Distributed Cognition as Epistemic Infrastructure: A Taxonomy of Collective Epistemic Systems

Abstract

The concept of 'distributed cognition' is routinely invoked to unify heterogeneous collective epistemic systems, including prediction markets, open-source software development, deliberative bodies, digital platforms, and regulatory institutions. These systems are often treated as interchangeable instances of 'crowd wisdom', whose epistemic virtues are presumed to arise naturally from decentralisation and aggregation. This article argues that this assumption rests on a category error: it conflates epistemic coordination architectures with epistemic closure architectures and treats descriptive claims about cognitive distribution as if they entailed claims about epistemic reliability. Distributed cognition is not a unitary epistemic good but a family of structurally distinct socio-technical systems that perform different epistemic functions under different closure mechanisms, incentive structures, and governance-mediated conditions. The article develops a formal taxonomy of distributed epistemic systems and shows that each class distributes a different epistemic resource, stabilises closure in a different locus, and exhibits characteristic failure modes. Through a paradigm contrast between prediction markets and open-source communities, the article demonstrates that epistemic failures in distributed systems are governance-mediated distortions of closure and salience rather than merely cognitive accidents. Distributed cognition is therefore best understood as an engineered epistemic infrastructure whose reliability depends on architectural design rather than on decentralisation alone.

Keywords

distributed cognition; collective intelligence; epistemic governance; epistemic closure; prediction markets; open-source software; socio-technical systems; epistemic reliability

1. Introduction: Distributed Cognition as a Governance Problem

Over the past three decades, 'distributed cognition' has migrated from a technical concept in cognitive science into a rhetorical unifier for a wide range of collective epistemic practices. What began, in Hutchins' foundational work, as an empirical thesis about how cognitive tasks are realised across socio-technical systems has been progressively re-described as a general epistemic virtue of decentralisation itself (Hutchins 1995, chs 1–2, esp. pp. 3–18; 49–70). In contemporary discourse, prediction markets, open-source software development, digital platforms, deliberative bodies, and even regulatory institutions are increasingly treated as interchangeable instances of a single phenomenon: collective intelligence. The prevailing assumption is that dispersing cognition across many agents and artefacts reliably generates epistemic improvement, whether in the form of more accurate forecasts, better solutions, or more robust judgements.

This article challenges that assumption. It argues that 'distributed cognition' has become an inflationary concept—invoked to unify heterogeneous collective systems without attending to the institutional architectures that structure their epistemic performance. The dominant discourse treats distributed cognition as a natural epistemic good: a property that emerges spontaneously from decentralisation and aggregation. What is systematically missing from this picture is a governance analysis of collective epistemic systems that structure belief formation and closure. These systems are being celebrated, deployed, and institutionalised without a corresponding account of how incentives, power, contestability, and credibility allocation are organised within them. The result is a form of epistemic naïveté: a tendency to attribute epistemic success or failure to cognitive mechanisms alone, while ignoring the governance conditions that enable or distort those mechanisms.

The intellectual roots of distributed cognition lie in a careful, empirically grounded account of cognition as a socio-technical achievement. Hutchins' analysis of navigation teams in *Cognition in the Wild* explicitly rejects the idea that cognition is located in individual heads alone; instead, it is realised across coordinated agents, material artefacts, and representational systems (Hutchins 1995, chs 1–2, esp. pp. 3–18; 49–70). The cognitive system, on this view, includes not only human operators but also charts, instruments, procedures, and spatial arrangements that structure information flow and constrain error. Clark and Chalmers' extended mind thesis generalised this insight, arguing that cognitive processes can extend into the environment when external artefacts play a functionally integrated role in reasoning and memory (Clark and Chalmers 1998, pp. 7–12). These accounts were not normative claims about epistemic virtue. They were descriptive and functional theses about how cognitive work is actually accomplished in practice.

The subsequent transformation of distributed cognition into a normative ideal owes more to crowd epistemology than to cognitive science. Surowiecki's *The Wisdom of Crowds* popularised the idea that aggregation across diverse, independent agents can outperform individual expertise, provided certain conditions are met (Surowiecki 2004, Introduction, esp. pp. xiii–xvii). Sunstein's *Infotopia* refined this claim, arguing that many minds, appropriately structured, can produce superior knowledge outcomes in both markets and deliberative contexts (Sunstein 2006, esp. pp. 1–9; chs 2–3). These interventions were explicitly optimistic. They framed decentralisation, diversity, and aggregation as epistemic goods, capable of disciplining expert overconfidence and institutional inertia.

What neither strand of literature adequately confronts is the governance problem that arises when collective epistemic systems are treated as epistemically self-justifying. Hutchins and Clark provide no account of how socio-technical cognitive systems ought to be governed. Surowiecki and Sunstein provide only thin institutional criteria—independence, diversity, decentralisation—without analysing how those criteria are systematically undermined by power asymmetries, capital concentration, algorithmic mediation, or institutional self-interest. The result is a discourse that celebrates distributed cognition while remaining largely blind to the conditions under which it tends to fail.

