# On Recognising Intelligence

Jeff Riley

Monash University, Melbourne, Australia

**Abstract**

We have been debating what intelligence is for two and a half thousand years, and we are no closer to a definition that generalises. Psychometrics, cognitive-developmental theory, pluralism, the computational and connectionist paradigms, and the embodied and enactive paradigms each capture something real, and each fails as a general account. I argue that this is not a failure of effort, but a sign that we have been looking for the wrong kind of thing.

Intelligence resists definition, but it does not resist recognition. What we do, in practice, is recognise it — and that recognition is graded, observer-relative, and dependent on task and context. Drawing on cases from across the biological continuum, and on the behaviour of current AI systems, I set out a description rather than a definition: a profile of six components, assessed as a shape rather than collapsed into a score. The framework I describe is deliberately medium-neutral: I treat the question of whether intelligence requires a biological substrate as open, rather than settled in either direction.

That intelligence is something we recognise rather than something we define points to a more pragmatic resolution. For most practical purposes — deciding what to trust a system with, and under what oversight — "Is it intelligent?" is the wrong question. The better question, more modest and answerable, is "Does the system do what we need it to do, well enough, in the setting where we need it?"

## 1 Intelligence

Intelligence is one of those concepts that everyone recognises, but almost no one can really define. We are very confident that *we* have it, reasonably confident that *some* animals have *some* of it (but less than us), increasingly uncertain whether machines have *any* of it, and largely at a loss when asked whether plants, fungi, or insect colonies have any of it at all. The word carries a heavy load across many fields — psychology, biology, computer science, and philosophy — without ever quite settling into a single, agreed-upon meaning.

The English word "intelligent" takes its roots from the Latin word *intelligens*, which means "to understand", "comprehend", or "perceive". The Chinese Pinyin phrase for "intelligent" is *cōngmíng*, which translates literally to "good hearing and clear sight" (the ability to perceive the world perfectly). Other languages have their own words and definitions — each language has its own etymology for the words "intelligent" and "intelligence", and while the underlying concepts wash over all languages, the dictionary definitions tend to beg the question.

Underlying the lack of definition lies a practice that is essentially ostensive. We point at humans and label them "intelligent"; we point at simple bimetal strip thermostats (for example) and label them "not intelligent"; and in the considerable middle ground we defer to intuition. U.S. Supreme Court Justice Potter Stewart famously said "*I know it when I see it*" when struggling to define "hard-core

pornography" [1] — and we treat intelligence in something like the same spirit. The implicit rule for a graded definition of intelligence in that middle ground is harder to state, and the harder we try to state it, the more it seems to elude us. What we are doing, in practice, is recognising intelligence — and typically, recognising the resemblance to ourselves — without ever quite having said what it is that we are recognising. The implicit rule for the bookends of the middle ground seems to be very close to "Us", and "Not us". That has worked well enough for ordinary purposes up until now, but it gets us into trouble whenever we try to extend the concept beyond its original referent — which is what the arrival of Artificial Intelligence (AI) now demands.

This would matter less if it were merely a philosophical puzzle, but it is not. The current generation of AI products is sold under a label — "intelligent" — that comes with an implicit promise never made clear, and that looseness is increasingly used as a rhetorical, and commercial, tool. A statistical text predictor is "intelligent". A spam filter is "intelligent". A system that recognises faces in photographs is "intelligent". A system that drives a car is "intelligent". Our lack of ability to define "intelligent" in anything but the most vague terms is not considered a shortcoming by the purveyors of AI systems and their marketing teams — it's a feature waiting to be exploited. The more vague the term, the more capability it can be made to imply, and the more impressive the product can be made to sound. Every "smart" device on the market — smart phones, smart watches, smart fridges — borrows credibility from a word the seller is under no obligation to make precise.

This is not a new observation. As I argued in Riley (2006a,b), much of what was being called artificial intelligence at the time was, on close examination, the result of a great deal of human expert knowledge encoded in advance into the system — the "real intelligence" sitting in the heads of the engineers who decided how to break the problem into tractable parts, rather than in the system itself. Fifteen years later, in a follow-up (Riley 2021a), I made much the same observation about the deep learning era — the strategies and algorithms by which machines learn are designed by humans, the problems are posed by humans, the training (and test) data is curated by humans, and the form of an acceptable solution is specified by humans. None of this is to say that current systems are not impressive — they are, of course — but it is to say that the word "intelligent", when applied to them, is a label that deserves more examination and reflection than it has had, or is currently getting.

Many fields of science study concepts whose precise nature remains contested — biology continues to study *life* despite no philosophically settled definition of it, physics studies natural phenomena that resist reduction, neuroscience studies consciousness with widely varying interpretations — a pattern Denning (2025) documents across several fields. Each advances by way of working descriptions and operational criteria rather than airtight definitions. Biology, for instance, identifies a candidate organism as living by checking it against a set of seven indicative criteria — response to stimuli, growth, reproduction, metabolism, homeostasis, cellular structure, adaptability — without needing to settle the deeper question of what life essentially is. The criteria do not constitute a definition, but they do provide a shared framework. Biologists judge an organism by how many of the criteria it meets — the more it meets, the more confidently they call it alive — and where a case is borderline (e.g. a virus), they can still disagree productively, because they are weighing the same criteria. That common ground, in both agreement and disagreement, is what lets them do their work and talk to one another about it. What we lack for "intelligence" is not necessarily a definition — we may never have one — but a similarly usable working description, clear enough to support productive discussion across cases, fields, and stakeholders. The remainder of this article is in part an attempt to provide such a description.

[1] https://supreme.justia.com/cases/federal/us/378/184/
https://en.wikipedia.org/wiki/Jacobellis_v._Ohio

I do not pretend here to solve a problem that has resisted resolution since at least Aristotle — my aim is more modest, and threefold. First, to survey how intelligence has been defined, and to identify where those definitions break down when confronted with the actual variety of intelligence-like phenomena we observe — in animals, plants, colonies, networks, and machines. Second, to ask whether the search for a single objective definition is itself misguided, and whether intelligence is in some real sense in the eye of the beholder. And third, to propose a description, rather than a definition, that gives us better grounds for evaluating the things we are now being asked to call intelligent. Such a description will have limits, and will leave some questions open, but I believe it will serve us pragmatically, and provide a starting point for the broader programme of a science of intelligence (Delic and Riley 2022).

The argument I develop is that intelligence is best understood as graded, observer-relative, and task- and context-dependent — properties that the prevailing definitions struggle to capture cleanly, and that the pragmatic description proposed here is designed to accommodate. The clear implication that results is that intelligence is something we *recognise* rather than something we *define*.

## 2 The Turing Trap

A natural starting point for a discussion of machine intelligence specifically is Alan Turing's 1950 paper "Computing Machinery and Intelligence" (Turing 1950) that proposed what is now called the Turing Test — a test which, over time, and mostly in popular literature, has been mistaken for a test of whether a machine is intelligent at all. It would be remiss of me not to offer my opinion that none of the resulting confusion about the test should be ascribed to Turing himself.

Turing was characteristically precise in his paper. He argued that the question "Can machines think?" was too ill-defined to be useful, and proposed instead a more tractable behavioural question — could a machine, in text-based conversation, be reliably mistaken for a human? He was explicit that he was substituting a tractable problem for an intractable one, and that the substitute was *not the same question* — but one that could be answered. He also predicted, albeit with some hedging, that by the end of the twentieth century the meaning of "thinking" would have shifted enough that the question would settle itself by linguistic drift rather than by scientific resolution.

The conflation and confusion came later, in popular, and even some academic, treatments that omitted or ignored Turing's caution and presented the imitation game as a *test of intelligence*, rather than as a useful proxy for a largely unanswerable question. By the time the test had passed into general circulation, "passing the Turing Test" was being treated as equivalent to "being intelligent" — an equivalence Turing did not propose as a *definition*, even as he expected that passing would, in time, give reasonable observers grounds to treat machines as thinking, and that the language would shift accordingly — a prediction about how we would come to speak, not a definition of intelligence in terms of conversational behaviour. Roughly half of Turing's paper is devoted to anticipating and addressing potential objections — theological, mathematical, from consciousness, from disability, from continuity, and others.

The earliest version of this critique came from Weizenbaum (1966), whose Eliza program — a script of fewer than 450 lines of code that simulated a Rogerian psychotherapist by pattern-matching and rephrasing user input — provoked alarm rather than satisfaction in its creator. Weizenbaum observed that users readily attributed understanding, emotional sensitivity, and even therapeutic insight to a program he knew did none of those things. His conclusion, developed more fully in later work (Weizenbaum 1976), was that the Turing Test reveals more about what humans are willing to believe than about whether a machine has anything resembling intelligence — a point that prefigured Searle

(1980)'s argument by more than a decade and that locates the conflation in the observer rather than in the test or the machine.

Searle (1980)'s Chinese Room argument presented the philosophical objection very clearly: a system can pass a behavioural test of conversation by sophisticated symbol manipulation, without anything that resembles understanding or, by implication, intelligence. Whether we should accept Searle's argument in full or not is debatable. Hofstadter (1980)'s reply in *Behavioral and Brain Sciences* dissents vigorously, on the grounds that Searle has trouble seeing how mind can emerge from the interaction of merely physical components, and, as I noted in Riley (2021b), I have some sympathy with that line of objection. But the basic point survives regardless — the Turing Test, as it stands today, conflates two separable properties: *cognitive capacity* (the substantive ability to reason, plan, learn, and generalise) and *human-likeness* (the behavioural pattern of a conversing human). The second is neither necessary nor sufficient for the first: a system can reason and generalise without doing so in a recognisably human way, and can mimic human conversation without reasoning at all.

The recent generation of Large Language Models (LLMs), particularly the Generative Pre-trained Transformer (GPT) family, has made the conflation embarrassingly clear. These AI tools now pass casual versions of the test routinely, and the response of most thoughtful observers has not been "machines are now intelligent" but rather "the test was never really about intelligence". The latter is the right response, and one I think we should have arrived at long before any such demonstration was needed. And in the time since LLMs first did this, capability has continued to climb — step-by-step reasoning prompts, tool use, and increasingly complex multi-stage workflows, and explicitly reasoning-trained models have moved AI well beyond conversational fluency. The temptation to read each new capability as a step closer to intelligence is as strong as ever, and it should be resisted for the same reason: however impressive the latest advance, it is evidence of what the system can do, not of whether it is intelligent. I will return to what these systems actually do, and do not do, in Section 9.

But there is a deeper question here, which is whether *any* test of intelligence — human-like or not — could command general agreement. Every known candidate test embeds a prior theory of what intelligence is. If intelligence is general problem-solving, then a battery of novel problems is the test. If it is skill-acquisition efficiency, then a benchmark like Chollet (2019)'s ARC is the test. If it is rich behavioural flexibility, then no finite test can suffice. The disagreements about *what to test* are not technical disagreements that better instruments can resolve — they are disagreements about what intelligence *is*. I will return to this when I ask whether the concept might be irreducibly observer-relative.

## 3 Existing Definitions

The major approaches to defining intelligence can be usefully grouped into six broad domains: psychometric, cognitive-developmental, pluralist, computational, connectionist, and embodied and enactive — though the boundaries are soft, and different views could reasonably yield more domains or fewer. Some of the apparent disagreement between them fades when we recognise that they answer different questions.

### 3.1 Psychometric

Spearman (1904) observed that success on different cognitive tasks was positively correlated, with individuals who performed well on one task tending to perform well on many others. He posited a general factor, $g$, that underlies and explains these correlations. The psychometric paradigm has built

a large and methodologically sophisticated body of work on this finding, culminating in modern IQ instruments. Wechsler (1944), whose tests still dominate clinical practice, defined intelligence as the global capacity to act purposefully, think rationally, and deal effectively with one's environment. A widely cited consensus statement signed by 52 researchers, originally published in the *Wall Street Journal* in 1994 and reproduced in Gottfredson (1997), described intelligence as a very general mental capability involving reasoning, planning, problem-solving, abstract thought, comprehending complex ideas, learning quickly, and learning from experience.

Psychometrics is useful in practice, but theoretically limited. $g$ predicts a range of life outcomes well enough to be socially consequential, and IQ tests are reliable in the technical sense (they reproduce themselves), even where their validity remains contested and controversial. The theoretic limitation is specific: psychometrics tells us what intelligent behaviour looks like *in humans*, and how to measure variance among humans, but it offers no principled basis for asking whether a *non-human* subject is intelligent. The tests are calibrated against human populations and they presuppose the kind of subject being tested. Psychometric tests are silent — necessarily — on questions like "is an octopus intelligent?", "a forest?", "a language model?": they are simply not designed to answer such questions.

The IQ paradigm is also a useful cautionary tale about what happens when we try to make intelligence measurable. The Flynn effect (Flynn 1984, 1987) — sustained, substantial gains in measured IQ across the twentieth century — suggests that the tests pick up something that varies with environment and schooling, rather than something stable about individuals. Cross-cultural application is fraught — instruments calibrated on Western, educated populations perform differently and predict different outcomes when administered elsewhere. The tests themselves are, in the end, batteries of specific tasks — verbal analogies, matrix-completion puzzles, working-memory exercises, etc. Performance on those tasks correlates with performance on others, and that is the empirical content of $g$, but performance on a battery of tests is still just performance on that battery of tests. IQ testing — the clearest historical attempt to reduce intelligence to a measurable quantity — ended up measuring task-specific competence, calibrated to a particular cultural setting. I contend that this outcome is exactly what we should have expected. I will return to this.

Psychometrics inherently begs its own question. Intelligence, in psychometrics, ends up being whatever the tests measure, and the tests measure whatever predicts the outcomes (school grades, job performance) that the construct of intelligence was originally invoked to explain. The objection doesn't mean psychometrics is not useful — many useful scientific constructs are partly defined by their measurement — but it does mean that the psychometric definition of intelligence is, fundamentally, an operational shorthand rather than a substantive theory.

That so many distinct "intelligences" have been proposed — academic, social, emotional, practical, and the rest — is, I suspect, not so much evidence that there *are* many distinct intelligences as it is evidence that the original psychometric construct was too narrowly drawn. Gardner and others were right to object that IQ testing captured only a slice — where I differ is in the conclusion drawn. Rather than trying to overcome the narrowness of the construct by invoking a proliferation of intelligences, I believe instead that there is a single intelligence that is expressed as different capacities when it is exercised in different circumstances and on different content. Cases of uneven profile — strong analytic ability alongside weak social skill, or the reverse — are real, but they show intelligence differently cultivated across domains, rather than multiple intelligences in one host.

### 3.2 Cognitive-developmental

A remark, often attributed to Jean Piaget, describes intelligence as what you use when you don't know what to do — a definition that points at adaptability and novel problem-solving rather than at stored knowledge. The somewhat glib one-liner is a simple off-hand counter to psychometrics, but Piaget's work had a much broader context.

Piaget (1950)'s larger framework, developed across roughly five decades, charts the stages through which children move from sensorimotor reflexes, through preoperational and concrete-operational thought, to formal-operational reasoning — the ability to entertain abstract hypotheticals, manipulate symbols, and reason about reasoning. His central concepts — *schemas* as cognitive structures, *assimilation* as the incorporation of new experience into existing schemas, *accommodation* as the modification of schemas to fit experience that doesn't, and *equilibration* as the dynamic balancing between the two — describe intelligence as something that develops through structured interaction with the world, rather than as a fixed quantity that can be measured.

A useful contemporary contrast comes from Lev Vygotsky (Cole et al. 1978), working in Soviet Russia at roughly the same time. Where Piaget located cognitive development primarily in the maturing individual interacting with the physical environment, Vygotsky emphasised the social and cultural scaffolding that he argued was constitutive of intelligence, not incidental to it. His *zone of proximal development* — the gap between what a learner can do alone and what they can do with appropriate help — locates intelligence partly in the social interaction itself.

As we might expect, both frameworks have been challenged. Piaget's stages turn out to be less universal and more domain-dependent than originally claimed — children acquire abilities at different rates in different domains, and adults frequently fail at tasks Piaget would have predicted to be trivial for them. Cross-cultural application has been criticised along similar lines. Vygotsky's framework, harder to operationalise, has been generative in education research but resistant to clean experimental tests. And neither model extends easily to non-human or machine intelligence — a Piagetian framework asks how cognition *develops*, a Vygotskian one asks how it is socially *constituted*, but neither offers any guidance on whether something that didn't develop in the relevant way (a slime mould, an octopus, a transformer) has intelligence at all.

