for approximately 26% of the variance in job performance; 74% is explained by other factors. The dimensions that account for the missing variance (practical judgment, social competence, self-regulation, moral reliability, collaborative capability) are the dimensions that the traditions in this book identified as central to institutional effectiveness and that the psychometric tradition's tools do not measure.

SJTs, assessment centers, and team composition research already move in the right direction. What is needed is not a new methodology but the institutional willingness to apply existing methods more broadly, to design selection for the dimensional profile the role requires rather than defaulting to the dimension that is easiest to measure.

Chapter 40: Alignment After Intelligence

AI alignment is currently framed around making systems "intelligent" and "safe," as if these were two separate properties that could be developed and assessed independently. But if intelligence is not one thing (Chapter 33), then the first term in this framing is incoherent, and if the dimensions of excellence interact with each other (Chapter 28 on wisdom as integration, Chapter 29 on moral reliability as inseparable from judgment), then treating safety as a separate add-on to capability is a structural error. What we actually want from AI systems is a specific profile of capabilities and dispositions, not maximum intelligence in some undifferentiated sense.

Not "how intelligent?" but "what profile?"

The dimensional framework proposed in Chapter 34 applies to AI systems as directly as it applies to human individuals or groups. The question for alignment is not "how intelligent should the AI be?" but "what dimensional profile should the AI have, given its intended purpose, deployment context, and the stakes involved?"

This reframing changes the conversation fundamentally.

Some dimensions should be maximized within scope. An AI system designed for scientific research should have high retained knowledge and strong reasoning within its domain. A medical diagnostic AI should have extensive clinical knowledge, pattern recognition capability, and the ability to integrate multiple information sources. A legal research tool should have comprehensive knowledge of case law and the capacity to identify relevant precedents. For these dimensions, more is generally better, subject to accuracy constraints.

Some dimensions should be constrained. Creative generativity in high-stakes contexts (an AI system that generates novel and surprising strategies for military operations, financial trading, or infrastructure management) may need to be constrained rather than maximized, because the novelty that makes creativity valuable in art and science makes it dangerous in contexts where unpredictability carries catastrophic risk. Autonomous practical judgment without human oversight is another dimension that should be constrained for the foreseeable future: a system that makes consequential decisions about people's lives (medical treatment, criminal sentencing, resource allocation) without human review is exercising a form of judgment that it does not possess in the sense the traditions described (Chapter 23). The system may produce statistically reasonable outputs, but it does not perceive the morally relevant features of particular situations, weigh competing considerations in light of experience, or accept responsibility for the consequences.

Some dimensions are not applicable in the same way. Embodied-ecological skill is not relevant for a text-based AI system in the way it is for a Pacific navigator or a surgeon. Moral reliability as a character trait (the disposition to act on what one knows to be right, developed through decades of moral formation) is fundamentally different from safety compliance as a design specification (the tendency to follow training objectives and refuse harmful requests). Both are important. But conflating them, treating an AI system's adherence to safety guidelines as equivalent to a person's moral character, commits the category mistake that Chapter 37 warned against.

Collective capability in the Ubuntu sense (participating in a community of persons through mutual recognition, shared responsibility, and relational attunement) is not something current AI systems possess or can meaningfully approximate. An AI system can be a useful component of a human-AI team (Chapter 31), but it does not participate in the team the way a person does: it does not take conversational turns out of social sensitivity, contribute to group cohesion through relational engagement, or sustain collective cognitive processes through mutual responsiveness. Designing AI systems to function well as components of human-AI teams is important and achievable. Claiming that AI systems are members of teams in the same sense that humans are is a category mistake.

The alignment problem is a judgment problem

Chapter 23 established that practical judgment is the dimension where the gap between what the traditions valued and what psychometrics measures is largest. This chapter argues that the same gap defines the central difficulty of AI alignment.

The alignment problem, at its core, is the problem of specifying what "acting well" means in situations that cannot be fully anticipated by rules. This is exactly the difficulty that Aristotle identified with *phronēsis* (practical wisdom cannot be captured by rules because applying rules requires judgment) and that Kant identified with the faculty of judgment (judgment is the capacity to subsume particular cases under general rules, and this capacity cannot itself be taught by rules, because any rule for applying rules would itself require judgment to apply).