Yet the systems routinely grouped under the label of distributed cognition are not epistemically homogeneous. Prediction markets aggregate probabilistic beliefs into a price signal, while open-source communities coordinate modular problem-solving through shared artefacts and informal authority hierarchies. Both are routinely described as instances of 'distributed cognition', yet they distribute different epistemic resources, operate through different aggregation mechanisms, and fail in systematically different ways. Markets are vulnerable to capital-weighted narrative capture and strategic manipulation. Open-source communities exhibit maintainer oligarchies, gatekeeping, and architectural stagnation. Digital platforms amplify salience rather than truth. Deliberative bodies are prone to conformity cascades and group polarisation. Institutions such as courts and regulators display epistemic closure and resistance to external critique. These pathologies are often described as contingent distortions of otherwise sound cognitive mechanisms.

This article rejects that diagnosis. It argues that these failures are more plausibly understood as governance-mediated rather than as purely cognitive accidents. Distributed epistemic systems are governed as if they exercised no distinctive form of epistemic control, and therefore bore no corresponding governance responsibilities. In practice, such systems routinely determine what counts as salient information, which forecasts are treated as credible, which solutions are adopted, which dissenting views are marginalised, and which errors are acknowledged or concealed. They structure belief formation and closure at scale. That structuring is not epistemically neutral in practice. It is mediated by capital, code, procedural control, reputational hierarchies, and institutional incentives.

The conceptual error at the heart of contemporary distributed-cognition discourse is therefore twofold. First, it treats heterogeneous collective systems as instances of a single phenomenon, despite the fact that they distribute different epistemic resources—belief, labour, judgement, attention—through radically different aggregation mechanisms—prices, artefacts, discourse, algorithms, hierarchical synthesis. Secondly, it attributes epistemic virtue to decentralisation itself, rather than to the governance conditions under which decentralised systems can function as reliable epistemic infrastructures.

This article responds to that error in three moves. It first develops a formal taxonomy of distributed epistemic systems, distinguishing belief-aggregation systems (such as prediction markets), problem-solving production systems (such as open-source software), deliberative synthesis systems (such as juries and peer review), signal-extraction platforms (such as Wikipedia and social media), and institutional epistemic governance systems (such as courts and regulators). The purpose of this taxonomy is not classificatory elegance but architectural clarity: to show that different systems perform different epistemic functions under different incentive and closure structures, and therefore fail in systematically different ways.

Secondly, it uses prediction markets and open-source communities as paradigm contrasts to demonstrate that distributed epistemic systems can be both epistemically powerful and governance-fragile. Markets aggregate belief through capital-weighted price formation and are structurally vulnerable to narrative capture under conditions of thin liquidity and strategic participation. Open-source systems distribute cognitive labour through artefact-centred coordination but tend towards informal authority concentration and epistemic closure under conditions of architectural path dependence. These cases illustrate why distributed cognition cannot be treated as a unitary epistemic good.

Thirdly, it argues that epistemic failures in distributed systems are governance-mediated pathologies. Incentive misalignment, authority concentration, and weak contestability systematically distort aggregation mechanisms that are otherwise cognitively sound. The implication is that distributed cognition is best understood not as a natural epistemic virtue, but as an engineered epistemic infrastructure whose reliability depends on governance architecture rather than on decentralisation alone.

Methodologically, this article develops a theoretical and structural analysis of distributed epistemic systems as socio-technical infrastructures whose epistemic performance is governance-mediated. It does not deny the empirical successes of collective epistemic systems. Prediction markets often outperform individual experts. Open-source development has produced some of the most robust software infrastructures in existence. Platforms can surface relevant information at unprecedented scale. Deliberative bodies sometimes correct individual bias. The critique advanced here is not that distributed cognition fails everywhere or always. It is that its successes are contingent on governance conditions that are rarely theorised, rarely protected, and increasingly undermined.

Distributed cognition, in short, is not a miracle. It is an engineered epistemic infrastructure. Like all infrastructures, it requires governance. Without attention to how incentives, power, and contestability are organised within it, collective epistemic systems do not merely risk failure. They systematically tend to generate it.

2. Why 'Distributed Cognition' Is Not a Unitary Concept

Contemporary invocations of 'distributed cognition' conceal a basic conceptual conflation. The term is routinely used to describe at least three distinct phenomena: (i) the functional extension of individual cognition into the environment; (ii) the coordination of cognitive labour across multiple agents within task systems; and (iii) the operation of institutional systems that aggregate, filter, rank, or otherwise synthesise information at scale. These phenomena are regularly collapsed into a single category and treated as instances of 'collective intelligence'. This section argues that the collapse is analytically misleading. It obscures decisive differences in what is being distributed, how epistemic work is performed, and where epistemic closure is produced. It also licenses an unearned rhetorical slippage from 'distributed' to 'democratic' that is not warranted by either cognitive science or social epistemology.

From a social-epistemological standpoint, the decisive question is not where cognition is functionally realised, but whether a collective epistemic practice is reliable under realistic social conditions. Goldman's veritistic framework evaluates social-epistemic systems by their propensity to promote true belief rather than by their architectural form alone (Goldman 1999, ch 3; cf. ch 5 for application to argumentation). Reliability is not a function of architectural form alone. It depends on how information is filtered, how incentives are structured, how dissent is handled, and how error-correction operates over time. Cognitive-science models of distribution and extension, important though they are, do not by themselves supply criteria for assessing the epistemic performance of institutional-scale epistemic arrangements.