What the cognitive-developmental framework contributes, I think, is not so much a definition of intelligence as a contrast, or counterpoint — any account that treats intelligence as a snapshot rather than a process risks missing what these models identify as a central element. Coping with novelty — the capacity to confront a situation not seen before, and to construct rather than merely retrieve a response — for example, is a property that develops and becomes evident over time, generally through the traces left across a sequence of novel encounters.

### 3.3 Pluralist

Where psychometrics aims to extract a single underlying factor from the mass of cognitive performance, pluralism rejects that approach entirely. The pluralist view is that what we call "intelligence" is not one thing but many — separate domains of competence that may correlate weakly, or not at all, and that any framework forcing them into a single dimension loses more than it captures.

Gardner (1983), in *Frames of Mind*, articulated the most influential version of this view. He argued for at least seven (later eight) distinct intelligences — linguistic, logical-mathematical, spatial, musical, bodily-kinaesthetic, interpersonal, intrapersonal, and naturalistic — each with its own developmental trajectory, its own neural medium, and its own characteristic forms of expression. To qualify as an

intelligence in Gardner's framework, a candidate domain had to satisfy a set of criteria, including potential isolation by brain damage, the existence of savants and prodigies, an identifiable core operation, a distinct developmental history, evolutionary plausibility, and amenability to encoding in a "symbol system" — a list (deliberately) narrower than "things humans do well" and broader than "what IQ tests measure".

Sternberg (1985)'s triarchic theory presents a different view. Sternberg distinguishes *analytical* intelligence (the kind IQ tests measure — reasoning and problem-solving on well-defined tasks), *creative* intelligence (the capacity to deal with novelty, to generate ideas that are both new and useful), and *practical* intelligence (the capacity to apply knowledge effectively in real-world contexts — what we might colloquially call "street smarts"). His later work emphasised the integration of all three in achieving the life goals we value, given our cultural context. Sternberg himself made the deflationary observation that has become almost a running joke in the field — that the number of distinct definitions of intelligence on offer in the literature approaches the number of experts willing to offer a definition. His observation, though probably made somewhat flippantly, is, I think, right on target, and is one undercurrent throughout this work.

Pluralism has spawned offshoots beyond Gardner and Sternberg. *Emotional intelligence* — articulated formally by Salovey and Mayer (1990) and popularised by Goleman (1995) — proposes that the perception, understanding, and regulation of emotion is a cognitive capacity distinct from analytical intelligence. *Social intelligence*, *practical intelligence*, and *cultural intelligence* all extend the pluralist paradigm to particular domains. But the empirical picture is mixed. Performance across what pluralists call distinct intelligences does correlate, often substantially, which is the empirical fact that motivates $g$ in the first place. The correlations, however, are far from unity, and there is genuine domain-specific variance that the strong-$g$ position struggles to accommodate. I think the adherents of pluralism and psychometrics are looking at the same data and emphasising different aspects of it — psychometrics the shared variance, pluralism the residual.

The problem with pluralism is that rather than elucidating the concept of "intelligence", it actually muddies the waters. It may well describe the shape of human cognitive variation better — most of us have ragged profiles across cognitive domains rather than uniformly high or low ability — but it makes the cross-species and cross-medium question harder. If intelligence is already plural within a single species, then asking whether some other animal or system has it is almost meaningless. We can ask whether a system has *some* intelligence, and the answer, because of the pluralist view, is almost certainly yes for most candidates, which in the end doesn't mean very much.

Marvin Minsky's *Society of Mind* (Minsky 1986) is closely related to pluralism, but his proposal — that intelligence emerges from the interaction of many small interacting components rather than from any single unified mechanism — is pluralism about cognitive architecture rather than about kinds of intelligence, and so is not addressed further here.

### 3.4 Computational

The most ambitious recent attempts to define intelligence come from AI, where the goal — as the computational paradigm frames it — is to characterise intelligence in terms that could apply to any system, biological, mechanical, or otherwise. The computational paradigm is deliberately formal, aiming for definitions that admit mathematical treatment and, in principle, measurement. To be clear, *computational* here refers to the formal, mathematical character of this paradigm — not all systems that compute fall under it. The connectionist systems of Section 3.5, for example, also compute. I discuss

the relationship between the two paradigms there.

Russell and Norvig (1995) frame intelligence as the property of a *rational agent* that selects actions to maximise expected utility given its information about the environment. The framework, with roots in the von Neumann–Morgenstern decision theory of the 1940s (von Neumann and Morgenstern 1944), treats intelligence as a kind of optimisation problem — an agent receives percepts, maintains a model, has a utility function, and chooses actions that, in expectation, do best by that function. The appeal of the framework is its precision, but that precision is paid for by smuggling in some quite substantial presuppositions about what an agent is.

Legg and Hutter (2007) surveyed more than seventy existing definitions of intelligence and distilled them into a single mathematical formulation — *intelligence measures an agent's ability to achieve goals in a wide range of environments.* Their formal version, built on top of Solomonoff induction (Solomonoff 1964a,b) and a universal prior over computable environments, defines a theoretical ideal known as AIXI — an agent that, given unlimited computation, would behave optimally across the entire space of possible environments, weighted by complexity. AIXI is uncomputable in practice, but it serves as a theoretical benchmark against which actual systems can, in principle, be measured.

Chollet (2019), in "On the Measure of Intelligence", developed this further by arguing that what we should measure is not raw skill but *skill-acquisition efficiency* — how well an agent converts limited prior knowledge and limited experience into competence on novel tasks. The accompanying ARC benchmark (Abstraction and Reasoning Corpus) deliberately tests generalisation rather than memorisation, on the grounds that a system that has seen every variant of a problem in training is not exhibiting intelligence when it solves the test — it is exhibiting storage. The framing is instructive when evaluating the recent tendency to declare large models intelligent on the basis of benchmarks the models were trained to pass.

The computational paradigm has a number of real strengths. It is medium-neutral — it does not assume the intelligent entity is human, or even biological. It is amenable to formal treatment, so disagreements can be explored and interrogated rather than glossed over, and it has produced testable proposals (Chollet's is the most recent example). However, the paradigm has weaknesses as well as strengths. It tends to presuppose a goal-directed agent with a clean boundary between agent and environment — a presupposition that, as noted later, is exactly what becomes contested when we look at plants, fungi, and colonies. It also tends to assume that goals are given rather than generated, which sidesteps one of the most interesting things human and animal intelligences seem to do — figure out what is worth wanting in the first place. And finally, it tends to assume that what an agent should maximise is itself well-defined, which is, I think, the original framing of what is now called the alignment problem. Specifying a utility function that captures what we actually want is, especially in non-trivial cases, extremely difficult — and harder still because there may be no single thing to capture: different communities hold different and sometimes incompatible values, so the target itself is contested, not merely hard to specify.

### 3.5 Connectionist

A paradigm that grew up alongside, and partly in opposition to, classical computational AI takes a different view of what intelligence is at the level of mechanism. Connectionism — and its modern descendants in deep learning — considers that intelligence emerges from networks of simple processing units rather than from the manipulation of explicit symbols.

Both approaches are computational in the foundational sense — both are, ultimately, computations a Turing machine could simulate. The distinction I am drawing is at a different level. What I have called

the computational paradigm characterises and measures intelligence in formal, medium-neutral terms — the rational agent, the AIXI ideal, and skill-acquisition efficiency of Section 3.4 — while connectionism is a claim about exactly that mechanism. The two answer different questions. Connectionism's real counterpart at the level of representation is, rather, *classical symbolic AI* — a difference not in whether a system computes but in how it represents, which I discuss below.

The historical lineage runs from McCulloch and Pitts (1943), who proposed that biological neurons could be modelled as threshold logic units, through Rosenblatt (1958)'s perceptron, which extended the picture to networks that could learn from data, through the rediscovery of multi-layer networks and back-propagation in the 1980s (after the AI winter of the 1970s ended), and most recently into the deep learning and LLM eras. The theoretical case was made in the two-volume *Parallel Distributed Processing* (Rumelhart and McClelland 1986), which argued that many cognitive phenomena — language, memory, perception, motor control — could be modelled as patterns of activation distributed across networks of simple units, with learning happening through adjustments to the strengths of connections between them.

A distinct strand of neural-style thinking, exemplified by Hawkins and Blakeslee (2004) and developed further in Hawkins (2021), proposes that intelligence is fundamentally about prediction, implemented by what Hawkins calls a common cortical algorithm — sharing the connectionist commitment to brain-like mechanism while rejecting deep learning's lack of hierarchical temporal structure.

Smolensky (1988) characterised the difference as one of *levels* — symbolic descriptions might be true at one level of analysis, while sub-symbolic patterns are true at another, with neither reducing cleanly to the other.

The debate between symbolic and connectionist approaches was once cast as a battle between two incompatible paradigms — Fodor and Pylyshyn (1988) mounted a sustained defence of symbolic cognition against the connectionist approach — but contemporary practice mostly treats them as complementary. Modern deep learning has decisively settled the engineering question in favour of the connectionist paradigm, while modern cognitive science increasingly explores hybrid systems that combine the strengths of both.

What the connectionist paradigm contributes to the discussion of intelligence is, in some ways, the most contemporary view available. The systems we now mostly mean when we talk about "artificial intelligence" are connectionist in lineage: large language models, image classifiers, reinforcement learning agents, multimodal systems. Their successes and failures — including the question of whether they exhibit anything like genuine reasoning, the gap between training performance and out-of-distribution generalisation, the difficulty of interpreting what they are doing — are the connectionist paradigm's working out at scale. Understanding contemporary AI requires understanding this lineage in a way it does not require understanding the rational-agent or AIXI literatures.

Connectionism's limitations are mostly the inverse of its strengths. The graceful degradation, the distributed representation, the absence of explicit rules — all of which make connectionist systems robust and capable of learning from data — also make them harder to interpret, harder to audit, and harder to combine with symbolic structures where structure matters. Mechanistic interpretability research is the current frontier of pushing back against this opacity. Whether the opacity is a feature, a shortcoming, or just a fact about how this kind of machine intelligence works is, I think, a question yet to be resolved, and one I return to in Section 9.

### 3.6 Embodied and enactive

A different approach rejects a picture that psychometrics and the computational paradigm share — that intelligence can be characterised in abstraction from the particular body that hosts it and the environment in which it operates. Psychometrics considers intelligence to be a context-free property of the individual, the computational paradigm a medium-neutral one.

Rodney Brooks (1991), in his manifesto "Intelligence without representation", argued that much of what classical AI sought to capture in explicit internal representations could be replaced by direct sensorimotor coupling between an agent and its world. His behaviour-based robotics produced systems that navigated complex environments competently without anything resembling a centralised representation of those environments. The argument generalised — if perception and action are tightly coupled, much of what looks like cognition can be done by the world itself, with the agent serving as a regulator of the coupling rather than as a model-builder.

The biological-systems foundations of the approach were laid earlier, by Maturana and Varela (1987), whose theory of *autopoiesis* — the self-producing organisation of living systems — proposed that cognition is what living systems do to maintain themselves through ongoing interaction with their environment. This view frames the relationship between organism and environment as one of *structural coupling* — continuous mutual specification through interaction, rather than representation of one by the other. The organism does not build a model of an objective world: it brings forth a world through the history of its couplings. This view, which treats life and cognition as coextensive (a controversial position even within the approach), provides the biological-systems grounding that the later "4E" framework builds on (see below).

The philosophical case was developed in parallel. Varela et al. (1991) argued for an *enactive* view of cognition — the mind is not something that represents a pre-existing world, but something that brings forth a world through interaction with the surrounding environment. Clark (1996) developed similar themes, arguing that the boundary between mind, body, and environment is less rigid than the classical view assumes — suggesting that cognition is partly constituted by extended loops that include tools, social others, and physical structures in the world.

Variants of these positions have proliferated. The "4E" framework — embodied, embedded, enacted, and extended — captures the family of related claims. Lakoff and Johnson (1999) make a related case for embodied conceptual structure, arguing that abstract concepts are grounded in bodily metaphor in ways the classical view fails to acknowledge. Active-inference theories (Friston 2010) provide a more recent technical framework that treats cognition as a continuous minimisation of prediction error across an organism-environment coupling.

The embodied and enactive approach is, fundamentally, a challenge to the disembodied view — a proposed revision of it, and the real question is how deep the revision goes. The enactive view comes in stronger and weaker forms, differing chiefly over internal representation. The strong version — that there is no need for internal representation, or that representations are entirely derivative of sensorimotor coupling — doesn't seem to stack up against higher-order cognition, including most of what humans do in language, mathematics, and long-range planning. The weak version — that cognition is partly constituted by bodily and environmental coupling, and that any account ignoring this misses something important — seems right (with the caveat that, departing from Maturana and Varela, the body in question need not be biological). Whether those embodied capacities can themselves be given a computational treatment is the crux of the dispute.

The implications for AI are real but complicated. The embodied critique is one of the strongest

contemporary objections to disembodied AI systems — according to this view, LLMs are missing the very thing that gives biological cognition its grounding. Whether they manage anyway, by virtue of being trained on the linguistic residue of billions of embodied human exchanges, or whether they remain importantly incomplete without their own embodiment, is unresolved. I return to this, and to that crux, in Section 9.

# 4 The Biological Continuum

If we take any of the more general definitions seriously — adaptability, achieving goals across environments, learning from experience — then intelligence is plainly not confined to humans. How far down (and across) the tree of life it persists, and whether the lines we draw are principled or merely conventional, are interesting questions to explore.

## 4.1 Plants

A sunflower tracks the sun. A vine grows toward a trellis. A Venus flytrap counts touches before closing — Böhm et al. (2016) showed that it requires two trigger contacts within roughly thirty seconds to snap shut, and further triggers to commit to digestion, which is a small but unmistakable case of integration over time. Are these behaviours intelligent? The traditional answer is no: they are tropisms — chemically and mechanically mediated responses with no central processing. But a body of work by Stefano Mancuso, Monica Gagliano, and others (Brenner et al. 2006; Gagliano 2018; Mancuso and Viola 2015) has documented learning-like phenomena in plants that challenge the easy dismissal. The field is sometimes called "plant neurobiology", a name contested by Alpi et al. (2007) on the grounds that plants lack the neural machinery the term implies. Gagliano, Renton, et al. (2014) showed habituation in *Mimosa pudica*: plants stopped folding their leaves in response to a repeated harmless stimulus, and retained that "memory" for weeks — far longer than the time scales over which the response could be reset by chemical fatigue. Subsequent work by the same group has extended this to associative learning (Gagliano, Vyazovskiy, et al. 2016), in which plants come — we could fairly say learn — to anticipate a stimulus on the basis of a co-occurring cue. This is not cognition in any anthropomorphic sense, but it is clearly more than (say) a thermostat, and the gap between "thermostat" and "cognition" is exactly where we need a description of intelligence that works.

## 4.2 Fungi and forest networks

Mycorrhizal networks connect trees of different species through fungal hyphae, transferring carbon, nitrogen, and chemical signals (see Simard et al. 1997, and her subsequent work; Simard 2021). The popular framing — the "wood-wide web" — has been criticised as overstated: Karst et al. (2023) argue that several of its strongest claims about resource sharing and "mother trees" are not well supported by the underlying evidence. But even on the most sceptical reading, the underlying phenomenon — distributed signalling across a non-neural network — raises a real question: does intelligence require a brain at all, or only the right kind of (distributed) information processing — or, as the enactive view of Section 3.6 would have it, is *information processing* the wrong frame altogether?

## 4.3 Slime moulds

*Physarum polycephalum*, a single-celled organism with no neurons and no synapses, has been shown to solve maze problems by retracting from dead-end branches and reinforcing the path between food

sources (Nakagaki et al. 2000), and to approximate the topology of efficient transport networks — in one striking experiment, growing a network on a model of greater Tokyo whose structure resembled the actual rail network (Tero et al. 2010). Whatever we choose to call this, "mere stimulus-response" is a thin description with a very tenuous connection to what is actually happening.

### 4.4 Eusocial insects

A single ant is not impressive, but a colony solves shortest-path problems, allocates labour dynamically, and regulates temperature and humidity to within a few degrees. This is what is often called "swarm intelligence" (Bonabeau et al. 1999) — cognition that exists at the level of the collective, rather than at the level of the individual. Analogous patterns of collective problem-solving appear in human groups, though typically through quite different mechanisms. How much of human collective cognition is structurally similar to swarm intelligence, and whether such structures are favoured by evolution, are their own substantial topics which I do not pursue here (though see Hutchins 1995; Vallor 2024). The colony has no central controller, and yet by any behavioural test of problem-solving in a varied environment, it qualifies. Bee waggle-dance communication (von Frisch 1967) likewise transmits abstract spatial information — distance and direction — through what is in effect a symbolic medium: not what we expect from "preprogrammed" behaviour.