AI systems that follow rules perfectly but lack judgment will fail in exactly the situations where judgment matters most: novel situations that the rules did not anticipate, situations where rules conflict with each other, situations where the right action requires perceiving morally relevant features that no rule specifies. The history of AI safety incidents already illustrates this pattern: systems that optimize for specified objectives while producing catastrophic side effects that the specifications did not address. These are failures of judgment, not failures of rule-following.

The alignment field has developed increasingly sophisticated approaches to this problem: reinforce-

ment learning from human feedback (RLHF), constitutional AI, debate-based alignment, and scalable oversight. These are genuine advances. But they all face the same structural limitation that the traditions identified: judgment cannot be fully specified in advance, and any system that works only within the scope of its specifications will fail outside that scope in ways that cannot be predicted from within the specification framework.

The dimensional framework does not solve this problem. But it clarifies it. What alignment researchers are trying to build into AI systems is something that the traditions tracked under the names *phronēsis*, *nyansa*, practical wisdom, and cultivated judgment: the capacity to act well in situations that exceed formal specification. The traditions' unanimous finding was that this capacity develops through experience, requires emotional and moral attunement alongside cognitive ability, and cannot be reduced to rule-following. If they were right, then the alignment problem is not a technical challenge waiting to be solved by better engineering. It is a permanent condition of any system (human or artificial) that must act well in a world more complex than any specification can capture.

The risk of wrong-dimension optimization

The most dangerous implication of the single-axis model for AI development is what might be called **wrong-dimension optimization**: building AI systems that maximize cognitive power and retained knowledge while possessing no practical judgment, no moral reliability, and no wisdom.

This is the psychometric tradition's mistake recapitulated at civilizational scale.

The psychometric tradition defined intelligence as abstract reasoning and processing speed, excluded practical judgment and moral character, measured the included dimensions with precision, and then used the resulting scores to allocate educational and economic opportunity. The result was a system that identifies and rewards a narrow profile of excellence while systematically neglecting the dimensions that most human traditions considered central to living well.

If AI development follows the same pattern, defining "intelligence" as benchmark performance on reasoning and knowledge tasks, excluding practical judgment and moral reliability, measuring the included dimensions with obsessive precision (MMLU scores to two decimal places), and then deploying the resulting systems in contexts that require judgment, character, and wisdom, the consequences will be proportionally larger. An educational system that narrows the conception of human excellence affects millions of students over decades. An AI system that narrows the conception of intelligence and then operates at global scale, making consequential decisions about medical treatment, legal outcomes, financial allocation, and information distribution, affects billions of people immediately.

The dimensional framework provides a structural defense against this risk. By requiring separate assessment on each dimension, it makes it impossible to claim that a system is "generally intelligent" on the basis of high performance on two dimensions while being absent on ten. It forces the question: does this system possess the specific capabilities and dispositions required for this specific deployment context? A diagnostic AI deployed in a hospital needs high domain knowledge, strong pattern recognition, and calibrated uncertainty estimation. It does not need creative generativity, and it should not exercise autonomous practical judgment. A conversational AI deployed for mental health support needs high emotional attunement (or at least a convincing approximation of it), strong social-relational competence, and careful self-regulation. It does not need mathematical reasoning ability, and it should not make clinical decisions without human oversight.

Dimensional profiles for AI evaluation

This book proposes that AI systems should be evaluated using dimensional profiles rather than single benchmarks. A responsible evaluation would report:

For each applicable dimension, a score or qualitative assessment (some dimensions may require expert evaluation rather than automated scoring). For each inapplicable dimension, an explicit statement that the dimension is not assessed and is not claimed to be present. For each deployment context, a specification of which dimensions are required at what levels and which should be constrained.

This is more complex than "it scored 92% on MMLU." It is also more honest. An LLM that scores at the 95th percentile on knowledge retrieval and the 90th percentile on certain reasoning tasks, but that cannot exercise practical judgment, does not possess moral reliability, has no emotional attunement, and cannot participate in collective intelligence, should be described exactly this way: as a system with extraordinary narrow capability and significant dimensional absences. The profile prevents the category mistake. It makes the system's actual capabilities visible and its limitations explicit.