The original theoretical locus of distributed cognition is Hutchins' empirical analysis of navigation teams in *Cognition in the Wild*. Hutchins' central claim is that cognitive processes are not confined to individual minds but are realised across coordinated systems of agents, artefacts, and representational structures (Hutchins 1995, chs 1–2, esp. pp. 3–18; 49–70). On his account, the navigation team constitutes a single cognitive system whose problem-solving capacity depends on patterned interaction among human operators and external artefacts such as charts, instruments, and procedural checklists. Two features of this model are salient. First, distributed cognition is not spontaneous: it presupposes dense coordination artefacts and procedural scaffolding that structure information flow and constrain error. Secondly, epistemic roles within the system are differentiated. Some participants generate information, others transform it, others record it, and others commit the system to action. The system's epistemic performance depends not merely on decentralisation, but on disciplined role differentiation and institutionalised routines.

Nothing in Hutchins' account entails that distributed cognition is, as such, epistemically reliable in the sense required by social epistemology. Hutchins provides a descriptive and functional analysis of task performance, not an evaluative theory of how collective epistemic systems should be assessed across social contexts. The navigation team is not a paradigm of democratic epistemic organisation. It is a paradigm of disciplined epistemic coordination. To abstract away from these scaffolds and treat 'distribution' itself as the operative epistemic virtue is to invert the lesson of the case study.

A parallel ambiguity afflicts the extended mind thesis. Clark's early work, and his joint paper with Chalmers, argue that cognitive processes can extend into the environment when external artefacts are functionally integrated into reasoning and memory (Clark 1997, ch 3, esp. §§3.1–3.3; Clark and Chalmers 1998, pp. 7–12). On this view, notebooks, calculators, and digital devices can become parts of an individual's cognitive system when they reliably perform functions that would otherwise be carried out internally. The core insight is again descriptive: cognition is not bounded by the skull but is realised across brain–body–world couplings.

However, the extended mind thesis concerns the functional architecture of individual cognition, not the epistemic reliability of institutional systems. Clark's canonical examples are primarily dyadic: a single agent coupled to an external artefact. Even when social interaction is involved, the analytic focus remains on how external resources scaffold an individual's reasoning, not on how institutional systems allocate epistemic roles, shape information environments, or govern collective belief formation (Clark 1997, ch 3, esp. §§3.1–3.3; Wilson and Clark 2009, pp. 58–65). Extended cognition therefore does not warrant treating markets, platforms, or regulatory bodies as epistemic agents in their own right, still less treating institutional outputs as reliable simply because they emerge from distributed inputs. Epistemic reliability does not follow from extended cognition.

Taken together, Hutchins and Clark establish two points. First, cognition can be functionally distributed across artefacts and agents. Secondly, such distribution typically presupposes scaffolding—procedures, artefacts, and role differentiation—that structures information flow and constrains error. What neither establishes is a criterion for assessing the epistemic reliability of institutional-scale epistemic systems. Cognitive distribution tells us where and how cognitive work is done; it does not tell us whether a given socio-technical arrangement will track truth, resist distortion, or remain robust under power asymmetries and strategic behaviour.

At this juncture, the social-epistemological stakes of closure become visible. Recent work on public epistemic authority converges on the point that what counts as fact, salience, and admissible interpretation is not stabilised by individual cognition or aggregate belief alone, but by durable socio-technical infrastructures that structure epistemic environments over time. Neame's analysis of factual authority reconstructs epistemic closure as an infrastructural achievement: public knowledge is stabilised through inherited institutional forms, material artefacts, and procedural regimes that mediate credibility, visibility, and interpretive uptake (Neame 2025, esp. pp. 416–426; 436–442). On this view, epistemic authority is not an emergent property of aggregation but a property of the systems that select, rank, and stabilise outputs. Closure is therefore not a cognitive endpoint but an architectural function of epistemic infrastructures.

This distinction becomes decisive once one moves from coordination systems to institutional systems that stabilise epistemic outputs. In task systems such as Hutchins' navigation team, cognitive labour is distributed, but closure is generated through procedural roles. In extended cognition, artefacts scaffold reasoning, but they do not themselves occupy epistemic roles in the sense of selecting, ranking, or filtering what counts as the system's output. By contrast, in institutional systems—prediction markets, open-source communities, platforms, peer-review regimes, regulatory organisations—closure is precisely what is at stake. These systems do not merely distribute cognitive labour. They determine salience, set thresholds for acceptance, select which contributions become canonical, and stabilise outputs in ways that shape downstream belief and action.

Conflating these architectures under a single label encourages a persistent analytic error: the rhetorical slippage from 'distributed' to 'democratic'. Because distributed cognition is associated with plurality, decentralisation, and diversity, it is tacitly treated as democratising. Collective epistemic systems are described as normatively superior simply because they are not overtly top-down. Yet nothing in either Hutchins' or Clark's theoretical frameworks supports that inference. A system can distribute contributions widely while concentrating closure in a small set of gatekeeping roles. It can aggregate signals while privileging capital, code, or institutional priorities. It can appear participatory while systematically filtering dissent. 'Distributed' describes an architecture; 'democratic' is an evaluative predicate. The inference from one to the other is not licensed by cognitive-science premises.