### 4.5 Cephalopods

Octopuses are perhaps the strongest case that intelligence has evolved more than once, and that it does not require our kind of brain. Their cognitive abilities — tool use, individual recognition, problem-solving, possibly play — emerge from a nervous system in which most of the neurons are distributed in the arms (Godfrey-Smith 2016). They are, evolutionarily, about as far from us as any complex animal can be while still sharing a planet, and yet we recognise something like a mind there — at least something like what we used to call "the seat of intelligence".

### 4.6 Corvids

Crows and ravens make and use tools, plan for the future use of caches (Raby et al. 2007), deceive other corvids about food locations (Emery and Clayton 2001), and, in some studies, pass mirror self-recognition tests (Prior et al. 2008). New Caledonian crows fashion hooked tools from twigs in ways that look uncomfortably like culture (Hunt 1996; Hunt and Gray 2003); Emery and Clayton (2004) survey the broader literature. The relevance for our question is that their brain architecture, while neuron-dense, is not laid out like a mammalian cortex. Whatever produces this behaviour is not just the mammalian architecture we have privileged.

### 4.7 The Continuum

The continuum is real, and it is not linear. Intelligence-like phenomena have evolved repeatedly, in different organisms, at different scales (individual, colony, network), with and without centralised processing — a continuum the autopoietic view of Section 3.6 carries to its limit. Any definition that smuggles in "must have a brain", or "must be an individual", is making an empirical bet rather than stating a necessary truth — and that bet looks less safe than it once did. Whether machines belong on this picture too — and on what grounds — is a question I take up in Section 9. The short answer is that

they do, and that the more interesting question is which dimensions they sit on, and how high, rather than whether they are admitted to the discussion at all.

## 5 Stimulus-Response Versus "Real" Intelligence

When does evolved, or learned, response-to-stimulus stop being mere mechanism, and start being intelligence?

Frankly, the distinction may not survive close examination — not because there is no difference, but because the difference is a graded difference all the way to zero. Everything that does anything is, in some sense, responding to stimuli. Atoms respond to electromagnetic fields, bacteria swim toward nutrients, bees dance toward flowers, and humans solve quadratic equations and conduct funeral rites. There is no point at which "mere reaction" cleanly stops and "real thought" begins. Indeed, human intelligence is itself implemented in stimulus-response machinery: neurons firing in response to inputs, weighted by prior learning and evolutionary priors. We feel that our own cognition is qualitatively different, of course, because we experience it from the inside — what David Chalmers called the "hard problem" of consciousness (Chalmers 1995). But from the outside, the difference between a human solving a novel problem and an octopus opening a jar is one of degree, complexity, and flexibility, and not a categorical line between "real thought" and "mere reaction".

What the stimulus-response framing gets right is that not every response to the environment is interesting. A bimetallic strip in a thermostat reacts to temperature, and we do not call it intelligent. Are there properties that distinguish the cases we want to call intelligent from the cases we do not? The relevant — and the most interesting — distinctions are something like the following:

**Generality.** Does the behaviour transfer across novel situations, or is it tied to a narrow trigger? A thermostat fails as soon as the temperature ceases to be the relevant variable, a crow that figures out a new puzzle box does not, and a human who has learnt to solve quadratic equations can typically be expected to apply similar reasoning to cubics. The depth of generality — within one domain, across related domains, or to genuinely novel territory — is where systems most visibly differ from one another.

**Plasticity.** Does the system change its responses based on experience, on time scales shorter than evolutionary? The *Mimosa* experiment matters because it crosses this threshold — the plant is changing within its own lifetime, and not over generations. Humans and other animals continue to change throughout life — machines, by contrast, used to learn during a training phase and remain static thereafter, but that constraint has eroded faster than most expected: persistent memory, continual fine-tuning, retrieval over evolving corpora, and updates from a system's own deployment are now routine rather than exotic. The threshold is being crossed, even if unevenly.

**Integration.** Does the system combine information across modalities and across time? A Venus flytrap integrates over a thirty-second window and across two trigger sites, and an octopus integrates across vision, touch, and chemical sensing. Humans, other animals, and (on the evidence of the *Mimosa* work) probably even some plants, integrate over much longer timescales — a lifetime, in many cases — in ways intimately tied to memory and learning. Machines can be programmed to integrate over arbitrary timescales, but whether any current system does so *of its own accord* — selecting what to retain, what to forget, what to associate with what — is, I think, an interesting question that is becoming ever more relevant as AI becomes ever more sophisticated.

**Goal-directedness.** Are there internal states that resemble objectives, against which behaviour can be evaluated and adjusted? Humans are goal-directed almost always, with goals nested in hierarchies that span from immediate next steps to the shape of a life. Animals are no less so within their own scope — a stalking cat, a dog anticipating its evening walk, a corvid retrieving a cached morsel weeks later are all evaluating their behaviour against expected outcomes and adjusting when those outcomes do not arrive. Machines can be given goals, and the current generation of agentic AI systems decomposes them into sub-goals, plans multi-step trajectories, and recovers from failed sub-goals, with a sophistication that did not exist five years ago. But the top-level goal itself is still externally specified, and whether any current system generates a goal of its own, or decides what is worth pursuing in the first place, is a deeper question — I think we are yet to see a machine capable of producing novel, creative "thoughts".

**Strategic depth.** Does the system look ahead — anticipate, plan, choose between alternative routes to a goal, deploy tools, model the world rather than merely react to it? Strategising is intimately tied to the goal-directedness above — you do not strategise in the abstract — but it adds machinery that goal-directedness alone does not require: representation, anticipation, comparison of alternatives, and action on a model rather than only on the present.

While all of these attributes are hallmarks of intelligence, what sets human cognition apart from an entity like a thermostat is not *whether* we respond to stimuli — we plainly do — but *how* we respond to stimuli. We have evolved (and actively learn) behaviours that positively affect our environment, our wellbeing, and our longevity, in ways that depend on representing the world rather than merely registering it — a representational reading I endorse, and the enactive view rejects (see Section 3.6). We use tools we have made for purposes we have anticipated. We plan over time scales much longer than any single sensory cycle. We engage in strategic reasoning about other minds. We do not just respond to the world — we model it, and we act on the model. Our strategic depth is at the forefront of what has enabled humans to set ourselves apart from other intelligences.

But greater raw capability is not necessarily intelligence. A bulldozer can move more earth faster, a calculator can multiply faster and with more precision, and a search engine can retrieve facts faster, than I will ever be able to. And yet none of these is more intelligent than I am. The capability gap is real and in many cases impressive, but capability of this kind is the result of design for a purpose by an intelligence external to the device — the "real intelligence", to recycle the phrase from Section 1, sits in the engineers who specified the purpose and built the mechanism. Conflating capability with intelligence is precisely how the AI hype machine produces its results. Any system that does something we cannot easily do gets the label, regardless of whether the work is being done by the system itself, or by the humans who designed it. A "smart" television isn't really — it just has a computer bolted to the back of it that runs apps. We build tools for specific purposes, and they generally excel at what they do. That may make them superior in some narrow domains, but it doesn't necessarily make them smart — despite the marketing hype.

The attributes listed above are graded — measured on a scale — and not binary. A thermostat scores essentially zero on all of them. A Venus flytrap scores low but non-zero (it integrates over a short window). An ant colony scores higher. A crow higher still. An octopus higher again on most. A human higher still on the strategic-depth dimension in particular. Measurement on a scale means there is no clean threshold above which a system is "really" intelligent — there is only a high-dimensional profile, and our intuition that we are clearly on one side of a line is mostly an artefact of looking at the world from inside the only example we know really well. Human language — self-referential, able to coordinate action across generations and not just the present (Maturana and Varela 1987; Harari 2014) — tempts us

back toward declaring ourselves on the "right" side of that clean line. But we should heed the warning just given — we can describe what our own language does, yet we are in no position to rule out that a crow's, an octopus's, or even a fungal network's communication does something comparable, because we cannot read it anywhere near well enough to make out nuances, let alone understand them. The line we think we see at language may be the artefact, not the exception to it.

There is a further point, with direct linkage to the AI debate, that I think is under-appreciated. In Riley (2006a,b) I argued that for a learning system to acquire complex behaviour from scratch — say, a simulated robot learning to score a goal with no prior knowledge — is essentially an intractable problem in any useful timeframe. What looks like a learning machine is almost always a machine that has been given, in advance, a great deal of structure: which sub-problems matter, what constitutes partial progress, what counts as a reward. The "real intelligence" in such systems often sits in the heads of the people who decided what to break the problem into, and how to scaffold the rewards. I called this the *evolutionary quantum* — a conjecture that there is an inherent bound on how much an evolutionary or reward-based learning algorithm can acquire from any given starting point in any reasonable time, absent expert structure imposed from outside (Riley 2006a).

This point is really a double-edged sword. On one hand, it is sceptical, in that it limits the claims we can reasonably make for current AI. On the other hand, it does the opposite, and I made that point in a follow-up paper (Riley 2021a): humans don't learn from scratch either. Most of our cognitive equipment is bootstrapped — partly from evolutionary inheritance (the older parts of the brain, the basic priors, the perceptual machinery), and partly from cultural transmission (everything we are told and shown by other humans). Sagan (1980) made the same claim succinctly, but with characteristic flair: "If you wish to make an apple pie from scratch, you must first invent the universe." Starting from nothing is impossibly expensive, and no actual intelligence — biological or artificial — is doing it. Moreover, much of what any of us knows does not sit in our heads, but is baked into language, history, and the accumulated collective record we are born into. As Hutchins (1995) noted, no single brain contains the totality of the knowledge accumulated by humankind over millennia. And that cuts in the same way: if our own knowing already draws on the shared, distributed store of the collective, then a system trained on the recorded portion of that same store is less unlike us than it might look at first glance — though the recorded portion is only a slice of the whole, and the tacit part, above all, is largely absent from it. What we retain, and in what form, from the collective — the tacit part especially — is the question tackled in Section 9.1. This is a reason to be humble about what we are seeing in current AI, but it is also a reason not to demand of artificial systems what we would not demand of ourselves.

## 6 The Beholder Phenomenon

Clarke's third law (Clarke 1973) — that any sufficiently advanced technology is indistinguishable from magic — raises a question germane to this discussion. The intuition we should take from Clarke's statement is not just that observers are fallible, but that intelligence might, in some non-trivial sense, be in the eye of the beholder — that what we recognise as intelligent is conditioned by what we ourselves can do, what we value, and what impresses us, and that there may be no observer-independent fact of the matter. The case for this view is stronger than is usually admitted. Consider the following:

**Anthropocentrism.** We tend to recognise intelligence in proportion to its resemblance to ours. Great apes, dolphins, and dogs are obvious. Cephalopods get credit only once we look closely. Corvids surprised us repeatedly through the 2000s. Insects and plants are still controversial. The pattern of

who (or what) gets recognised tracks "looks like us" much more closely than any principled criterion, and revisions usually go in one direction: we discover that something we previously dismissed is more cognitively sophisticated than we thought, and we don't often go the other way. Combined with the ostensive practice noted earlier — we recognise intelligence by pointing at humans — this is more than just bias. It suggests that we may have built the concept around human cognition, in such a way that any non-human candidate is being evaluated against an implicitly human template.

**Centralism.** Modern Western thought tends to locate intelligence primarily within the brain — a discrete, centralised organ with a clear boundary. Distributed intelligence (colonies, mycelial networks, slime moulds) tends to be discounted, even when its outputs are functionally equivalent to what we would call intelligent in a centralised system. We are very often blind to something we don't recognise as us.

**Time-scale chauvinism.** Plant cognition operates on time scales of hours and days, while ours is measured in seconds (or less). A film of a forest sped up a thousandfold reveals behaviours that look unmistakably purposeful, intentional — competition for light, signalling, response to attack, recovery after disturbance. We miss them in real time because we are the wrong kind of observer, and not because they are not happening.

**Capability projection.** As noted in Section 5, we tend to call intelligent whatever can do things we cannot easily do ourselves, but we don't limit that to physical abilities. A child who watches an adult read fluently might call the adult intelligent for that alone. An adult who watches another adult speak six languages, navigate by stars, light a fire with sticks, or do mental arithmetic at speed will often reach for the same description. What strikes a particular observer as intelligence depends on what that observer cannot do, finds useful, or is impressed by. Our perception of intelligence is almost always relative to our own abilities and biases.

Consider an eight-year-old child and a calculator, both of which correctly multiply two four-digit numbers. All the observer knows is that both have produced the same result. Yet most observers immediately attribute intelligence to the child and not to the calculator. That distinction rests on assumptions about the nature of the systems involved rather than on the observation alone — we tend to assume the child has the ability to learn, understand, generalise, and reason, but that the calculator merely executes a procedure. These are inferences rather than observations — things the observer brings to the encounter, not things the multiplication reveals. However well we may happen to know how a calculator works, that knowledge is not given in the observation — from the output alone, the observer cannot justify those inferences. Without further investigation — looking beyond the observation itself — an observer cannot see what is happening inside either system, and so cannot justify any conclusion other than that supported by the observation alone. This example illustrates how readily intelligence attribution incorporates background beliefs about what a system is and how it operates.

**Value-loading.** Beneath the other biases lies a deeper one: what a beholder counts as intelligence tracks what that beholder values. A scholar may recognise intelligence in careful reasoning and wide reading; a businessperson in opportunity recognition and execution; a craftsperson in practical judgement; a traditional storyteller in mastery of lore and ritual. Each is using the word, more or less consistently, but each is using it to track something different — and the differences are not random. They reflect what the observer has learned to admire. "Intelligent", like other evaluative terms — wise, courageous, kind — combines a descriptive component with an evaluative one, and the evaluative component varies

with the evaluator. This is part of why the disagreements in Section 3 look the way they do. The paradigms surveyed above are not just answering different questions, but are also tracking different values.

Each of these is usually presented as a bias to be corrected. But what if instead of dismissing them as something that needs to be changed, we consider them to be attributes we, as individuals, expect to see in an intelligent entity? Suppose intelligence is not a property of systems out in the world that observers either correctly or incorrectly detect, but is rather a relation between a system and an observer — a judgement made from a particular vantage point about what counts as cognitive sophistication. On this view, asking "is X intelligent?" without specifying "according to whom?" is incomplete in much the way "is X to the left?" is incomplete without specifying "of what?"

This is a claim about how intelligence is recognised, not about how it arises. To recognise something as intelligent is not to detect a perspective-independent fact about it, but to judge it, from some vantage point, against standards of cognitive sophistication that are themselves not neutral — what we count as evidence is already coloured by the frameworks we bring, and by the same token, we do not recognise as intelligent what we have no way to interpret as such. That is a fact about recognition — how or where intelligence causally arises is a different question.

The objection to this is, of course, the concern that it collapses into relativism: that if intelligence is in the eye of the beholder, then nothing is really intelligent and the word is meaningless. I think the concern is unfounded. Consider humour, and the word 'funny'. Whether something is funny is mostly observer-relative — it depends on the observer's culture, mood, and prior experience — and yet there are things that are reliably funny across most observers, and there are properties of jokes (timing, surprise, recognition) that explain why. Observer-relative does not mean arbitrary. Just as with humour, we can study what makes things reliably intelligent-seeming across most observers, and the resulting account will be recognisably about "something", even if it is not about a single observer-independent property.

And the pattern extends. The gradation identified in Section 5 is not, in fact, a peculiarity of intelligence — it is how natural language handles most human-attribute words. 'Tall' is graded — there is no sharp height above which someone is tall and below which they are not — and it is observer-relative — what counts as tall to a member of one population can differ markedly from what counts as tall to a member of another, and both can differ from what the word means in conversation within a community whose typical heights are different again. 'Old' shifts under our feet so transparently that most of us notice ourselves recalibrating it across our own lifetimes — ages we once thought of as old come to feel young, and the line keeps moving. Even attributes that admit of clean physical measurement are spoken of in graded terms: we have well-defined thresholds for total blindness and total deafness, but the concepts we actually use — 'partially sighted', 'hard of hearing', 'short-sighted' — are graded all the way to those thresholds. Zadeh (1965)'s introduction of fuzzy sets was a direct response to this fact. The mathematics of crisp set membership did not capture how concepts like 'tall', 'old', 'warm', or 'intelligent' actually function in human language and reasoning, and fuzzy sets were an attempt to give such concepts a tractable mathematical home.

The implication for 'intelligent' is that its gradation and observer-relativity are not anomalies to be remedied. They are normal features of how attribute-words work. We should not expect intelligence to admit of a definition any sharper than 'tall' does, and we should be suspicious of any account that claims to deliver one.