The alignment field has already begun moving in this direction. Safety evaluations now assess models on multiple dimensions: toxicity, bias, factual accuracy, instruction-following, refusal of harmful requests, and susceptibility to adversarial manipulation. These are dimensional assessments. But they are typically treated as separate safety checks rather than as part of an integrated capability-and-disposition profile. The framework this book proposes would unify capability assessment and safety assessment into a single dimensional profile, making it clear that safety is not an add-on to intelligence but a dimension (or cluster of dimensions: practical judgment, moral reliability, self-regulation) that is assessed alongside and on equal footing with cognitive capability.

Summary

AI alignment needs a richer target than "intelligent and safe." What we want from AI systems is a specific profile of capabilities and dispositions, tailored to the deployment context, transparent about what is present and what is absent, and designed to prevent the category mistake of treating narrow capability as general excellence.

The alignment problem is a judgment problem, and judgment, as the traditions and Kant agreed, exceeds any formal specification. Wrong-dimension optimization, building systems that maximize cognitive power without practical judgment, moral reliability, or wisdom, recapitulates the psychometric tradition's narrowing at a scale that makes the consequences civilizational. The dimensional framework does not solve alignment. But it provides the vocabulary for being honest about what AI systems can and cannot do, and honesty about capabilities is the prerequisite for deploying them safely.

Chapter 41: A New Language for Human Excellence

This book began with a word and ends with a proposal to retire it. Not because the phenomena "intelligence" points at are unreal (reasoning is real, memory is real, judgment is real, and they matter), but because the word compresses too much, obscures too much, and has accumulated too much ideological weight to serve any longer as the master category for human excellence.

The full arc

Movement I showed that humanity has never had one concept here. Across nineteen chapters and more than a dozen civilizational traditions, the survey established that what different cultures tracked, cultivated, and rewarded overlaps substantially but bundles differently. The Greek tradition distinguished five intellectual virtues and placed *phronēsis* (practical wisdom) at their integration. The Confucian tradition built the world's largest meritocratic examination system around textual mastery and moral cultivation, treating these as inseparable. The Indian Nyāya school developed a formal logic of comparable sophistication to Aristotle's, using a fundamentally different architecture. The Buddhist tradition insisted that wisdom without compassion is not yet wisdom. The Yoruba tradition distinguished intellectual wisdom (*ogbón-ori*) from emotional wisdom (*ogbón-inú*) and made character (*iwà*) the paramount quality. The Ubuntu tradition challenged whether the individual extracted from relational context is a meaningful unit of analysis at all. The Māori and Pacific traditions preserved and transmitted vast bodies of ecological, navigational, and genealogical knowledge through memory systems of extraordinary sophistication and precision.

Five civilizations independently developed ideals of the excellent person that integrate cognitive and moral excellence: the Aristotelian *phronimos*, the Confucian *junzi*, the Islamic *adīb*, the Yoruba *omolúàbí*, and the Nahua *tlamatini*. Three civilizations independently developed formal logical systems using different architectures. Eight or more traditions refused to separate intellectual excellence from moral character. The convergence confirms that the phenomena are real. The different bundling confirms that the category "intelligence," as a single unified kind, is constructed rather than discovered.

The Western narrowing (Chapters 15-18) traced how a rich, plural field of capacities was progressively compressed: Bacon's emphasis on method over judgment, Descartes's elevation of mathematical reasoning, Locke's reduction of the mind to a processor of experience, and then the nineteenth-century collapse into measurable, testable, heritable "intelligence" that Galton initiated and Spearman formalized. Binet designed the first intelligence test to identify children who needed educational support, defining intelligence as "judgment, good sense, practical sense, initiative, the faculty

of adapting oneself to circumstances.” The tradition that followed measured abstract reasoning and processing speed instead. The dimension Binet named as central was too difficult to operationalize, so the tradition measured what it could measure and then defined intelligence as what it measured.

Movement II mapped twelve dimensions of human excellence across traditions and modern measurement, and the pattern that emerged was devastating for the single-axis model. The dimensions the psychometric tradition measures best (reasoning, knowledge) are the ones most traditions treated as one component among many. The dimensions most traditions placed at the center (practical judgment, wisdom, moral character, social and relational competence) are the ones the psychometric tradition measures least well or not at all. One dimension (embodied-ecological skill) has no standardized measurement whatsoever. The measurement landscape revealed three-model confusions (emotional intelligence), ecological validity gaps (executive function tests predicting only 18-20% of real-world self-regulation), psychometric collapses of widely used instruments (the RMET), and wisdom measures that disagree on whether wisdom increases with age.