The crowd-epistemology literature exemplifies this slippage. Surowiecki's and Sunstein's optimism about many minds is often presented as if it naturally follows from insights about distributed cognition (Surowiecki 2004, Introduction, esp. pp. xiii–xvii; Sunstein 2006, esp. pp. 1–9). The implicit suggestion is that if cognition is widely distributed across populations and artefacts, then collective epistemic systems that aggregate inputs across many people will tend to be superior, or at least reliably self-correcting. Yet the move from descriptive functionalism to institutional optimism requires more argument than is typically supplied. Independence, diversity, and decentralisation are not self-stabilising properties. They are design-sensitive and routinely undermined by incentives, social influence, and structural asymmetries.

Goldman's reliability framework clarifies the point. Epistemic systems must be evaluated by their truth-tracking performance under realistic social conditions, not by the mere fact that many participants contribute inputs (Goldman 1999, ch 3; cf. ch 5 for application to argumentation). A system that aggregates many perspectives is not thereby reliable. Its performance depends on how information is filtered, how aggregation is weighted, what constitutes admissible evidence, how dissent is handled, and how errors are recognised and corrected. Distributed cognition, in the cognitive-science sense, therefore does not entail distributed epistemic virtue in the sense of social epistemology.

The upshot is that 'distributed cognition' is not a unitary concept. It names a family of structurally distinct phenomena that perform different epistemic functions under different governance conditions. To treat these phenomena as instances of a single epistemic kind is best understood as a category conflation: the conflation of extended cognition (individual-plus-artefact systems), coordinated task cognition (multi-agent systems with procedural scaffolding), and institutional epistemic infrastructures (systems that aggregate inputs into stabilised outputs).

The purpose of the taxonomy developed in the next section is to correct this conflation. By distinguishing belief-aggregation systems, problem-solving production systems, deliberative synthesis systems, signal-extraction platforms, and institutional epistemic governance systems, the taxonomy makes explicit what is being distributed in each case, how aggregation occurs, and where closure is produced. The point is not classificatory refinement for its own sake. It is to show that epistemic reliability cannot be inferred from distribution alone. It must be assessed in relation to the governance architecture of the system in question.

3. A Taxonomy of Distributed Epistemic Systems

The preceding sections have argued that 'distributed cognition' is not a unitary epistemic phenomenon, but a family of structurally distinct socio-technical architectures that perform different epistemic functions under different governance conditions. The purpose of this section is to make that claim analytically precise by introducing a formal taxonomy of distributed epistemic systems. The taxonomy is not offered as a classificatory exercise for its own sake. It is a diagnostic instrument designed to expose the category conflations that underwrite contemporary crowd epistemology and to show why treating heterogeneous collective systems as interchangeable instances of 'collective intelligence' is a conceptual error.

A closely parallel methodological problem has been identified in adjacent literatures on participatory knowledge practices. Chastin et al. show that contemporary discourse on 'co-creation' exhibits the same inflationary tendency: heterogeneous epistemic activities are rhetorically unified under a single participatory label and treated as epistemically virtuous in virtue of decentralisation alone. Their analysis argues that co-creation is not an epistemic good as such, but a family of governance-sensitive architectures whose epistemic performance depends on procedural design, role differentiation, and institutional scaffolding (Chastin et al. 2025, esp. pp. 2–4; 5–6). The conceptual move is formally identical to the one required here. Just as participatory research practices cannot be evaluated independently of their governance architecture, distributed epistemic systems cannot be evaluated independently of the mechanisms through which closure is produced and error is corrected. The taxonomy developed below should therefore be understood as part of a broader methodological correction against architectural flattening across domains of collective knowledge production.

The taxonomy is grounded in three core distinctions. First, distributed epistemic systems differ in the epistemic resource they distribute: probabilistic belief, cognitive labour, perspective and judgement, micro-contribution and attention, or delegated expert judgement. Secondly, they differ in their aggregation mechanism: price formation, artefact-centred coordination, discursive convergence under procedure, algorithmic ranking and visibility metrics, or hierarchical synthesis and procedural filtering. Thirdly, they differ in their closure locus: the site at which distributed inputs are stabilised into an output treated as settled for purposes of action. Systems that differ across all three dimensions are not variants of a single epistemic kind. They are distinct epistemic architectures.

The taxonomy therefore maps five analytic dimensions: (i) what epistemic resource is distributed; (ii) how aggregation occurs; (iii) how incentives are aligned; (iv) where closure is produced; and (v) how epistemic failure tends to manifest. The five classes introduced below are ideal types. Real-world institutions often hybridise them. The point is not to deny hybridity but to prevent conceptual flattening.

1. Belief-Aggregation Systems (Prediction Markets)

Distributed resource:	probabilistic belief.
Aggregation mechanism:	price formation.
Incentives:	financial.
Closure locus:	market price.
Failure mode:	capital-weighted narrative capture.

Belief-aggregation systems distribute private beliefs across a population of agents and aggregate them into a forecast signal. Prediction markets are the paradigm case. Participants buy and sell contracts contingent on future events, and the market price is interpreted as a collective probability estimate.