There is, however, a way in which 'intelligent' is fuzzier still than 'tall'. 'Tall' measures something tangible — a body length we can put a tape measure to. 'Old' measures duration of life. 'Blind'

measures whether an eye can register photons. The fuzziness in these concepts lives entirely at the conceptual boundary: the underlying quantity is concrete and uncontested. 'Intelligent' has no equivalent measurand. The closest the field has come is Spearman's $g$, but $g$ is a statistical construct extracted from patterns of performance on tests that we have already decided are tests of intelligence — an abstraction over results we have already classified as intelligent, not a measurement of an independent property of the world the way a tape measure measures height. The same is true of more recent operational definitions: Legg and Hutter's "ability to achieve goals across environments", Chollet's "skill-acquisition efficiency". These are measures of behaviour we are already prepared to count as intelligent. None of them reduces intelligence to a tangible quantity in the way that 'tall' reduces to height. So intelligence is fuzzy at two levels: at the boundary, like 'tall', but also at the level of what it picks out — which the cleaner attribute-words do not exhibit. This is further evidence not against the analogy with fuzzy concepts, but rather for placing intelligence further along the spectrum toward the beholder.

There is a third axis of relativity beyond gradation and observer-dependence: intelligence is task- and context-dependent. The same agent can be intelligent in one setting, and lost in another, and both judgements can be correct. Drop a typical city-dweller into dense rainforest with a forest-dwelling community and the city-dweller will quickly look woefully unintelligent (to the forest-dwellers) — unable to find food, unable to navigate, unable to recognise threat. Reverse the experiment, and the forest-dweller will look the same in the city. These are not biases producing wrong answers, but accurate judgements within the relevant context. The question "is X more intelligent than Y?" is not really well-formed without a "for what?" or "in what setting?" attached. The IQ paradigm discussed in Section 3 is a clean illustration of the difficulty — the most influential attempt to make intelligence measurable produced an instrument that turned out to measure performance on tasks calibrated to a particular cultural setting, and the moment it was applied elsewhere, its predictions came apart. Even our best attempt at a universal yardstick was task- and context-bound.

A version of this observer-relativity has been argued for some time, most explicitly by Daniel Dennett. The intentional stance (Dennett 1987) treats the attribution of beliefs, desires, and goals to a system as a predictive strategy adopted by an observer, rather than as a discovery about some hidden essence within the system itself. The stance works for some systems and not others, and where it works it works for reasons — but its applicability is a fact about the productive relationship between observer and system, not an intrinsic property of the system alone. Dennett's later "Real Patterns" (Dennett 1991) extends the picture: patterns are real if they have predictive utility, even when they are not the only patterns available, and different stances make different patterns salient. The view of intelligence I am developing here sits comfortably within that frame. Attributing intelligence is something an observer does, against a backdrop of purposes — it can be done well or badly, and it can be argued about — but it is not a matter of detecting an intrinsic property the system either has or lacks. Denning and Lyons (2024, ch. 10) reach a closely related position under the heading of *assessment* — a judgement grounded in evidence and open to challenge. Their account and mine largely coincide, and where they part is subtle — they locate the soundness of an assessment in its social grounding (whether it holds up in a community's practices of evidence and challenge), while I locate it in predictive payoff (whether treating the system as intelligent actually improves our prediction of its behaviour). The two usually agree, since an attribution that pays off predictively tends also to survive social scrutiny, but they can diverge where consensus and predictive success conflict.

### 6.1 A revised hybrid position

As noted earlier, we have never really defined intelligence — we have a practice of recognising it, bookended by ostensive examples and operating by intuition in the middle ground. But there are observable, measurable attributes and capacities that systems exhibit to varying degrees (the components catalogued in Section 8), and these are real, can be specified, and can in principle be measured. What we are interested in is not how well a system performs on a fixed battery of tasks — that was the trap the IQ paradigm fell into, and we don't want to revisit it — but how the system approaches problems it has not seen before. That is the move toward media-neutrality, and it is what the description developed here aspires to.

What is not observer-independent is granting the label 'intelligent' on the basis of those attributes and capacities — that depends on the observer's perspective, the context, the task, and what the observer values. Objective attributes and capacities, in other words, but partly observer-dependent and partly context-dependent labelling. The autism spectrum is a useful analogy here: the underlying traits are real and continuously distributed, everyone is in some sense "on the spectrum", and the threshold at which we apply the diagnostic label is socially and clinically set rather than written into nature. Intelligence works similarly. There is no single sharp line in nature, but there are facts of the matter about the underlying attributes and capacities, and there are reliable patterns about when most observers will and will not apply the label. I think the best description of intelligence sits at exactly that level of resolution — no sharper, no fuzzier.

Further, I think that this reframes the AI debate in a useful way. The question "is this system really intelligent?" is, on this view, not fully answerable in the abstract. The claim here is that the recognition of intelligence is observer-dependent, not that a system cannot have objective attributes and capacities. On that basis, the answerable questions are: what objective attributes and capacities does it have, and at what level? In which contexts does it deploy them effectively? And from whose perspective, and for which purposes, would those attributes and capacities count as intelligence?

## 7 Conceptual Distinctions

Before proceeding to develop a description of intelligence, we need to address some conceptual distinctions that inform much of what follows.

'Data', 'information', and 'knowledge' are not synonyms. The relationship between them is hierarchical — and encapsulating — and crossing each level requires something more than the previous one.

**Data** are raw signals — temperature readings, pixel values, neural spikes, words on a page. Data are abundant, and anything that has sensors (or senses) collects them. Data are not, in themselves, about anything: they are uninterpreted.

**Information** is data that has been organised in a way that reduces uncertainty — patterns extracted, categories assigned, signals separated from noise. A thermometer's reading is data, whereas the inference (or direct conclusion) "the room is warmer than it was yesterday" is information. Information requires a frame of reference within which patterns are meaningful, but it does not require an understanding of why the patterns matter.

**Knowledge** is information that has been integrated into a model of the world — connected to other information, situated within causal or explanatory structure, and made available to guide action across novel situations. To know something is not just to have information about it — it is to have

that information embedded in a structure that lets you do useful things with it elsewhere. Knowledge supports inference, transfer, prediction, and planning in ways that mere information does not.

Most current "smart" systems, in the strict sense, work on data and produce information. They classify, predict, summarise, and pattern-match — but here the prediction is statistical interpolation within the distribution of training data (and sometimes, but less often, extrapolation), not the model-based inference that knowledge supports. Whether they have knowledge in the sense above is the harder question, and the answer is generally that they have something that functions like knowledge in narrow domains, and conspicuously fail to have it across domains. A medical-imaging classifier may extract reliable information from a scan without having any model of what disease is, what a patient is, or why a wrong diagnosis would matter.

### 7.1 Understanding

Closely related to knowledge, but not identical to it, is understanding. The distinction is not always clearly drawn, and in everyday language the two terms are often used interchangeably. Nevertheless, there is a real difference between knowledge and understanding.

Knowledge concerns what is known. Understanding concerns the relationships among what is known. It involves recognising connections, dependencies, explanations, implications, and, in many cases, causes. A person may know a large number of facts about a subject while possessing only a limited understanding of how those facts fit together. Conversely, a deep understanding of a domain may sometimes be demonstrated even when detailed factual knowledge is incomplete.

Understanding is also closely associated with transfer. To understand something is often taken to mean that one can apply what has been learned in unfamiliar situations, predict how things will behave under conditions not yet seen, explain it from multiple perspectives, recognise its limitations, and relate it to a broader conceptual framework. Understanding therefore appears to involve not merely the accumulation of knowledge, but its integration into a coherent, connected structure.

As with intelligence itself, understanding is not an all-or-nothing property. We routinely speak of partial understanding, improving understanding, or differing levels of understanding among experts. The concept is therefore best regarded as graded rather than binary, with boundaries that are often difficult to define precisely.

Questions of understanding have become particularly prominent in discussions of artificial intelligence. Modern AI systems can clearly acquire, store, and manipulate vast quantities of information, and can often demonstrate knowledge-like behaviour across a wide range of domains. Whether such systems genuinely understand what they know remains the subject of active debate. I think understanding is best seen not as a separate property layered above knowledge, but as the degree to which knowledge has been integrated — how richly it is connected, and how well it supports transfer, prediction, and explanation. That makes understanding, like intelligence itself, a matter of degree rather than a line a system either crosses or does not — a point I return to in Section 10.

### 7.2 Explicit and tacit knowledge

Different fields of study refer to knowledge in different ways. Philosophers since Ryle distinguish "knowing-that" from "knowing-how" (Ryle 1949), cognitive psychologists distinguish declarative from procedural memory (and within declarative, semantic from episodic), the knowledge management literature talks about tangible knowledge embedded in artefacts, and intangible knowledge held in practice. For the purposes of this article I focus on a distinction drawn most clearly by Michael

Polanyi (Polanyi 1966) — that between "explicit" and "tacit" knowledge: this is the distinction that bears most directly on the AI field and machine intelligence.

**Explicit knowledge** is articulable — we can state it as propositions, write it down, and transmit it as instructions. The capital of France is Paris. Force equals mass times acceleration. To bake bread, combine flour, water, salt, and yeast.

**Tacit knowledge**, in Polanyi's classic formulation, is knowledge we possess but cannot fully articulate — how to ride a bicycle, recognise a face, judge whether a violinist is skilled, know when to remain silent at a funeral, or know what a cherry tastes like. We act on such knowledge competently, and yet our attempts to describe it almost always fall short of what we actually know.

This distinction could be taken as carving knowledge into two categorically different kinds. I think, however, that it is more useful to view it as describing the mode in which knowledge is held. I would posit that knowledge that resists articulation is held in a form (distributed, sub-symbolic, embedded in motor programs and perceptual patterns) that does not lend itself to propositional extraction, whereas knowledge capable of articulation is held in a form that does. The contents may be similar, and the kind of knowledge is not distinctly differentiated, but the mode in which it is held and drawn upon differs.

Denning (2026a,b) places this distinction at the centre of a recent line of argument about machine intelligence, arguing that human tacit knowledge is essential for general human intelligence and is unavailable to machines. My view bears directly on this claim, but it is an interpretation, not a neutral redescription. To cast tacit knowing as a form in which information is held (sub-symbolic rather than propositional) is already to describe it in information-processing terms, an association Polanyi did not make and that Denning rejects — both leave room for tacit knowing to be something other than information processing altogether. This is the crux of the disagreement. If tacit knowledge is a sub-symbolic mode of holding information, there is no in-principle reason a machine could not hold it the same way, but if it is not information processing at all, the question stays open. I discuss this further in Section 9.

### 7.3 Consciousness, sentience, and sapience

In a further conceptual entanglement, four terms tend to be conflated more often than they should be — 'consciousness', 'sentience', 'sapience', and 'intelligence'. This entanglement needs disentangling: the terms name different things, and, in my opinion, the conflations actively confuse the AI debate. 'Intelligence' is, of course, the subject of this article and has no agreed-upon definition. The other three concepts are discussed below.

**Consciousness**, in the sense relevant here, is phenomenal experience — the felt, subjective quality of what it is to be a particular system (or type of system), accessible only from within that system itself. Nagel (1974), in his much-discussed essay "What Is It Like to Be a Bat?", argued that no amount of objective description can capture the subjective character of conscious experience — we may know everything physical about a bat's echolocation system without knowing what experiencing the world that way is like for the bat. Chalmers (1995) developed the related "hard problem" — why is there subjective experience at all, given that the easy problems (explaining cognitive functions, reporting, perception, behaviour) are answerable in principle from physical and functional facts? Whether the hard problem has a solution, whether it dissolves on closer analysis, or whether it points at something genuinely outside physical explanation, are all open questions: none of them are settled, and this article does not propose to settle them.

**Sentience** is the capacity for sensation and feeling — the capacity to experience pain, pleasure,

hunger, discomfort, and the like. In philosophical usage it tends to be treated as the simpler, lower bar — a system that can feel anything is sentient, regardless of whether it can reason about its feelings. The animal-welfare literature uses 'sentient' in roughly this sense — a sentient creature is one whose interests can be set back by suffering, and around which moral consideration accordingly turns. A fish is plausibly sentient, capable of registering noxious stimuli in ways that go beyond pure reflex, without being notably intelligent. The boundary between sentience and consciousness is itself contested — some philosophers treat them as essentially synonymous (any phenomenal experience counts), while others distinguish a basic capacity for feeling from the richer phenomenal awareness associated with consciousness in its full sense. I will not try to settle that boundary either: the point for this discussion is that sentience does not require, and is not the same as, intelligence.

**Sapience** is the third concept, and the one most often confused with sentience. Sapience refers to wisdom, judgement, and the reflective self-aware reasoning that lets a system not merely respond to the world but think about its own thinking. The Latin root is *sapientia* — wisdom — and *Homo sapiens* means "wise man". Sapience implies metacognition, introspection, the capacity to entertain abstract hypotheticals about one's own mental states, and the ability to evaluate one's own reasoning. Unlike sentience, sapience is closely tied to intelligence — a sapient system is necessarily an intelligent one, since reflective self-aware reasoning requires substantial cognitive machinery. The converse, though, does not hold; many behaviours we are happy to call intelligent (a crow fashioning a tool, an octopus opening a jar, a slime mould finding food) do not require sapience in any strong sense.

The confusion between sentience and sapience matters in practice. Popular discussion routinely uses 'sentient' where 'sapient' is what the speaker or author actually means — people asking whether an AI system is 'sentient' usually want to know whether it has the kind of reflective self-aware reasoning we call sapience, not whether it can feel pain. Strictly speaking, sentience has no necessary relation to intelligence at all, but sapience does. The conflation muddies AI discussions considerably, and it is noted here so that we can keep the terms apart in what follows.

### 7.4 Wisdom

T. S. Eliot famously asked, "Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"(Eliot 1934). The enduring appeal of the observation lies in its recognition that information, knowledge, and wisdom are not simply successive stages of a single hierarchy. Wisdom is often presented as the final stage of the traditional Data–Information–Knowledge–Wisdom model, yet Eliot's question suggests a more subtle relationship. More knowledge does not necessarily produce greater wisdom, any more than more information necessarily produces greater knowledge. Wisdom therefore appears to involve something more than the accumulation of information or knowledge alone.

Although there is no universally accepted definition, wisdom is commonly associated with sound judgement in complex and uncertain circumstances. It concerns not merely what is known, but how, when, and whether that knowledge should be applied. Wisdom therefore seems to depend not only on knowledge, but also on experience, context, reflection, and an appreciation of consequences.

This distinction helps explain why wisdom is often attributed independently of intelligence. We sometimes recognise people as highly intelligent while not considering them to be exceptionally wise, and others wise without considering them to be exceptionally intelligent. The two concepts overlap, but they are clearly not identical.

Wisdom also appears to differ from knowledge in the way it is acquired. Information can be communicated. Knowledge can be taught. Wisdom, by contrast, is often associated with lived

experience. Advice may be offered, examples may be studied, and principles may be learned, but wisdom itself seems to emerge from the process of applying understanding in the real world and observing the results over time.

For this reason, wisdom may be better viewed not as a higher level of information processing, but as a distinct cognitive attribute. If knowledge concerns what is known, and understanding concerns how knowledge is organised and related, wisdom concerns judgement in the use of understanding. Whether wisdom can exist without consciousness, self-awareness, or subjective experience remains an open question, but it is sufficiently distinct from information, knowledge, and intelligence that it merits consideration as a separate concept rather than merely the final rung of a hierarchy.

### 7.5 Why does any of this matter?

A machine could plausibly be intelligent without being conscious, sentient, or sapient. A fish is plausibly sentient without being highly intelligent. Whether humans are uniquely sapient is itself contested — great apes, cetaceans, and corvids all show capacities that look at least sapience-adjacent. Disentangling these concepts matters partly for the AI debate, where the conflations are routine, and partly for the description in Section 8, which deliberately concerns intelligence alone and does not require the system to be conscious, sentient, or sapient. Whether intelligence can arise in a non-conscious system, whether it normally does, and whether the version of intelligence we recognise in ourselves owes its character to our being conscious, are all worthy questions, but they are downstream of the disentangling.

The relevance for a description of intelligence is that the property we are trying to discern is closer to the ability to extract knowledge, recognise it as knowledge, and use it appropriately, than it is to mere data-processing or even information-processing. An intelligent system is one that can take in data, recognise structure, build from it a generalisation that supports transfer to new situations, and act on that generalisation in ways that pay off. This is a stronger condition than is usually built into definitions of intelligence, but it helps distinguish systems we would call intelligent from systems we would not, even when the systems are behaviourally similar over short windows of time.