Movement III drew the conclusion and proposed the replacement. Chapter 33 argued that intelligence is not a single natural kind, using evidence from within psychometrics (the mutualism model’s alternative explanation of the positive manifold), from across civilizations (convergent phenomena, divergent bundling), and from the measurement fragmentation documented in Movement II. Chapter 34 proposed a twelve-dimensional framework. Chapter 35 argued for profiles over single scores. And Chapters 36-40 applied the framework to the domains where it matters most: civilizationally plural assessment, human-AI comparison, education, institutions, and AI alignment.

The three-move distinctive contribution

Chapter 19 identified what distinguishes this book from its predecessors. The argument bears restating here, at the end.

First, the cross-civilizational survey. Sternberg’s triarchic theory, Gardner’s multiple intelligences, Peterson and Seligman’s character strengths, and Baltes’s Berlin wisdom paradigm all challenged the single-axis model, and this book is deeply indebted to each. But none conducted a systematic, chapter-by-chapter survey of how distinct civilizational traditions (Greek, Roman, Chinese, Indian, Buddhist, Islamic, Akan, Yoruba, Ubuntu, Māori, Inuit, Pacific, Nahua, Japanese) conceptualized, cultivated, and assessed human excellence. The survey is this book’s empirical foundation, and the findings (five integrative ideals, three independent logical traditions, the thick-formation convergence on character, the Gf/Gc value inversion, the five recognition-method families) could not have been generated without it.

Second, honest psychometric engagement. This book is not an anti-IQ polemic. It concedes what the psychometric tradition gets right: cognitive abilities are real and measurable; they vary across individuals; they predict important outcomes; the positive manifold is a genuine empirical regularity; and standardized testing provides comparability, a specific kind of fairness, and scale that no other assessment method can match. The book's argument is not that IQ is wrong but that it is incomplete, and that the incompleteness is not a minor gap waiting to be filled but a structural narrowing that excludes the dimensions most human traditions placed at the center of their accounts of excellence.

Third, the proposed replacement framework. Many books critique the concept of intelligence. Few propose a specific, empirically grounded alternative. The twelve-dimensional framework is not the final word (dimensions may need to be added, split, or merged as evidence accumulates), but it is a concrete proposal that can be tested, refined, and applied. It is more useful than critique alone because it provides a vocabulary for the practical domains (education, hiring, AI evaluation, alignment) where the single-axis model currently dominates.

What is lost

Honesty requires acknowledging what is lost when you give up a single simple concept.

Parsimony. One number is easier to use than twelve. A Full Scale IQ can be computed, communicated, and compared in seconds. A twelve-dimensional profile requires more computation, more communication, and more cognitive effort from the user. For some purposes (quick screening, rough comparison, resource allocation under severe time pressure), the single number will remain useful even if it is known to be incomplete. The plural framework does not eliminate the need for summary statistics; it insists that summary statistics should not be mistaken for complete descriptions.

Communication ease. "She's intelligent" is quicker and more intuitive than "she has high cognitive power, moderate practical judgment, and exceptional social-relational competence." Language compresses for a reason: life is too complex to describe in full at every moment. The concept of intelligence, even if it is a false unity, serves a communicative function. Retiring it from technical and institutional use does not mean banning it from everyday conversation. It means being precise when precision matters and acknowledging the compression when it does not.

The political convenience of a single ranking. A single axis of intelligence produces a single ranking: this person is smarter than that one, this school is better than that one, this nation outperforms that one. Rankings are politically powerful. They justify resource allocation, legitimate social hierarchies, and provide a simple narrative for complex realities. Giving up the single ranking means giving up this convenience. It means acknowledging that a person who ranks low on one dimension

may rank high on another, that a school that produces strong reasoners but poor collaborators is not simply "better" than one that produces the reverse, and that national rankings on PISA scores tell you about reading and mathematics performance but not about the practical judgment, moral character, or collaborative competence of a nation's citizens. The political convenience of single rankings is also their greatest danger: they compress a multidimensional reality into a one-dimensional hierarchy and then treat the hierarchy as natural rather than constructed.

What is gained

What the plural framework gains is worth the cost.