What is distinctive here is the closure locus. The price functions as the epistemic output by closing uncertainty into an action-guiding signal. The system's epistemic service consists in converting dispersed information into a continuously updated forecast.

A characteristic failure mode arises under thin liquidity, capital asymmetry, or strategic participation. The price signal can be distorted by coordinated trading or by concentrated capacity to withstand volatility. In such contexts, the market ceases to aggregate dispersed information and instead amplifies salient narratives or strategic positioning. As Surowiecki himself notes, aggregation reliability depends on independence and resistance to herding; correlated beliefs and imitation dynamics systematically degrade collective accuracy (Surowiecki 2004, ch 2, esp. pp. 41–65). Sunstein further shows how informational cascades and reflexive deference to price signals can anchor dominant narratives even when private information remains heterogeneous (Sunstein 2006, esp. chs 2–3).

2. Problem-Solving Production Systems (Open Source)

Distributed resource:	cognitive labour.
Aggregation mechanism:	artefact-centred coordination.
Incentives:	reputational, intrinsic, career-based.

Closure locus:	maintainer merge decisions.
Failure mode:	oligarchy and architectural stagnation.

Problem-solving production systems distribute cognitive labour rather than belief. Open-source software development is the paradigm case. Contributors propose code and fixes, which are integrated into a shared artefact through maintainership and version control.

Aggregation occurs through modular decomposition and artefact-centred coordination. The evolving codebase functions as an externalised working memory and constraint system.

The closure locus is informal and maintainer-centred. Closure is achieved through merge decisions and architectural direction. Core maintainers determine what becomes canonical.

A characteristic governance-mediated failure mode is the concentration of informal integration authority and the emergence of path-dependent architectural trajectories. In Raymond's account of open-source 'homesteading', founders acquire de facto control over projects through Lockean property norms, reputation accumulation, and continuity of stewardship, while dominant 'category killer' projects suppress rival initiatives by attracting extensions rather than competition. Over time, this produces strong path dependence and informal authority structures that stabilise ownership, control, and integration rights independently of formal governance (Raymond 2001, esp. 'Homesteading the Noosphere', pp. 88–92; 105–106; cf. Benkler 2006, esp. pp. 60–63; 100–104; 375–379; 445–447).

3. Deliberative Synthesis Systems (Juries, Peer Review)

Distributed resource:	perspective and judgement.
Aggregation mechanism:	discursive convergence under procedure.
Incentives:	civic, professional.
Closure locus:	procedural verdicts or acceptance decisions.
Failure mode:	conformity cascades and group polarisation.

Deliberative synthesis systems distribute perspective rather than belief or labour. Juries, peer review, and citizen assemblies are paradigm cases. Participants deliberate over evidence and arguments and converge on a collective judgement.

Aggregation occurs through discursive convergence under procedural constraints. Closure is achieved through verdicts or acceptance decisions treated as settled because they result from an established process.

A characteristic failure mode is conformity pressure leading to polarisation and convergence on rhetorically salient positions rather than epistemically superior ones (Sunstein 2002, esp. pp. 178–184). Procedural regularity does not guarantee epistemic reliability under asymmetrical expertise or reputational pressure.

4. Signal-Extraction Platforms (Wikipedia, Social Media)

Distributed resource:	micro-contribution and attention.
Aggregation mechanism:	algorithmic ranking and visibility metrics.

Incentives:	visibility and status.
Closure locus:	algorithmic salience ranking.
Failure mode:	epistemic drift.

Signal-extraction platforms distribute micro-contributions and attention. Users contribute fragments of content; algorithms determine which content becomes salient.

Aggregation occurs through ranking and visibility metrics. Closure is achieved through salience: what is treated as epistemically relevant is what is made prominent.

A characteristic failure mode is epistemic drift: the divergence between salience and epistemic merit. Popularity substitutes for truth, coordinated brigading distorts visibility, and marginalised perspectives are structurally suppressed (Gillespie 2018, esp. chs 1, 6–7).

5. Institutional Epistemic Governance Systems (Courts, Regulators)

Distributed resource:	delegated judgement.
Aggregation mechanism:	hierarchical synthesis and procedural filtering.
Incentives:	reputational, career-based, political.
Closure locus:	binding decisions and settled interpretations.
Failure mode:	epistemic closure and self-protection.

Institutional epistemic governance systems distribute delegated judgement. Courts and regulators synthesise evidence into authoritative interpretations.

Aggregation occurs through hierarchical synthesis and procedural filtering. Closure is formal and legally constituted.

A characteristic failure mode is epistemic closure and resistance to external critique. Institutions simplify reality in order to render it administratively legible and then defend those simplifications against inconvenient evidence (Scott 1998, Introduction, esp. pp. 1–7; ch 1, esp. pp. 11–24; cf. ch 9, esp. pp. 309–318).

Comparative Summary

The taxonomy distinguishes infrastructural, institutional, and hybrid epistemic systems along axes of agency distribution, normativity, and epistemic authority (see Table 3.1).