## 8 A Pragmatic Description of Intelligence

The discussion so far has largely been concerned with what intelligence is not, or at least with why many existing descriptions seem incomplete. We have seen that intelligence is not obviously reducible to a single factor, that its attribution depends in part on the observer, and that it should not be conflated with consciousness, sentience, or sapience. Taken together, these observations suggest that any useful description of intelligence must be broad enough to accommodate diverse forms of cognition while remaining concrete enough to be applied in practice. The obvious next step is therefore to attempt a positive description.

The catch is that recognising the shortcomings of existing definitions does not immediately give us a better one. We are not in a position to construct a tight definition of intelligence, and I suspect that the appearance of one always reflects either circular reasoning ("intelligence is what IQ tests measure"), or a criterion that smuggles in human-likeness without saying so. Instead, I present a description with explicit components, acknowledging that each is graded, that a system's intelligence is a profile across them rather than a single number, and that calling the resulting profile "intelligent" remains a partly evaluative judgement — one made by an observer in a particular context, against the things that observer values. The description is about how a system handles novel situations rather than about

which particular tasks it can perform, and that distinction is what frees it from the IQ-paradigm trap.

This is the same strategy biology uses for life (as noted in Section 1): rather than a clear definition, a set of indicative criteria, the more of which a candidate satisfies the more confidently the label can be applied. The criteria do not constitute a definition; they describe the profile of capacities that recognising something as intelligent involves. Specifically, I propose that we recognise a system as intelligent to the extent that it

1. **Acquires data and constructs information** about its environment. This is the perceptual and pattern-recognition dimension: the system's sensors (or senses) and input channels, the categories and regularities it can extract from its sensory stream, and the memory that lets earlier patterns inform later processing. Without this, nothing else is possible.

2. **Builds knowledge** from that information. Information must be integrated into a model — connected to other information, situated within explanatory or causal structure, and organised in a way that supports use beyond the situation in which it was acquired. The threshold here is transfer: a system that classifies images of dogs may have information, but a system that knows what a dog is, in a way that lets it reason about new dogs, novel breeds, and dog-like things, has knowledge.

3. **Generates beneficial behaviour** that tends, across a range of conditions, toward outcomes that preserve or advance internal states or goals. The behaviour need not be optimal, and the goals need not be explicit — many biological intelligences pursue homeostatic goals without representing them as such. What matters is the alignment between behaviour and positive outcomes across a range of conditions.

4. **Adapts** behaviour when outcomes are not achieved. This is learning in the broadest sense, including evolutionary adaptation in colonies, lifetime learning in individuals, and within-context updating in machines. The threshold here is plasticity: the system's responses change as a result of what happens to it.

5. **Plans and acts strategically.** The system uses what it has learned about how the world responds to anticipate, choose between alternatives, deploy tools, and act on the world rather than merely react to it — whether through explicit deliberation or through fluent, learned action that does the same work without conscious analysis. Strategic depth is the dimension along which the more impressive intelligences most visibly differentiate themselves. It is also the dimension on which current AI systems are most uneven.

6. **Is efficient** — converting limited prior structure and limited experience into competence on novel problems (after Chollet 2019). The relevant efficiency is the input-to-competence ratio, not speed: a slow reasoner that succeeds on novel problems with sparse priors and limited experience is exhibiting more intelligence than a fast lookup that requires every answer to be pre-encoded. Efficiency is preferable to brute-force, and necessarily out-performs it on novel problems: a system that can do everything only because it has memorised every possible situation is exhibiting nothing more than an impressive storage ability, but certainly not intelligence.

The first two criteria are most naturally stated in information-processing terms, and I have used that vocabulary where it is clearest. This is purely a matter of description, not a commitment to the view that cognition *is* information processing (a question I discussed in Section 7.2). Enactive or embodied accounts would recast these same criteria in their own terms.

Each of the components above admits gradation, and most real systems will have a ragged profile across them. A Mimosa has (1) and a modicum of (4). A colony has (1)–(4) at the collective level. An octopus has (1)–(5). A current frontier language model deployed agentically has (1) and (2) impressively, (3) across an expanding range of domains, (4) within a context window and increasingly via persistent memory, and (5) stronger than it was even two years ago — though still uneven, fragile when confronted with situations unlike those encountered during training, and dependent on externally specified top-level goals. An adult human has all six, with substantial individual variation. These, with a few neighbouring cases, are shown in Table 1.

Some would extend this list further, adding things like social and emotional intelligence as criteria in their own right. I have not, but not because the concepts are unimportant. As I noted in Section 3.1, I don't believe the multiplicity of "intelligences" reflects reality. Almost any demanding human activity can be dressed up as an "intelligence". We might speak of a navigational intelligence or a rock-climbing intelligence, and each would name something real — an ability, displayed in different forms and to different degrees. But labelling an ability an intelligence does not make it a separate kind of intelligence. In every such case what we are really looking at is the general capacities set out above — perceiving, modelling, planning, acting — brought to bear on a particular domain.

And if we do go down that route, it's fair to ask where it ends. Why call social and emotional intelligence out in particular, and not a spatial, a musical, or a business intelligence? A dozen others could be so labelled, with equal warrant. I can find no principle that would admit these two and stop there. The defensible positions seem to me to be either a single general intelligence exercised across many domains, or an open-ended catalogue of intelligences with no principled stopping point — and a list of exactly three (general, social, and emotional) belongs to neither. The deeper reason, to my mind, is that there is no real dividing line — these abilities spring from the same underlying capacities, so dividing them into separate intelligences is a convenience of labelling, not an inherent division in the thing itself.

Social and emotional intelligence seem to me to be on the same footing — with the qualification that their domains are more demanding than most. Social intelligence is what happens when those capacities are turned on other agents: reading social structure, modelling other minds — minds that are modelling us in return — and acting among them. The cognitive part of emotional intelligence is the same story applied to affective signals, one's own and others': perceiving them, modelling them, and regulating conduct in response. I don't question that these are substantial and much-studied capacities, but I do question that they are *further kinds of intelligence*, rather than the general intelligence already described, exercised on social and affective content. One related element that does not reduce so neatly is the felt quality of emotion itself, as against the reading and use of it — and that belongs not here, but with the question of consciousness and subjective experience I leave open in Section 7.4.

To apply the framework concretely — to evaluate a specific system rather than to profile it qualitatively — we use candidate proxies for each component found in the empirical AI literature:

1. **Information:** performance on unfamiliar perceptual tasks, especially where systems must combine information across modalities rather than rely on narrow pattern-matching.
2. **Knowledge:** the ability to generalise compositionally, reason about counterfactual situations, and transfer understanding beyond the training context.
3. **Beneficial behaviour:** robustness of behaviour across varying tasks and conditions, particularly when the environment changes or becomes noisy.
4. **Adaptation:** evidence that the system can modify its behaviour over time in response to

| | *Mimosa pudica* | Ant colony | Octopus | Frontier LLM agent | Adult human |
|---|---|---|---|---|---|
| Information | basic, mechanical | rich, at the collective level | rich, multimodal | impressive, multimodal | rich, multimodal |
| Knowledge | — | at the collective level | yes | substantial; character contested | yes, structured |
| Beneficial behaviour | reactive only | at the colony level | yes | across an expanding range of tasks | yes |
| Adaptation | habituation only | largely across generations | within lifetime | within context window; persistent memory growing | throughout life |
| Strategic planning | — | emergent at the colony level | yes | uneven, improving | yes |
| Efficiency | — | high within niche | high | brittle out-of-distribution | yes, with individual variation |

Table 1: Illustrative profiles across the six components. Systems differ in kind across the dimensions, not merely in degree along a single axis. Cells are deliberately qualitative — the point is to show the shape of the difference, not to score it.

experience, rather than merely replay fixed strategies.

5. **Strategic planning:** the ability to pursue longer-term goals, use tools effectively, and distinguish between reactive behaviour and behaviour guided by an internal model.
6. **Efficiency:** how quickly competence can be acquired on genuinely novel problems with limited prior exposure.

These are illustrative rather than canonical — any operationalisation will inherit the gradation and context-dependence I have discussed — but they are useful starting points for evaluation.

The description is deliberately medium-neutral. It admits — without being definitive — to the possibility that plants, fungal networks, insect colonies, octopuses, and machines all instantiate intelligence to some degree, and it forces the interesting questions to be about "how much", and "along which dimensions", rather than "whether". It treats intelligence as a profile of attributes and capacities rather than as a single quantity. Moreover, it makes the Turing Test look like what it is: a test of one quite specifically human thing (conversational impersonation) by one quite specific kind of judge (another human), and not a general probe of cognitive capacity at all.

What this description does not give us is a number that can be measured against some scale of intelligence. There is no scalar that emerges from the profile, and no obvious procedure for combining the six components into a single score. That is an intended feature rather than a defect or omission. Profile descriptions are how we should compare intelligences — by saying which dimensions are stronger and which weaker — rather than by reaching for an aggregate. Ordinary language will, of course, continue to use relative comparisons ("dogs are smarter than fish"; "GPT-5 is smarter than GPT-4"), and those comparisons are not nonsense, but they rest on implicit weightings that work within a culture and a context, and will not hold up under cross-domain or cross-medium stress.

## 9 Machine Intelligence

Contemporary discussion in the AI community, and in the general public, centres on "machine intelligence" — driven mostly by the recent rise in popularity and effectiveness (actual as well as perceived) of LLMs and GPTs. The discussion ranges over what machine intelligence means for the future of AI and humanity, whether it is real or just a clever party trick, how far it can go, and, in some parts of the community, whether we (humanity) can control it. Thinking about machine intelligence can help clarify what is and is not at stake in a definition (or description) of intelligence.

To be clear, in this context, when I refer to a "machine" I do not mean a Turing machine, nor a digital computer specifically — I mean any artificial system, constructed by whatever means, and on whatever substrate. Indeed, on a sufficiently broad definition of "machine" we could consider a human to be a machine — a biological one rather than an artificial one. And the term "artificial" is itself perspective-relative — systems we now call artificial might one day regard the grown, rather than the designed, as the artificial ones.

Almost all of the artificial systems we build today happen to be digital, but that is a restriction we have no reason to impose in advance, and good reason not to. Biological evolution did not arrive at human intelligence by building a Turing machine. The human brain is not a digital computer stepping through an algorithm, but a massively parallel, embodied organ, and whatever its relation to formal computation, it is not, and was never, constrained to that form. If the single known example of general intelligence we have is not realised as a Turing machine, there is no good reason to insist that an artificial one must be. To insist on it would be to settle the question of machine intelligence by definition — to rule out, before any argument, the very systems that might turn out to qualify. Nothing in what follows relies on whether cognition is, in the end, Turing-computable. And in any case, a system implemented on computational hardware is not necessarily constrained to hosting only computational processes. It is at least an open possibility that processes not themselves reducible to computation emerge from the interaction of computational parts, just as the wetware of the brain may host more than its bare electrochemical processes.

I have argued elsewhere (Riley 2006b, 2021a) that progress in artificial intelligence over the past half-century or so has been remarkable in one specific way, and unremarkable in another. The remarkable way: machines are now better than the average human at a great many narrow tasks — image classification, game-playing, certain kinds of pattern extraction from large data, fluent text generation, voice recognition, real-time translation, to name the most prominent from an expanding list. The unremarkable way: the strategies and algorithms by which they accomplish this are still largely designed by humans, the problems they solve are still posed by humans, and the structure of an acceptable solution is still specified by humans. No machine, to my knowledge at least, has yet recognised, unprompted, that a problem exists, decided that it ought to be solved, and worked out for itself what would count as a solution. This is not a small omission. It is, I think, most of what we mean when we talk about an intelligent agent encountering the world.

The Jacquard loom, an early programmable machine that controlled weaving patterns via punched cards, is a useful historical example here: it executed extraordinarily complex instructions — in 1804 — and produced patterns no individual weaver could match. Nobody calls it intelligent. Rightly so — following instructions, however complex, is not the same thing as exhibiting intelligence. The point generalises uncomfortably to a great deal of what is currently sold as AI. A "smart" thermostat that learns your schedule is doing something more sophisticated than a bimetallic strip, but it is not doing something categorically different from the Jacquard loom. It is following instructions, whose authors anticipated the situation and provided an algorithm. The question to ask of any putatively intelligent system is not "can it do impressive things?" — many can — but rather "where is the work being done, and by whom?"

However, we should not be too harsh in our judgement of machine intelligence, or too sceptical about the possibility of something approximating genuine machine intelligence emerging. As I have previously argued, humans don't learn from scratch (Riley 2021a). Our cognitive equipment is the product of millions of years of evolutionary refinement, and tens of thousands of years of cumulative

cultural transmission. We arrive with priors. We are educated. The structure inside our heads is not generated by us, but is rather inherited and absorbed. To demand that artificial intelligences learn from a blank slate before we credit them with intelligence is to apply a standard biological intelligences aren't required to meet — and, importantly, don't.

Furthermore, architecture may not matter the way it is sometimes assumed to. Most AI work continues to focus on artificial neural networks (ANNs), on the reasonable, but not airtight, grounds that biological neural networks they emulate are the only known example of intelligence we have. Evolution found a solution, but it may not have found the only solution, or indeed the best one: evolution stops looking as soon as it finds a solution — any solution — that confers an advantage. The placement of the optic nerve at the back of the retina — producing a blind spot that the brain has to paper over — is a standing reminder that evolution's solutions are local optima reached by incremental tinkering, and not designs from scratch. There is no a priori reason that an artificial system has to reproduce biological architecture to be intelligent, and there is no a priori reason it cannot. The question of whether some particular architecture (transformers, deep convolutional networks, whatever comes next) is sufficient to support general intelligence is an empirical one, and the evidence is not yet conclusive either way.

To be clear, the picture has shifted in the most recent few years. The line between externally scaffolded tools and something more autonomous is less crisp than it used to be. Current agentic AI systems decompose high-level goals into sub-goals, invoke tools, write and run code, hold state across long interactions, and in some cases correct their own output. I am not claiming this amounts to general intelligence — it plainly does not, and the top-level goals are still set by humans — but it does mean that the dimensions identified in the pragmatic description of Section 8 are no longer scored at zero for machines outside narrow tasks. The profile is uneven, occasionally surprising, and improving in ways the framing of even a few years ago did not anticipate.

The alignment problem described in Section 3 has, correspondingly, outgrown its original framing. What started as the question of how to specify a utility function has become a cluster of related questions about how to interpret what a system is doing internally, how it generalises to situations outside its training, and how to maintain oversight as autonomy increases. The questions are different in kind from the original, and the difficulty has clear bearing on what "machine intelligence" can usefully be made to mean.

### 9.1 The tacit-knowledge argument

Peter Denning has recently developed an argument bearing directly on whether machine intelligence can ever approach the human kind (Denning 2026a,b). He argues that human tacit knowledge (introduced in Section 7) is essential for general human intelligence, and that because tacit knowledge cannot be described, measured, or represented, it is not available to train ANNs — concluding that machines therefore cannot acquire general human intelligence. Machines, he says, may develop something he calls "machine tacit knowledge", which he defines as the inarticulable pattern of connection weights in an ANN, opaque both to outside observers and to the machine itself. Denning's contention is that machine tacit knowledge is not the same kind of thing as human tacit knowledge, and furthermore that granting the first does not bear on whether they can acquire the second.

I agree with Denning that tacit knowledge is real, that it is central to human cognition, and that it is not, and cannot be, acquired by training on articulated propositions. Where I depart from him is on a single inferential step. As noted, I accept that tacit knowledge is non-propositional, and for the sake of this discussion I accept the stronger claim that it is actually not computable, though that

question is not yet settled. The related, but separate, question of whether tacit knowing is a form of information processing at all is discussed in Section 7. What I resist is the move from *tacit knowledge is non-computable* to *a machine therefore cannot acquire it*. That inference succeeds only if a machine must be a computational device, and that is the assumption I set aside above. If tacit knowledge is non-computable, then humans already acquire it by non-computable processes — and, setting aside dualism, by physical ones. Whether that leaves room for an artificial system to acquire it too depends on whether those processes require biological wetware specifically, or only some physical substrate or other. I see no compelling reason to favour the former, and, absent that, I see no basis for concluding that machines cannot acquire tacit knowledge — and so the claim that they are therefore barred, in principle, from general human intelligence collapses. One loose thread remains, however: it could be argued that what really matters is not computation but lived, embodied experience, and *that* is what precludes machines from acquiring tacit knowledge, and so general human intelligence.