Honesty. The framework describes what is actually there rather than compressing it into a false unity. Human excellence is plural: it consists of cognitive, regulatory, relational, creative, moral, embodied, and collective capacities that develop along different trajectories, are valued differently across civilizations, and are measurable to different degrees. A framework that acknowledges this plurality is more accurate than one that denies it, and accuracy is the precondition for every other intellectual and practical virtue.

Cross-cultural fairness. The single-axis model systematically disadvantages populations whose strengths lie in dimensions that the psychometric tradition does not measure. If practical judgment, social-relational competence, embodied skill, and collaborative capability are genuine dimensions of excellence (and the evidence assembled in this book says they are), then a selection system that measures only abstract reasoning and vocabulary systematically excludes people who are excellent in ways the system cannot see. Plural assessment does not guarantee fairness, but it reduces the structural disadvantage imposed by measuring only one civilization's priorities.

Better science. Dimensions measured separately can be studied, developed, and related to outcomes more precisely than a single composite. The Gf/Gc distinction has been one of the most productive analytical tools in intelligence research precisely because it separates two things that develop differently, predict different outcomes, and respond to different interventions. Extending this analytical separation to additional dimensions (practical judgment, emotional attunement, self-regulation, moral character) would enable more precise research on how these capacities develop, what predicts them, and how they can be cultivated.

Better education. Children cultivated across multiple dimensions develop more fully than children selected on one. The Confucian, Buddhist, and apprenticeship traditions demonstrated that character, judgment, self-regulation, and relational competence can be systematically developed, not merely selected for. A plural educational framework that takes these dimensions seriously would produce

not just better test-takers but more capable, more complete, more adaptable human beings.

Better AI alignment. AI systems evaluated on dimensional profiles are deployed more safely than systems evaluated on single benchmarks. The dimensional framework makes it impossible to claim that a system is "generally intelligent" on the basis of high performance on reasoning and knowledge while being absent on judgment, character, and wisdom. It forces the question that the single-axis model evades: does this system possess the specific capabilities and dispositions required for this specific deployment context?

The compressed synthesis

This book has traveled a long distance: from Homer's *mētis* to Aristotle's five intellectual virtues, from the Confucian *junzi* to the Nyāya five-step syllogism, from Galton's anthropometric laboratory to the WAIS and Raven's Matrices, from the Yoruba *omolúàbí* to the Berlin Wisdom Paradigm, from the Ubuntu challenge to the unit of analysis to the Woolley/Malone c-factor, from the Mohist Canons to the mutualism model of the positive manifold.

The findings can be compressed into a single statement:

Humanity has never had one concept here. What we have is a plural field of capacities, cognitive, regulatory, relational, creative, moral, embodied, and collective, that different civilizations have tracked, cultivated, bundled, and validated in different ways. The task ahead is not to find the one true definition of intelligence but to build frameworks worthy of the full range of human excellence.

This book is one attempt to begin that work.

Appendices

Appendix A: The Master Comparative Matrix

The full scored matrix from the research phase: traditions × dimensions × recognition methods. Provided as both a table and a downloadable dataset.

Appendix B: Glossary of Key Terms Across Traditions

A reference glossary of every major non-English term used in the book, with tradition, original script, romanization, and gloss.

Appendix C: Current Measurement Instruments. A Quick Reference

A table of major current tests and scales discussed in the book, organized by dimension: what each measures, how, and main limitations.

Appendix D: Further Reading by Tradition

Annotated bibliography organized by cultural/intellectual tradition rather than by academic discipline.

Book Statistics (Estimated)

Section	Chapters	Estimated Words
Movement I: History	18 chapters (1 to 18)	~87,000
Movement II: Map	13 chapters (19 to 31)	~54,500
Movement III: Replacement	9 chapters (32 to 40)	~46,000
Appendices	4	~15,000
Total	40 chapters + 4 appendices	~202,500

This is a full-length scholarly-but-accessible book: roughly 200,000 words, comparable in scope to major works of intellectual history.

Working Title Options

- *Beyond Intelligence: A Human History of Mind, Wisdom, Judgment, and the Measurement of Excellence*
- *The Measure of Mind: How Humanity Defined Intelligence, Wisdom, and Human Excellence*
- *After Intelligence: Toward a Plural Theory of Human Excellence*
- *The Plural Mind: What Humanity Really Meant by Smart*
- *Twelve Dimensions: A Cross-Civilizational Account of Human Excellence*