Analytic Payoff

The analytic payoff of this taxonomy is threefold. First, it shows that different collective systems perform different epistemic functions—forecasting, artefact production, judgement synthesis, salience ranking, interpretive closure—and therefore cannot be evaluated by a single theory of ‘crowd wisdom’. Secondly, it makes explicit that epistemic control is not uniformly distributed. In some systems closure is implicit; in others it is informal; in others it is procedural; in others it is algorithmic; in others it is formal. Thirdly, it reveals that failure modes are structurally specific. Each system tends to fail in ways predictable from its incentive structure and closure locus.

Two caveats are in order. First, this taxonomy is not exhaustive. Hybrid systems exist. Secondly, it is ideal-typical. Real-world systems often instantiate multiple classes simultaneously.

The point of the taxonomy is not classificatory tidiness. It is to discipline evaluation. It makes clear why treating heterogeneous collective systems as interchangeable instances of ‘collective intelligence’ is a category conflation. Epistemic reliability cannot be inferred from distribution alone. It must be assessed in relation to the governance architecture of the system in question.

System Class	Distributed Resource	Aggregation Mechanism	Incentives	Authority Locus	Characteristic Failure Mode
Belief-Aggregation	Probabilistic belief	Price formation	Financial	Implicit, capital-weighted (price as closure)	Narrative capture under capital asymmetry
Problem-Solving Production	Cognitive labour	Artefact-centred coordination	Reputational, intrinsic	Informal, maintainer-centred (merge as closure)	Oligarchy and stagnation under path dependence
Deliberative Synthesis	Perspective, judgement	Discursive convergence under procedure	Civic, professional	Procedural, role-defined (verdict/ acceptance as closure)	Polarisation and conformity cascades
Signal-Extraction Platforms	Micro-contribution, attention	Algorithmic ranking/ visibility metrics	Visibility, status	Algorithmic, opaque (ranking as salience closure)	Epistemic drift and salience distortion
Institutional Governance	Delegated judgement	Hierarchical synthesis/procedural filtering	Career, political	Formal, legal (binding decisions as closure)	Epistemic closure and self-protection

Table 3.1 — Taxonomy of Distributed Cognitive Systems

4. Prediction Markets and Open Source as Paradigm Contrasts

The taxonomy in §3 shows that distributed epistemic systems differ not merely by domain or scale, but by the epistemic function they perform and the architecture through which they stabilise outputs. This section sharpens that claim by examining two flagship cases that are routinely cited as paradigms of epistemically decentralised ‘collective intelligence’: prediction markets and open-source software development. These cases are instructive because they are often grouped together under a single rhetorical banner of ‘crowd wisdom’, despite exhibiting structurally distinct epistemic architectures. The aim is to show, first, that prediction markets and open-source systems perform different epistemic functions—belief aggregation versus problem-solving production—and secondly, that both are epistemically powerful yet governance-sensitive.

A brief clarification will keep the analysis disciplined. ‘Governance fragility’ here refers to the sensitivity of a system’s epistemic performance to background conditions such as participation structure, concentration of control, incentive alignment, and contestability. It does not imply that these systems are generally defective. The question is not whether these systems can work—they often do—but which architectural features make their epistemic outputs robust, and which features expose them to predictable failure modes.

4.1 Prediction Markets: Belief Aggregation Through Price

Distributed resource:	probabilistic belief.
Aggregation mechanism:	price formation.
Incentives:	financial.
Closure locus:	market price.
Characteristic fragility:	capital-weighted narrative capture.

Prediction markets are the canonical belief-aggregation system. Participants trade contracts contingent on future events, and the prevailing price is interpreted as a collective probability estimate. For well-specified events with verifiable outcomes and sufficiently liquid participation, prediction markets can perform competitively with, and sometimes outperform, individual experts and survey-based forecasts (Wolfers and Zitzewitz 2004, pp. 112–118).

The core epistemic mechanism is price formation. Traders bring heterogeneous priors and information fragments into the market. Those who believe the current price understates the likelihood of an event buy; those who believe it overstates that likelihood sell. In stylised designs, participants are rewarded for accuracy and penalised for error. This can filter out costless assertion because expressing a view requires exposure to loss (Hanson 2013, esp. pp. 359–362). In this limited sense, prediction markets externalise private belief into a public signal and update that signal as information changes.

The locus of epistemic authority in a prediction market is the price signal itself. The price closes uncertainty into an action-guiding estimate and is therefore treated as the authoritative epistemic output for purposes of coordination and decision-making. However, price does not remain an epistemically neutral aggregator once payoff asymmetry is introduced. Formal analysis shows that the interaction between dispersed information and asymmetric asset payoffs generates a systematic updating wedge and a risk-neutral distortion that biases equilibrium prices away from underlying dividend expectations, even in the absence of manipulation or strategic distortion (Albagli, Hellwig and Tsyvinski 2024, esp. pp. 2717–2724; 2726–2728; 2731–2732).

The epistemic performance of this mechanism is governance-sensitive. The closure locus—price—both aggregates belief and shapes belief. Participants rationally treat the price as information, producing reflexive dynamics in which early price movements anchor subsequent behaviour. Sunstein shows that information markets are vulnerable to cascades and manipulation, and that price functions not merely as an output but also as an input that reconfigures participant behaviour (Sunstein 2006, esp. ch 4; cf. chs 2–3).