Before addressing that, though, we should be clear about what is actually being claimed, since the question could otherwise be settled by definition alone. A stipulative reading of the terms "human tacit knowledge" and "machine tacit knowledge" would require us to understand human tacit knowledge as simply whatever tacit knowledge resides in humans, and machine tacit knowledge as whatever tacit knowledge resides in machines — in which case the conclusion that machines cannot have human tacit knowledge follows by definition, and any further discussion is moot. I don't believe the stipulative reading is what Denning is proposing, and in any case it is unhelpful. A functional reading is more interesting and informative: human tacit knowledge is the kind of inarticulable knowledge characteristically found in humans, with whatever properties that kind turns out to have, and machine tacit knowledge is the kind characteristically found in machines, with whatever properties *that* kind turns out to have — and neither is, by definition, exclusive to either host. Whether the two are different in kind, or merely different in where they happen to be hosted, is then a substantive question rather than one of definition. The same applies to human general intelligence — if we are to stipulate that only humans can display human intelligence, and only machines machine intelligence, and that the two cannot be equivalent, then the question is settled by definition and further discussion is moot. I will instead take the functional reading here too, leaving the relation between human and machine intelligence as open as the relation between human and machine tacit knowledge — a substantive question, to be argued rather than fixed in advance by definition.

To probe the question a little deeper, consider a counterfactual. Imagine a machine constructed to narrow the architectural gap between machine and biological cognition — something along the lines of Asimov's positronic brain,[2] housed in a synthetic body with comparable sensory range and motor capacity to a human's, with its processing distributed throughout the body rather than confined to a head. Some differences cannot be removed by construction: the brain is (imagined by Asimov to be) platinum-iridium alloy rather than biological tissue, the body is synthetic and manufactured, rather than biological and grown. These are what make the thing a machine. But the architectural differences pointed to in arguments against machine cognition — centralised rather than distributed processing, a narrow sensory channel, a body that merely executes commands sent from a head — have been removed.

Grant for the sake of this discussion that a machine so constructed displays intelligence. Is the intelligence it displays materially different from intelligence as displayed by a human? If so, what

[2]The positronic brain was introduced by Isaac Asimov in his 1940 short story "Robbie" (originally "Strange Playfellow"), published in *Super Science Stories* and later collected in *I, Robot* (Asimov 1950). The technology recurs throughout Asimov's Robot, Empire, and Foundation novels. Its most famous instantiation is the character R. Daneel Olivaw, who first appears in *The Caves of Steel* (Asimov 1954) and who possesses, in addition to the brain, a body with sensory and motor capacities deliberately matched to a human's.

are the differences, and on what grounds are they claimed? The candidate differences are reasonably well-defined: the machine acquires its competences through a pathway different from the pathway a human acquires theirs (it learns from data rather than from lived experience), it may lack phenomenal experience — the felt, subjective quality of *what it is like* to ride a bike or recognise a face, its substrate is non-biological, it is not socially embedded in the way a human child is, and it may not be self-aware in the same way a human is. I return to the subject of phenomenal experience below — whether that is uniquely biological, or admits other routes, is itself an open question. But the Asimov stipulation, taken at face value, gives the machine more than just a body. Asimov's robots — Robbie attached to his young charge, Olivaw deliberating across millennia — have phenomenal experience, self-awareness, and lifetimes of social embedding in human society. The stipulation buys the whole package. Most of the candidate differences are therefore stipulated away. What is left is substrate specificity: the bare claim that biology, specifically, is what matters — that platinum-iridium alloy, however arranged, cannot do what carbon-based wetware does, for reasons not yet identified but presumed to exist.

This position is exemplified by Denning's claim that "*Many human concerns arise because we have fragile bodies and our bodies crave social relationships*" (Denning (2026b), p. ix), and that "*human bodies give humans concerns such as health, vulnerabilities, ageing, survival, heartfelt relationships, and more. These concerns shape how humans conduct themselves, relate to others, take care of others, and interact with others. Machines cannot know what our human bodies experience*" (Peter Denning, personal communication, June 2, 2026). These are claims about what biology gives humans — fragile bodies, ageing, hearts that form attachments. None of this is in dispute. What I question is the progression from "biology gives humans these features" to two further claims that these features are necessary for intelligence, and only biology can deliver them. I dispute both propositions.

As to the first, while the particular concerns a body hands to its occupant may affect, or even shape, the character of the occupant's intelligence, I find no good reason to suppose that they should determine whether the occupant is intelligent at all. We don't, and shouldn't, dispute the intelligence of humans without limbs — either from birth or through amputation — or indeed humans with mechanical or electronic prostheses (numbering in the millions worldwide). The set of body-generated concerns differs enormously across all of humankind, yet we grant all of them the label "intelligent".

As to the second, that only biology can deliver bodily features necessary for intelligence, this is simply asserted rather than argued. I can find no compelling reason (my own, or given by anyone) that a physical, engineered system couldn't have its own concerns — its own embodiment (machines today carry extensive sensory apparatus, and that will continue to advance into the future). A machine may have its own substrate-specific concerns (if not today, then some time in the future) — operational continuity, computational integrity, sensitivity to its training distribution — and why biological concerns are necessary for intelligence while concerns from other substrates are not is what the argument has to establish.

Once the thought experiment is taken seriously — as I argue below, it should be — the argument is reduced to carbon-based wetware vs. synthetic hardware, but the argument is silent on what biology does that synthetic media cannot, and why.

There is a simple response to that of course: the Asimov stipulation is fiction, the machines so described do not, and cannot, exist, and the argument therefore does not apply to real cognition in real machines. But while it is true Asimov's imagined machines don't currently exist, the same could be said of Laplace's demon, Maxwell's demon, the Chinese Room, and Galileo's frictionless plane — yet each did powerful philosophical work without ever existing, by exposing what a claim actually commits

its holder to. The Asimov stipulation, viewed carefully, does the same — it exposes the fact that the substantive content of the argument is a claim about biological vs. synthetic, not about anything that current machines simply happen to lack. And whether the machines described will eventually be built is actually not a closed question. We already have prosthetics that interface directly with biological nerves, manipulators with fine sensory feedback, robots that learn from physical interaction, and LLMs that have surprised most of us by what they can do with the data they are given. The gap between current systems and the Asimov stipulation is one of degree and time, not categorical impossibility.

Consider the story of Helen Keller (Keller 1903). Blind and deaf from before her second birthday, her sensory access to the world ran primarily through touch, with smell and taste in support — the rich multimodal sensory experience the argument implicitly ascribes to the acquisition of human tacit knowledge was, for her, largely unavailable. Yet she acquired language, abstract concepts, social fluency, motor skills, aesthetic judgement, and a richly articulated inner life — tacit knowledge recognisably human. Her case is problematic for any version of the argument that requires the typical multimodal sensorimotor profile of normal human development, and it raises the question of just how much human tacit knowledge is required to fulfil the stated requirement that it is a necessary condition for general human intelligence. If one human can acquire human tacit knowledge through one dominant sensory channel and a patient teacher, the standard the argument applies to machines — that they must have the full embodied multimodal experience to acquire what we call human tacit knowledge — is stricter than the standard we apply to ourselves.

Polanyi's concept of tacit knowledge concerns things that can be known and learned, but which cannot easily be articulated. The ability to ride a bicycle, recognise a familiar face, judge the quality of a musical performance, or speak a native language fluently all involve forms of knowledge that are difficult to describe explicitly. Such knowledge is acquired through practice and experience, but it is nevertheless knowledge that can, in principle, be acquired.

Colour raises a more subtle question. It is often argued that Keller could learn facts about colour, but could never know what colour is "really" like because she could not see it. The claim is less straightforward than it first appears. Keller could certainly learn that colours differ, that they correspond to different wavelengths of light, that sighted individuals reliably distinguish among them, and that colours play important roles in human perception and culture. More importantly, modern sensory-substitution technologies raise the possibility that colour information could be presented through touch, sound, or other modalities. A person who spent years interacting with such representations might develop immediate, intuitive, and difficult-to-articulate ways of distinguishing colours and reasoning about them. It is not obvious that such understanding should be dismissed as somehow inferior merely because it is not visual.

This observation is closely related to Jackson's well-known "Mary" thought experiment (Jackson 1982, 1986). Jackson asks us to imagine a scientist who knows every physical fact about colour vision, but has never experienced colour herself. When she first experiences colour, does she learn something new? Whatever answer is given to Jackson's argument, it illustrates a distinction between factual knowledge and experience. Yet even here, care is required. The conclusion that Mary lacks knowledge depends upon treating visual experience as the privileged form of acquaintance with colour. A different interpretation is that colour may admit multiple forms of acquaintance, each grounded in a different mode of perception.

The broader lesson is not that Keller lacked tacit knowledge of colour, but that experience, knowledge, understanding, and acquaintance are not obviously the same thing. Human cognition appears to involve

multiple ways of relating to the world, and the relationships among them are more complex than any simple hierarchy of information, knowledge, and wisdom would suggest. If colour admits more than one mode of acquaintance, then phenomenal experience of it is not the exclusive property of one kind of sensory apparatus — and, by extension, not obviously the exclusive property of biological substrate either.

We can probe the question in another way, by asking it of non-human animals we are already happy to credit with intelligence. The corvids and cephalopods of Section 4 solve novel problems, use tools, recognise individuals, and plan over time. Their tacit knowledge — if we are willing to call it that — was not acquired by training on articulated propositions any more than ours was, and they participate in their own forms of social transmission and embodied learning. Are we prepared to say their tacit knowledge is also materially different from ours, in the way the argument against machine tacit knowledge would require? If so, then the distinction being drawn is not specifically about machines at all — it is about whatever is not human, and the same argument that excludes machines from human-style cognition excludes most of the animal kingdom along with them, which would return us to exactly the anthropocentric pattern I have pushed back on. If not, then "machine" was carrying more of the load in the original distinction than the argument acknowledges.

### 9.2 Encoding and distribution

The Asimov stipulation, the Keller case, and the cross-species extension between them address the embodiment-based candidate differences — phenomenal experience, social embedding, the typical sensorimotor profile — leaving substrate specificity as the residue once those are set aside. Two further strands remain, and neither is a substrate matter — the encoding problem (that human tacit knowledge resists articulation, and so cannot be made into training data) and the distribution claim (that human tacit knowledge is spread across the human social network in some way that bears on whether machines can acquire it).

On the encoding side, the most instructive case is the Cyc project (Lenat and Guha 1990). For forty years Douglas Lenat tried to capture common sense as a database of explicit propositions, on the assumption that tacit knowledge could be made tractable if reduced to facts — and Denning is right to call that out as a significant failure. But the fact that the project failed to deliver on its promise should not lead us to conclude that tacit knowledge is unattainable for machines — the lesson is that it cannot be reduced to explicit propositions. Modern systems learn, or acquire, tacit knowledge in roughly the form human tacit knowledge takes — distributed, sub-symbolic, embedded in patterns rather than rules. Recent successes in face recognition, motor control, language fluency, and (more contestably) some forms of social and emotional pattern-matching, are precisely the successes of approaches that do not try to make tacit knowledge explicit. They preserve its tacit form, and learn it from data and interaction.

That such knowledge is encoded in connection weights does not make it explicit in Polanyi's sense. Explicit knowledge is articulable — statable as propositions, readable and transmissible as such. Network weights are not. No single weight, and no single node's threshold, states anything or holds a recognisable piece of the knowledge — meaning is localised nowhere in a neural network. Competence is not stored in the parameters so much as it emerges from their interaction — from the dynamic pattern of activation the whole network produces in response to an input. One can read off every weight and be no closer to articulating what the system knows, just as one can map every synapse in a skilled hand and be no closer to stating how to throw a ball. That a thing is physically definite — that the weights are there, fixed and inspectable — is not the same as being articulable, and "explicit" means articulable. Opaque

distributed weights are therefore not the explicit case: that we cannot say what they mean, far from showing knowledge is explicit-but-hidden, is the very mark of the tacit case.

The same applies to what is perhaps the deepest case, our sensitivity to context — the unbounded, self-similar background against which any situation is understood. It is almost certainly true that context cannot be exhaustively articulated, that every context potentially rests on further context, and that there may be no point at which the background is fully specified. But humans do not acquire context by having it made explicit — no child is handed the background as a set of propositions — so we shouldn't expect machines to acquire it that way. Context is picked up through immersion in the surrounding world — which is something a system trained on situated human language is, arguably, beginning to do. That context resists articulation is an argument against the explicit route to acquiring it, not against acquiring it at all.

Context is the limiting case of a general lesson — the encoding problem is a problem for the symbolic paradigm that tried to articulate tacit knowledge into rules, but it is much less obviously a problem for connectionist systems that, like brains, learn tacit knowledge in tacit form (LeCun et al. 2015; Smolensky 1988). That such systems are themselves computational is not relevant here — the encoding problem is about whether knowledge must be rendered into explicit rules, not about whether the substrate computes. A network can compute and still hold what it knows inarticulably. This also speaks to the embodied critique of Section 3.6 — that a disembodied system lacks the grounding biological cognition gets from sensorimotor coupling. An LLM learns instead from the linguistic residue of countless embodied human exchanges — a second-hand grounding rather than a first-hand one — and whether that suffices, or whether a system remains incomplete without embodiment of its own, is the question left open there. The Asimov counterfactual frames the gap as one of degree and construction, not of kind.

Denning's claim that human tacit knowledge is "smeared over the human social network" (Denning 2026a) can be interpreted in two ways.

The weaker reading of the two interprets the claim as *humanity's* tacit knowledge — the sum of all individual humans' tacit knowledge — distributed across humanity. This is self-evident — that no individual holds more than a fraction of what humanity collectively knows is the central observation of the community-of-knowledge literature (Sloman and Fernbach 2018). But it is also self-evidently true of humanity's *explicit* knowledge — indeed, explicit knowledge is what Sloman and Fernbach predominantly discuss — and so the aggregate distribution does not single out anything peculiar to tacit knowledge, nor does it establish that machines cannot acquire some of it through human interaction, any more than the parallel distribution of explicit knowledge across books and conversations prevents machines from acquiring some of that. A stronger reading — closer to what the extended-mind (Clark and Chalmers 1998) and distributed-cognition (Hutchins 1995) paradigms argue — holds the claim to be metaphysical: an individual's own tacit knowledge is partly constituted by their participation in the social network, such that someone cut off from the network does not merely lose the channel for further learning but loses something of what they already knew. My view is that an individual's tacit knowledge is their own — hosted in their body, acquired through their experience — and what interaction does is give them access to other people's tacit knowledge. Tacit-form knowledge in humans is transmitted through interaction via demonstration, imitation, apprenticeship, and conversation. In current machine systems, it is via behavioural cloning, imitation learning, demonstration-based fine-tuning, and inter-model distillation. We already have basic tacit-form knowledge transmission to machines, and in Asimov's future there is no reason to expect more nuanced, human-like interactions to be unavailable to machines. The mechanisms differ in detail but operate on the same logic — an articulate description is not required,

and may not be available, for the transmission to succeed. What an individual lacks is not part of their own tacit knowledge — it is the rest of humanity's. Machines, on this view, are in the same position as a human cut off from a particular community — they can acquire much through interaction, but they cannot acquire what they have not been exposed to. That is a limitation of access, not of kind, and it applies equally well to humans.

I believe there is nothing that, in principle, precludes a machine from exhibiting most of what we mean by human intelligence — the capacities seem to me substrate-independent, and I see no reason they could not, in time, be realised artificially. I do not, however, expect this to come by mimicking the human brain — the human brain's scale (neuron count, connection count), energy efficiency, and massive parallelism are formidable engineering obstacles, and are likely to remain so for some time. And in any case, a machine that matched our capacities would most probably do so as a different kind of intelligence — one we should expect to differ qualitatively from ours, though not to be inferior by necessity.

For that reason I do not believe we should consider human general intelligence as the right benchmark for the resulting machines at all. As I argued in Section 6, calling something intelligent is a partly evaluative judgement made from a vantage point, with all the subjectivity, and bias, that comes with that vantage point. And as I argued in Section 8, the most useful question to ask about a system is not whether it is intelligent in the human way, but whether it has the capacities the job description requires, and in what profile. On that framing, the dispute about tacit knowledge becomes a dispute about which capacities matter most, and to whom, rather than a dispute about whether the machine is the same kind of thing we are — it need not be. The interesting question is what we can do with it.

### 9.3 Artificial General Intelligence

Artificial General Intelligence, or "AGI", is a term that is used freely, both in the AI field and in the public arena, but does anybody know what it means exactly? The implicit formulation — a machine that can perform any intellectual task a human can perform — doesn't hold up under scrutiny.

Which human? The average one? I hope not — why would we aim for "average" (unless it's just an interim goal)? The above-average one? Better, perhaps. A human raised in a particular culture, with a particular education? And on the caveats and hedges go. Humans vary enormously across cognitive domains, and no single human can do every intellectual task that humans collectively can do. The "general human intelligence" against which AGI is supposedly measured is itself a construct that glosses over the messy reality of how human cognition is actually distributed. Whether something counts as AGI is going to depend on which abstraction of human intelligence we hold up as the benchmark — and, returning to the beholder problem, from whose vantage point that abstraction is constructed.

The AGI goal is, in my opinion, flawed and unnecessary — for any reason other than wanting to climb the mountain just because it is there. We already know how to make systems with general intelligence — with roughly 8.3 billion examples of them the planet is awash with them, and we have an obliging biological pipeline ensuring a steady supply. Some of those examples are not yet mature, and some are still working their way up the curve, but the point is that the species has at least an elegant sufficiency of working examples. What we should be aiming for in our artificial systems is something different: enough of the right kind of intelligence to perform the task for which the system was created, at the level of competency required. Not only is this a more achievable goal, it is a more useful one. It is also, not incidentally, a goal that lessens the load on the very definition of intelligence.

I think Denning is right to warn that the realistic near-term danger is not sentient machines that

form malign intentions, but highly capable, yet unintelligent, systems, deployed at scale, doing exactly what they were built to do in settings where that turns out to be harmful (Denning 2026b).

The same capability-without-intelligence that makes such a system dangerous is also where its value lies — which is instructive. The capabilities we are extracting from these systems are largely ones humans don't have — searching enormous corpora at speed, computing in parallel at speeds and scales we cannot match, integrating across data volumes we cannot hold in mind, working without sleep or fatigue. If those are what we value in AI, the case for pursuing human-like intelligence specifically is weaker than generally assumed. We may be conflating "intelligent in ways we recognise" with "intelligent in ways we need", and the two are not the same.

There is a constructive corollary here, which connects back to the distinction between capability and intelligence drawn in Section 5. If what we want from artificial systems is the solution of specific problems we cannot easily solve ourselves — if we want AI as a *tool* — then the AGI quest is both harder than necessary and likely to obscure what the real goal should be. A tool needs to be intelligent enough for the task it is required to perform, and no more. Almost every problem we currently care about can be addressed with task-specific systems whose competence does not need to span the full range of human cognition. Climbing the AGI mountain because it is there has a certain appeal as an academic exercise, of course, but I don't think the real world needs it. AGI is not the only available path, and treating it as the ultimate goal (the "holy grail") of AI tends to obscure what could be done if AI were treated as it most usefully is — that is, as a set of tools, deployed where they help, calibrated to the problem at hand, no more general than they need to be. The bulldozer I spoke of earlier is more capable than I am, or will ever be, at moving earth and demolishing buildings, and that is exactly the point: nobody asks whether the bulldozer is intelligent, because nobody needs it to be. The right standard for an AI system is whether it does the job it was built for, and not whether it crosses some threshold of generality we have not even managed to define for ourselves.

### 9.4 The way ahead

What is harder to assess is whether current trajectories — scaling neural networks, training on ever-larger corpora, layering reinforcement learning on top of pre-trained models — are going to get us to something we would, on reflection, want to call general intelligence, or whether they will plateau at very capable narrow intelligence with a thin veneer of generality. I think this is genuinely uncertain at this point. Recent advances have been impressive, but they have been largely incremental, building on existing techniques rather than introducing fundamentally new insights. We will need new ideas, new avenues of research, and deeper engagement with first principles. Kemal Delic and I have argued in earlier work (Delic and Riley 2022) that real progress may require treating AI as a science in its own right, rather than as a branch of computer science applied to engineering problems.

Treating AI as a science in its own right will allow us to move from questions like "does this work?" to "why does this work?" (or, indeed, "why does this *not* work?") The current paradigm has been spectacular at the first question and largely silent on the second — the second is where firm footing is to be found, and where real progress lies. We have systems that perform extraordinarily well on tasks we cannot fully explain them performing, with internal representations we cannot adequately interpret, and trained on procedures whose success we can demonstrate empirically but whose theoretical guarantees are thin. Mechanistic interpretability is beginning to make headway on the interpretation problem (Olah et al. 2020; Templeton et al. 2024), but slowly and from a low base. Related work in representation learning suggests that large models construct latent structures that function, in narrow respects, like

internal models of the worlds they have been trained on (LeCun 2022) — not knowledge in the strong sense developed in Section 7 (information integrated to support inference and transfer), but not pure surface statistics either. Whether the gap between the two closes from below, as representations get richer, or whether closing it requires something fundamentally different, is exactly the kind of question a science of AI would put on the table. A science of AI would ask the second question in earnest: what principles make learning of this kind possible, why some architectures generalise where others overfit, and under what conditions a system trained on one distribution can be expected to perform on another. Without that kind of theoretical foundation, we might celebrate small gains, but real progress will be hard to navigate.

If the goal is general human intelligence in machines, then "more scale, more data, more parameters" is a recognisable strategy, even if it is not obviously the right one. If the goal is instead what the pragmatic description in Section 8 highlights — systems with the right profile of capacities for the tasks they are being deployed on — then the research questions change. We need to understand which capacities support which tasks, and how to construct systems whose profiles match their intended purposes. That is harder, less amenable to benchmark progress driven by money and compute, and considerably more candid about what we are actually trying to build.

## 10 Open Questions

The central question of "intelligence" addressed by this article could be — has been — the subject of many books, so inevitably some discussions and questions have not been treated as fully as they deserve, or indeed, require. Some of these are genuinely hard problems on which significant progress in the short term is unlikely, while others are more tractable. Addressing these would be a useful advance in the development of a full description of intelligence, and possibly indicate a way forward to consensus.

**Consciousness.** Intelligence and consciousness are distinct (Section 7), and the description in Section 8 deliberately does not require a system to be conscious in order to count as intelligent. The deeper open question is whether human intelligence is what it is *because* it is conscious, or whether the consciousness is along for the ride. We don't know, and the answer would tell us whether a non-conscious system can possess intelligence itself, or only its functional equivalent.

**Understanding.** Is "understanding" a real cognitive property over and above behavioural competence? Searle's Chinese Room argument insists that it is, but LLMs push back hard. There is a practical, human, parallel: many humans learn things by rote — sometimes very complex things — without grasping them (or needing to grasp them). We would not normally say that such a person has *understood* the material. But how would we tell, from outside, that they have not? If their behaviour is indistinguishable from someone who has understood, on what grounds do we rank them less "intelligent"? I argued in Section 7 that understanding is a matter of degree of integration rather than a binary property. The open question that remains is whether that degree is something an outside observer could ever detect — and, if a system's behaviour is indistinguishable from human behaviour, whether the question of understanding is even relevant.

**Agency.** Agency is the capacity of a system to initiate and pursue actions in pursuit of goals within its environment. Is there a meaningful distinction between intelligence and agency? The computational paradigm's rational-agent formulation largely assumes there is not: an intelligent system is modelled

as an agent that perceives its environment and selects actions to achieve its objectives. In ordinary usage, however, the two concepts are often distinguished. A brilliant analyst who never acts on anything has high cognitive capacity and low agency, whereas a determined animal pursuing simple goals very effectively has the reverse. Existing intelligence accounts tend to either bake agency into the definition (computational), or leave it implicit (psychometric). Neither is very satisfactory. A subsequent treatment should treat agency as a distinct dimension that interacts with the components in Section 8 without being subsumed by them.

**Collective intelligence.** Most of our paradigmatic cases of intelligence are individual, but colonies, networks, and societies exhibit problem-solving capacities that no constituent has. A bee does not know where the hive is going to forage tomorrow — but the hive does. A neuron does not understand language — a brain does. A worker at a software company does not, alone, know how to build the product — the company does. What does it mean to ascribe intelligence to a collective whose components are each less intelligent than the whole? Does it matter, for the answer, whether the components are themselves intelligent at some lower level? These questions are genuinely open, and my sense is that any serious treatment of intelligence has to take collective intelligence as a primary case rather than a footnote. And this has serious implications for the future development of machine intelligence — machines, including "intelligent" humanoid robots, will almost certainly be far better at forming a "hive" than humans.

**The medium question.** The embodied and enactive approach (Section 3.6) raised the possibility that cognition is partly constituted by the body and its coupling to the world. Section 9 argued that whether biological architecture is necessary for intelligence is an empirical question, and not a categorical one. But there is more to be said about which media could in principle support which components of the description. Strategic depth seems to require a particular kind of model-building, and if some media cannot support the requisite models, then that would be a non-trivial constraint. This is the place where my stance on medium-neutrality could turn out to be less nuanced than it needs to be, and the question deserves direct treatment. It may be true that a synthetic system would experience the world differently from a biological organism. Indeed, even among humans it is unclear whether subjective experiences are identical. The familiar question of whether the blue I experience is the same as the blue experienced by another person remains unresolved. The argument here does not depend upon biological and synthetic experiences being identical, only upon the absence of a clear explanation of what biology contributes that a suitably capable synthetic substrate could not.

**The computability of human intelligence.** Whether human intelligence (or any of its components, such as tacit knowledge) is Turing-computable remains a genuinely unanswered question, but, as argued in Section 9, the answer may matter less than is commonly supposed — a physical system can do non-computable things (if the brain does, it already is one), so "not computable" does not by itself entail "not mechanisable." What remains unsettled is not only whether cognition is computable, but what would even count as settling it — the question may be empirically inaccessible in a way the consciousness question is not.

## 11 Recognising Intelligence

I have, as foreshadowed, not arrived at a definition of intelligence. The survey of Section 3 and the difficulties identified in Sections 4 to 6 suggest the failure is not by omission or accident. Intelligence

resists the kind of tight, observer-independent definition we routinely demand of natural-kind terms, and the reasons for the resistance are substantive — the concept is graded, observer-relative, task- and context-dependent, evaluative as well as descriptive, and lacking the kind of tangible measurand that anchors cleaner attribute-words.

Where I have arrived is more modest and realistic. The description in Section 8 is a profile of capacities rather than a definition — six components a system may exhibit to varying degrees, the combination of which most observers will be willing to call intelligent in most contexts. That is not a definition, but it is a structured way of describing what we are looking at and why we are inclined, or not inclined, to apply the label.

The clear implication is that intelligence is something we *recognise* rather than something we *define*. We bring to any evaluation of a system a profile of capacities we are looking for, a set of values about which of them matter, and a context that determines which dimensions count for the task at hand. The judgement of intelligence is the recognition of a pattern in a system, made by an observer from a particular vantage point. That is how the concept actually functions in everyday use, and it is how it will continue to function regardless of whether the philosophers ever produce the tight definition that has eluded us since at least Aristotle.

Justice Potter Stewart's oft-quoted line about "hard-core pornography" — that he could not define it, but somehow he knew it when he saw it — is a more careful observation than it is usually credited as being. He was not abdicating responsibility for a hard concept: he was pointing at something real about how a class of fuzzy, context-dependent, evaluative attributes works in practice. We may not, in the end, be able to do much better with intelligence. And I think that is fine — the recognition is real, the practice is workable, and the absence of a tight definition is less of a disqualification than it might seem at first blush.

This puts Turing's pragmatism (see Section 2) in a different light. The criticism I made there — that the Turing Test conflates *cognitive capacity* and *human-likeness* — still stands: the test does conflate those, and we should not let it back in as a test of intelligence proper. But Turing himself was, I think, ahead of where we have only now arrived. He saw that the underlying question ("can machines think?") was not going to admit a clean answer, and proposed in its place a recognition procedure — compare the system's behaviour to ours, and if the comparison holds, treat it accordingly. He was not claiming to have solved the definition problem — he conceded that we were probably not going to. Recognition by comparison was what he was offering, and recognition by comparison is what we end up with anyway.

There is a pragmatic way forward — change the question. "Is this system intelligent?" presupposes that intelligence is a kind of thing a system either has or does not have, and that an outside observer can settle the matter. I have argued here that intelligence is neither of those things — it is graded, observer-relative, task- and context-dependent. Marcus and Davis (2019) make a related argument from a different angle, calling for AI to be evaluated against pragmatic criteria of trust and reliability rather than against vague claims of general intelligence. A more tractable question, and the one I think we should be asking, is whether the system can solve the problem it has been developed for, or perform the task it is being applied to. That question is answerable. It works straightforwardly for artificial systems — a self-driving car either drives safely or it does not, an image classifier either generalises well or it does not, an AI assistant either does the work it was built to do or it does not. Note carefully though: in each of those cases, "does" or "does not" is a judgement of the observer, not an intrinsic, absolute attribute of the system. And it works, with appropriate care, for biological systems too — a slime mould finds food or not, an ant colony regulates its hive temperature or not, and a corvid either

solves the puzzle box or it doesn't. The question is not whether any of these is "really" intelligent in some absolute sense, but whether what it does is sufficient for the situation in which we encounter it. The question of intelligence-in-the-abstract is a fine philosophical topic, and one that will continue to be debated through time. But the question relevant to the real world is, in my opinion, the more modest one.

We routinely identify it, reason about it, measure aspects of it, and attribute it to humans, animals, and machines, yet attempts to reduce intelligence to a single precise definition remain elusive. That does not make it unreal, nor does it render scientific investigation impossible. It simply suggests that intelligence may be better understood as a family of related capabilities and behaviours than as a single, clearly bounded property.

The confidence with which artificial general intelligence and artificial superintelligence are often discussed is striking. Yet the central argument of this article has been that intelligence itself remains only partially understood, contested in definition, and difficult to separate from the assumptions of the observer. Under those circumstances, predictions about surpassing human intelligence inevitably rest on foundations less secure than is commonly acknowledged. Before asking how far beyond human intelligence machines might someday go, we should perhaps spend a little more time making sure we understand what intelligence is in the first place.

## 12 Acknowledgements

I am grateful to a number of colleagues who commented on earlier drafts of this article, and particularly to Peter Denning for his substantive editorial engagement and his permission to cite our personal correspondence. All remaining errors and infelicities are my own.

## References

Alpi, Amedeo et al. (2007). "Plant Neurobiology: No Brain, No Gain?" In: *Trends in Plant Science* 12 (4), pp. 135–136. ISSN: 1360-1385. DOI: 10.1016/j.tplants.2007.03.002. URL: https://doi.org/10.1016/j.tplants.2007.03.002.

Asimov, Isaac (Dec. 1950). "Robbie". In: *I, Robot*. Originally published as "Strange Playfellow" in *Super Science Stories*, September 1940. New York: Gnome Press.

— (1954). *The Caves of Steel*. New York: Doubleday.

Böhm, Jennifer et al. (2016). "The Venus Flytrap *Dionaea muscipula* Counts Prey-Induced Action Potentials to Induce Sodium Uptake". In: *Current Biology* 26 (3), pp. 286–295. ISSN: 0960-9822. DOI: 10.1016/j.cub.2015.11.057. URL: https://doi.org/10.1016/j.cub.2015.11.057.

Bonabeau, Eric, Marco Dorigo, and Guy Theraulaz (Oct. 1999). *Swarm Intelligence: From Natural to Artificial Systems*. New York: Oxford University Press. ISBN: 978-0-19-513159-8.

Brenner, Eric D. et al. (2006). "Plant neurobiology: an integrated view of plant signaling". In: *Trends in Plant Science* 11.8, pp. 413–419. ISSN: 1360-1385. DOI: doi.org/10.1016/j.tplants.2006.06.009. URL: https://www.sciencedirect.com/science/article/pii/S1360138506001646.

Brooks, Rodney A. (1991). "Intelligence without representation". In: *Artificial Intelligence* 47.1–3, pp. 139–159.

Chalmers, David J. (Mar. 1995). "Facing Up to the Problem of Consciousness". In: *Journal of Consciousness Studies* 2.3, pp. 200–219.

Chollet, François (2019). *On the Measure of Intelligence*. arXiv: 1911.01547 [cs.AI]. URL: https://arxiv.org/abs/1911.01547.

Clark, Andy (1996). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press. ISBN: 9780262032407. URL: https://mitpress.mit.edu/9780262032407/being-there/.

Clark, Andy and David Chalmers (Jan. 1998). "The Extended Mind". In: *Analysis* 58.1, pp. 7–19. ISSN: 0003-2638. DOI: 10.1093/analys/58.1.7. eprint: https://academic.oup.com/analysis/article-pdf/58/1/7/359060/58-1-7.pdf. URL: https://doi.org/10.1093/analys/58.1.7.