Experimental evidence reinforces this governance sensitivity. Chen, Fine and Huberman show that even well-designed decentralised information markets systematically misaggregate once public signals induce correlated beliefs, producing double-counting effects and false probability peaks (Chen, Fine and Huberman 2004, pp. 985–987; 991–992). Their results further demonstrate that governance intervention in mechanism design—through a two-stage aggregation architecture and explicit public-signal separation protocols (GPIC, PPIC, SPIC)—is required to restore epistemic reliability (Chen, Fine and Huberman 2004, pp. 984–989; 991–993). This supports the mediation thesis of this section: aggregation alone does not guarantee epistemic adequacy; reliability depends on governance-sensitive design features that discipline how closure is produced through price.

This sensitivity becomes salient under conditions of thin liquidity, concentrated participation, or strategically motivated trading. Because the signal is generated through capital commitment, it is structurally weighted by the ability to deploy resources and withstand volatility. This does not imply that markets are generally unreliable. It implies that their epistemic output is produced through a capital-weighted closure mechanism that can be distorted when contestability is weak. Under such conditions, the market may cease to function primarily as an information aggregator and instead begin to amplify narrative salience or strategic positioning. The price then becomes less a stabilised estimate of probability and more a reflection of which actors can sustain positions through uncertainty.

The key point is architectural. Prediction markets distribute epistemic voice through a mechanism that translates belief into price via capital exposure. That design can be instrumentally valuable. It also introduces governance-mediated epistemic sensitivities: the stronger the incentive to move the signal—because the signal is consequential or institutionally relied upon—the stronger the incentive to attempt distortion, and the more the market's epistemic performance depends on participation conditions and market design. If epistemic performance were a spontaneous consequence of decentralisation, design would be marginal. In practice it is central.

4.2 Open Source: Artefact-Centred Cognition and Modular Labour

Distributed resource:	cognitive labour.
Aggregation mechanism:	artefact-centred coordination.
Incentives:	reputational, intrinsic, career-based.
Closure locus:	maintainer merge decisions.
Characteristic fragility:	oligarchy and architectural stagnation.

Open-source software development exemplifies a different distributed epistemic architecture. It does not aggregate belief into a scalar signal. It distributes cognitive labour across a population of contributors and coordinates that labour through shared artefacts, procedural norms, and informal role structures. The epistemic promise of open source lies in its capacity to harness dispersed technical competence and local problem knowledge.

Two features are structurally central. First, the system is artefact-centred. The evolving codebase functions as a shared coordination medium: it stabilises partial solutions, exposes incompatibilities, and allows contributors to align work through iterative integration rather than continuous deliberation. Secondly, the system is modular. Work can be decomposed into relatively independent tasks, enabling distributed contributions without comprehensive central planning. These architectural features help explain why open-source systems have produced robust technical infrastructures across multiple domains, as described in Raymond's contrast between cathedral-style development and bazaar-style parallel tinkering (Raymond 2001, esp. 'The Cathedral and the Bazaar', pp. 23–30; 'Release Early, Release Often', pp. 45–50; cf. Benkler 2006, esp. pp. 60–63; 90–94; 100–104; 375–379).

The closure locus in open-source systems is informal and maintainer-centred. Closure is achieved through merge decisions, release schedules, architectural direction, and vetoes over what counts as an admissible contribution. While contribution can be highly decentralised, integration cannot be. Absent some closure mechanism, projects fragment into incompatible forks. In this sense, concentrated closure is an architectural necessity rather than a moral defect.

The governance sensitivity arises from the way this closure is organised and sustained. Because maintainer roles often emerge informally and are reinforced by reputation and accumulated contextual knowledge, the locus of epistemic control can become insulated from contestation. Over time, and especially in mature projects with high downstream dependency, this can generate characteristic failure modes: maintainer oligarchy, gatekeeping, and architectural stagnation under path dependence. Benkler explicitly notes the vulnerability of peer production to leadership domination and governance capture (Benkler 2006, esp. pp. 90–94; 375–379), and Kelty shows how governance in open-source communities is frequently performed through tacit norms and informal authority structures that are difficult to render transparent or contestable (Kelty 2008, esp. pp. 196–204; 217–224).

The epistemic risk is not that maintainers are generally incompetent or ill-intentioned. It is that closure through informal gatekeeping can systematically filter which signals the system treats as admissible—whether technical concerns, design alternatives, or warnings about architectural risk. When authority becomes concentrated and pathways for contestation narrow, the system's capacity to exploit distributed expertise can be reduced. A production architecture that is highly generative under conditions of plural contribution can become comparatively conservative under conditions of entrenched maintainership and high dependency. The resulting epistemic sensitivity is therefore governance-mediated: it is a function of how roles, incentives, and closure are structured, not of distributed contribution as such.

4.3 Structural Divergence and the Assumption of Democratic Distribution

The contrast between prediction markets and open-source systems exposes the conceptual error of treating them as interchangeable instances of distributed cognition. They distribute different epistemic resources—belief versus labour. They aggregate in different ways—price formation versus artefact-centred coordination. They stabilise outputs through different closure loci—capital-weighted price signals versus maintainer-centred integration. They exhibit different characteristic fragilities—distortion through manipulation and reflexivity versus closure through informal oligarchy and path dependence. These are differences in epistemic function and governance architecture, not superficial differences in implementation.