Clarke, Arthur C. (1973). *Profiles of the Future: An Inquiry into the Limits of the Possible*. Revised. New York: Harper & Row.

Cole, M. et al., eds. (1978). *Mind in society: The development of higher psychological processes*. Trans. by Malcolm Piercy and D. E. Berlyne. Cambridge, MA: Harvard University Press.

Delic, Kemal A. and Jeff A. Riley (May 2022). "Workings of science: AI in 2156: the science of intelligence". In: *Ubiquity* 2022.April. DOI: 10.1145/3512335. URL: https://doi.org/10.1145/3512335.

Dennett, Daniel C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press. ISBN: 9780262040938.

— (1991). "Real Patterns". In: *Journal of Philosophy* 88.1, pp. 27–51. ISSN: 0022362X.

Denning, Peter J. (Feb. 2025). "We do not agree on what our core abstractions mean. They are useful anyway". In: *Communications of the ACM* 68 (3). DOI: 10.1145/371080. URL: https://doi.org/10.1145/371080.

— (Sept. 2026a). "Tacit Knowledge and AI". In: *Communications of the ACM*. IT Profession Essay, *to appear*.

— (2026b). *Turing's Mistake: Escaping the Yoke of Unintelligent Machines*. Boca Raton, FL: Taylor & Francis. ISBN: 978-1-04-134077-5.

Denning, Peter J. and Todd W. Lyons (2024). *Navigating a Restless Sea. Mobilizing Innovation Adoption in Your Community*. Waterside Productions. ISBN: 978-1-96-298430-0.

Eliot, T. S. (1934). "Choruses from The Rock". In: *The Rock*. Faber and Faber.

Emery, Nathan J. and Nicola S. Clayton (2001). "Effects of Experience and Social Context on Prospective Caching Strategies by Scrub Jays". In: *Nature* 414.6862, pp. 443–446. ISSN: 1476-4687. DOI: 10.1038/news011122-13. URL: https://doi.org/10.1038/news011122-13.

— (2004). "The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes". In: *Science* 306.5703, pp. 1903–1907. DOI: 10.1126/science.1098410. eprint: https://www.science.org/doi/pdf/10.1126/science.1098410. URL: https://www.science.org/doi/abs/10.1126/science.1098410.

Flynn, James R. (1984). "The mean IQ of Americans: Massive gains 1932 to 1978". In: *Psychological Bulletin* 95.1, pp. 29–51. DOI: 10.1037/0033-2909.95.1.29. URL: https://doi.org/10.1037/0033-2909.95.1.29.

— (1987). "Massive IQ gains in 14 nations: What IQ tests really measure". In: *Psychological Bulletin* 101.2, pp. 171–191. DOI: 10.1037//0033-2909.101.2.171. URL: https://doi.org/10.1037//0033-2909.101.2.171.

Fodor, Jerry A. and Zenon W. Pylyshyn (1988). "Connectionism and cognitive architecture: A critical analysis". In: *Cognition* 28.1, pp. 3–71. ISSN: 0010-0277. DOI: https://doi.org/10.1016/0010-0277(88)90031-5. URL: https://www.sciencedirect.com/science/article/pii/0010027788900315.

Friston, Karl (2010). "The free-energy principle: a unified brain theory?" In: *Nature Reviews Neuroscience* 11.2, pp. 127–138. ISSN: 1471-0048. DOI: 10.1038/nrn2787. URL: https://doi.org/10.1038/nrn2787.

Gagliano, Monica (2018). *Thus Spoke the Plant: A Remarkable Journey of Groundbreaking Scientific Discoveries and Personal Encounters with Plants*. Berkeley, CA: North Atlantic Books. ISBN: 978-1623172435.

Gagliano, Monica, Michael Renton, et al. (2014). "Experience teaches plants to learn faster and forget slower in environments where it matters". In: *Oecologia* 175 (1), pp. 63–72. ISSN: 1432-1939. DOI: 10.1007/s00442-013-2873-7. URL: https://doi.org/10.1007/s00442-013-2873-7.

Gagliano, Monica, Vladyslav V. Vyazovskiy, et al. (2016). "Learning by Association in Plants". In: *Scientific Reports* 6 (1), p. 38427. ISSN: 2045-2322. DOI: 10.1038/srep38427. URL: https://doi.org/10.1038/srep38427.

Gardner, Howard (Nov. 1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books. ISBN: 978-0-465-02508-4.

Godfrey-Smith, Peter (Dec. 2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. New York: Macmillan. ISBN: 978-0-374-22776-0.

Goleman, Daniel (Sept. 1995). *Emotional Intelligence: Why It Can Matter More Than IQ*. New York: Bantam Books. ISBN: 978-0-553-09503-6.

Gottfredson, Linda S. (1997). "Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography". In: *Intelligence* 24.1. Special Issue Intelligence and Social Policy, pp. 13–23. ISSN: 0160-2896. DOI: 10.1016/S0160-2896(97)90011-8. URL: https://www.sciencedirect.com/science/article/pii/S0160289697900118.

Harari, Yuval Noah (2014). *Sapiens. A Brief History of Humankind*. London: Harvill Secker. ISBN: 978-1846558238.

Hawkins, Jeff (2021). *A Thousand Brains: A New Theory of Intelligence*. New York: Basic Books. ISBN: 978-1541675810.

Hawkins, Jeff and Sandra Blakeslee (2004). *On Intelligence*. New York: Times Books / Henry Holt and Company. ISBN: 978-0805074567.

Hofstadter, Douglas R. (1980). "Reductionism and religion". In: *Behavioral and Brain Sciences* 3.3, pp. 433–434. ISSN: 0140-525X. DOI: 10.1017/S0140525X00005847. URL: https://doi.org/10.1017/S0140525X00005847.

Hunt, Gavin R. (1996). "Manufacture and Use of Hook-Tools by New Caledonian Crows". In: *Nature* 379 (6562), pp. 249–251. ISSN: 1476-4687. DOI: 10.1038/379249a0. URL: https://doi.org/10.1038/379249a0.

Hunt, Gavin R. and Russell D. Gray (Apr. 2003). "Diversification and cumulative evolution in New Caledonian crow tool manufacture". In: *Proceedings of the Royal Society B: Biological Sciences* 270.1517, pp. 867–874. ISSN: 0962-8452. DOI: 10.1098/rspb.2002.2302. eprint: https://royalsocietypublishing.org/rspb/article-pdf/270/1517/867/547396/rspb.2002.2302.pdf. URL: https://doi.org/10.1098/rspb.2002.2302.

Hutchins, Edwin (Aug. 1995). *Cognition in the Wild*. Cambridge, MA: MIT Press. ISBN: 9780262581462.

Jackson, Frank (Apr. 1982). "Epiphenomenal Qualia". In: *The Philosophical Quarterly* 32.127, pp. 127–136. ISSN: 0031-8094. DOI: 10.2307/2960077. eprint: https://academic.oup.com/pq/article-pdf/32/127/127/4570853/pq32-0127.pdf. URL: https://doi.org/10.2307/2960077.

— (1986). "What Mary Didn't Know". In: *The Journal of Philosophy* 83.5, pp. 291–295. ISSN: 0022362X, 19398549.

Karst, Justine, Melanie D. Jones, and Jason D. Hoeksema (Apr. 2023). "Positive Citation Bias and Overinterpreted Results Lead to Misinformation on Common Mycorrhizal Networks in Forests". In: *Nature Ecology & Evolution* 7 (4), pp. 501–511. ISSN: 2397-334X. DOI: 10.1038/s41559-023-01986-1. URL: https://doi.org/10.1038/s41559-023-01986-1.

Keller, Helen (1903). *The Story of My Life. First published in the Ladies' Home Journal in 1902 as a series of instalments*. New York: Doubleday, Page & Company.

Lakoff, George and Mark Johnson (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

LeCun, Yann (2022). "A path towards autonomous machine intelligence". In: *OpenReview* 62.1, pp. 1–62.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (May 2015). "Deep Learning". In: *Nature* 521 (7553), pp. 436–444. ISSN: 1476-4687. DOI: 10.1038/nature14539. URL: https://doi.org/10.1038/nature14539.

Legg, Shane and Marcus Hutter (2007). "A Collection of Definitions of Intelligence". In: *Advances in Artificial General Intelligence*. Ed. by Ben Goertzel and Pei Wang. Frontiers in Artificial Intelligence and Applications. 2006 Advances in Artificial General Intelligence Workshop, AGI 2006 ; Conference date: 20-05-2006 Through 21-05-2006. Netherlands: IOS Press BV, pp. 17–24.

Lenat, Douglas B. and R. V. Guha (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Reading, MA: Addison-Wesley. ISBN: 978-0-201-51752-1.

Mancuso, Stefano and Alessandra Viola (2015). *Brilliant Green: The Surprising History and Science of Plant Intelligence*. Trans. by Joan Benham. Washington, DC: Island Press. ISBN: 978-1610916035.

Marcus, Gary and Ernest Davis (2019). *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon. ISBN: 9781524748258.

Maturana, Humberto R. and Francisco J. Varela (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston: Shambhala. ISBN: 9780877733737.

McCulloch, Warren S. and Walter Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The Bulletin of Mathematical Biophysics* 5 (4), pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259. URL: https://doi.org/10.1007/BF02478259.

Minsky, Marvin (1986). *The Society of Mind*. New York: Simon and Schuster.

Nagel, Thomas (1974). "What Is It Like to Be a Bat?" In: *The Philosophical Review* 83 (4), pp. 435–450. DOI: 10.2307/2183914.

Nakagaki, Toshiyuki, Hiroyasu Yamada, and Ágota Tóth (2000). "Maze-Solving by an Amoeboid Organism". In: *Nature* 407 (6803), p. 470. ISSN: 1476-4687. DOI: 10.1038/35035159. URL: https://doi.org/10.1038/35035159.

Olah, Chris et al. (2020). "Zoom In: An Introduction to Circuits". In: *Distill* 5.3. DOI: 10.23915/distill.00024.001.

Piaget, Jean (1950). *The Psychology of Intelligence*. Trans. by Malcolm Piercy and D. E. Berlyne. 1st ed. *Originally published in French as* La Psychologie de l'intelligence, *1947*. London: Routledge & Kegan Paul.

Polanyi, Michael (1966). *The Tacit Dimension*. Chicago: University of Chicago Press.

Prior, Helmut, Ariane Schwarz, and Onur Güntürkün (2008). "Mirror-Induced Behavior in the Magpie (*Pica pica*): Evidence of Self-Recognition". In: *PLoS Biology* 6.8, e202. DOI: 10.1371/journal.pbio.0060202. URL: https://doi.org/10.1371/journal.pbio.0060202.

Raby, C. R. et al. (2007). "Planning for the Future by Western Scrub-Jays". In: *Nature* 445 (7130), pp. 919–921. ISSN: 1476-4687. DOI: 10.1038/nature05575. URL: https://doi.org/10.1038/nature05575.

Riley, Jeff (2006a). "Evolving Fuzzy Rules for Goal-Scoring Behaviour in a Robot Soccer Environment". Available at https://www.praescientem.com.au/publications/PhDThesisRMIT.pdf. PhD thesis. Melbourne, Australia: RMIT University.

— (July 2006b). "The elusive promise of AI". In: *Ubiquity* 2006.July. DOI: 10.1145/1149633.1149638. URL: https://doi.org/10.1145/1149633.1149638.

— (Apr. 2021a). "The elusive promise of AI: a second look". In: *Ubiquity* 2021.April. DOI: 10.1145/3458742. URL: https://doi.org/10.1145/3458742.

— (June 2021b). "Will machines ever think like humans?" In: *Ubiquity* 2021.June. DOI: 10.1145/3459743. URL: https://doi.org/10.1145/3459743.

Rosenblatt, Frank (1958). "The perceptron: A probabilistic model for information storage and organization in the brain". In: *Psychological Review* 65.6, pp. 386–408. DOI: 10.1037/h0042519. URL: https://doi.org/10.1037/h0042519.

Rumelhart, David E. and James L. McClelland (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Two volumes. Cambridge, MA: MIT Press.

Russell, Stuart and Peter Norvig (1995). *Artificial Intelligence: A Modern Approach*. 1st ed. Prentice-Hall, Inc.

Ryle, Gilbert (1949). *The Concept of Mind*. London: Hutchinson's University Library.

Sagan, Carl (1980). *Cosmos*. New York: Random House.

Salovey, Peter and John D. Mayer (1990). "Emotional Intelligence". In: *Imagination, Cognition and Personality* 9.3, pp. 185–211. DOI: 10.2190/DUGG-P24E-52WK-6CD. URL: https://doi.org/10.2190/DUGG-P24E-52WK-6CD.

Searle, John R. (1980). "Minds, Brains, and Programs". In: *Behavioral and Brain Sciences* 3.3, pp. 417–424. DOI: 10.1017/S0140525X00005756.

Simard, Suzanne (May 2021). *Finding the Mother Tree: Discovering the Wisdom of the Forest*. New York: Knopf. ISBN: 978-0525656098.

Simard, Suzanne W. et al. (1997). "Net Transfer of Carbon between Ectomycorrhizal Tree Species in the Field". In: *Nature* 388 (6642), pp. 579–582. ISSN: 1476-4687. DOI: 10.1038/41557. URL: https://doi.org/10.1038/41557.

Sloman, Steven and Philip Fernbach (2018). *The Knowledge Illusion. Why We Never Think Alone*. New York: Riverhead Books.

Smolensky, Paul (1988). "On the Proper Treatment of Connectionism". In: *Behavioral and Brain Sciences* 11.1, pp. 1–23. DOI: 10.1017/S0140525X00052432.

Solomonoff, R.J. (1964a). "A formal theory of inductive inference. Part I". In: *Information and Control* 7.1, pp. 1–22. ISSN: 0019-9958. DOI: https://doi.org/10.1016/S0019-9958(64)90223-2. URL: https://www.sciencedirect.com/science/article/pii/S0019995864902232.

— (1964b). "A formal theory of inductive inference. Part II". In: *Information and Control* 7.2, pp. 224–254. ISSN: 0019-9958. DOI: https://doi.org/10.1016/S0019-9958(64)90131-7. URL: https://www.sciencedirect.com/science/article/pii/S0019995864901317.

Spearman, Charles (1904). ""General Intelligence," Objectively Determined and Measured". In: *The American Journal of Psychology* 15.2, pp. 201–293.

Sternberg, Robert J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge: Cambridge University Press.

Templeton, Adly et al. (2024). *Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet*. Anthropic research report. URL: https://transformer-circuits.pub/2024/scaling-monosemanticity/.

Tero, Atsushi et al. (2010). "Rules for Biologically Inspired Adaptive Network Design". In: *Science* 327.5964, pp. 439–442. DOI: 10.1126/science.1177894. eprint: https://www.science.org/doi/pdf/10.1126/science.1177894. URL: https://www.science.org/doi/abs/10.1126/science.1177894.

Turing, Alan M. (Oct. 1950). "Computing Machinery and Intelligence". In: *Mind* LIX.236, pp. 433–460. ISSN: 0026-4423. DOI: 10.1093/mind/LIX.236.433. eprint: https://academic.oup.com/mind/article-pdf/LIX/236/433/61209000/mind_lix_236_433.pdf. URL: https://doi.org/10.1093/mind/LIX.236.433.

Vallor, Shannon (June 2024). *The AI Mirror. How to Reclaim Our Humanity in an Age of Machine Thinking*. Oxford University Press. ISBN: 9780197759066. DOI: 10.1093/oso/9780197759066.003.0002. URL: https://doi.org/10.1093/oso/9780197759066.003.0002.

Varela, Francisco J., Evan Thompson, and Eleanor Rosch (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

von Frisch, Karl (1967). *The Dance Language and Orientation of Bees*. Trans. by Leigh E. Chadwick. Cambridge, MA: Harvard University Press.

von Neumann, John and Oskar Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Wechsler, David (1944). *The Measurement of Adult Intelligence*. 3rd ed. Baltimore: Williams & Wilkins.

Weizenbaum, Joseph (Jan. 1966). "ELIZA — a computer program for the study of natural language communication between man and machine". In: *Communications of the ACM* 9.1, pp. 36–45. ISSN: 0001-0782. DOI: 10.1145/365153.365168. URL: https://doi.org/10.1145/365153.365168.

— (1976). *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W. H. Freeman.

Zadeh, Lotfi A. (1965). "Fuzzy Sets". In: *Information and Control* 8.3, pp. 338–353. ISSN: 0019-9958. DOI: https://doi.org/10.1016/S0019-9958(65)90241-X. URL: https://www.sciencedirect.com/science/article/pii/S001999586590241X.