What unites the cases is therefore not democratic epistemic participation but the production of stabilised collective outputs through socio-technical mechanisms that are often treated as self-stabilising. Both systems generate outputs that guide downstream belief and action. Both are often described as 'what the crowd thinks' or 'what the community built'. Yet in each case the output is shaped by how closure is organised: by capital-weighted trading in markets, and by discretionary integration in open source. This makes the rhetorical slippage from 'distributed' to 'democratic' unsupported. Markets weight epistemic voice by capital exposure. Open source concentrates closure in maintainer discretion. Both can be described as epistemically decentralised in participation; neither distributes closure on equal terms.

4.4 Non-Universal Conclusions

It would be a mistake to infer from these structural observations that either system is epistemically defective as such. Both instantiate genuine epistemic achievements. Both solve real coordination problems. Both can outperform traditional alternatives in many contexts. The point is that their successes are contingent on governance conditions that are rarely theorised and rarely secured as the systems scale and become consequential.

Belief aggregation through price is most epistemically useful under conditions of diverse participation, sufficient liquidity, and high costs of manipulation, and where market prices are treated as fallible indicators rather than as self-validating signals (Wolfers and Zitzewitz 2004, esp. pp. 112–118). Artefact-centred coordination is most epistemically generative where closure roles are transparent in practice, contestation channels remain viable, and maintainership remains oriented towards correction rather than reputational consolidation, as suggested by Benkler's analysis of governance fragility and leadership capture in peer production (Benkler 2006, esp. pp. 60–63; 90–94; 375–379) and Kelty's account of tacit governance and informal authority in open-source communities (Kelty 2008, esp. pp. 196–204; 217–224). These observations are not offered as a design manual. They establish the narrower point needed for the paper's central argument: distributed cognition is not an epistemic miracle. It is an engineered epistemic infrastructure whose epistemic performance depends on governance architecture.

Prediction markets and open source are therefore not exceptions to the governance problem. They are its clearest illustrations.

5. Failure Modes as Governance-Mediated Epistemic Failure Modes

The preceding sections have argued that distributed epistemic systems differ structurally in the epistemic functions they perform, the resources they distribute, and the loci through which epistemic closure is stabilised. This section consolidates that architecture by mapping characteristic epistemic failure modes across the five system classes identified in §3. The central claim is mediating rather than monocausal. Cognitive limitations, social dynamics, reputational incentives, and strategic behaviour do not generate epistemic distortion ex nihilo. Governance architectures shape how these familiar tendencies scale into stable distortions of salience, credibility, and closure.

By 'governance-mediated epistemic failure modes', this section refers to systematically recurrent patterns of epistemic distortion that arise when closure mechanisms, incentive gradients, and contestation pathways are structured in ways that convert ordinary human and organisational tendencies into durable epistemic harm. The mapping is offered as diagnostic and ideal-typical rather than exhaustive. It does not deny empirical successes or multi-causality. It identifies governance as a recurrent mediating structure that determines whether known dynamics remain episodic malfunctions or become stable features of collective epistemic systems.

5.1 Markets: Narrative Capture under Capital Asymmetry

In belief-aggregation systems, the epistemic function is the aggregation of dispersed probabilistic belief into a public forecast signal. The closure locus is the market price. A characteristic governance-mediated failure mode is narrative capture under capital asymmetry.

Sunstein's account of group polarisation shows how individuals rationally defer to salient social signals and converge on dominant positions even when private information remains heterogeneous (Sunstein

2002, esp. pp. 176–184; 185–188; cf. Sunstein 2006, chs 2–3). In market contexts, the analogue of a social signal is the price. Because participants treat the price itself as information, early movements—whether information-driven or strategically induced—can anchor subsequent trading and belief formation. The mechanism is reflexive: the signal both aggregates belief and reshapes belief.

The governance mediation occurs through the capital-weighted nature of closure. Because epistemic closure is stabilised through price, and because price formation depends on the capacity to sustain positions under uncertainty, actors with greater capital or risk tolerance can exert disproportionate influence over the epistemic output. Under conditions of thin liquidity, concentrated participation, or weak safeguards against coordination, the price signal can drift from aggregating dispersed information to amplifying the most capital-supported narrative. Contestability is reduced when early moves are costly to challenge, and error correction becomes institutionally expensive when deviation from the prevailing signal requires sustained capital exposure.

Formal analysis shows that this distortion is not merely strategic or behavioural. When asset payoffs are asymmetric, the interaction between dispersed information and payoff heterogeneity generates a systematic updating wedge and a risk-neutral distortion that biases equilibrium prices away from underlying dividend expectations. Aggregation becomes structurally biased because belief heterogeneity interacts with asymmetric exposure to weight closure toward payoff-salient narratives rather than toward probabilistic accuracy (Albagli, Hellwig and Tsyvinski 2024, esp. pp. 2717–2724; 2726–2728; 2731–2732).

In such settings, epistemic closure shifts from belief aggregation to payoff-weighted narrative dominance. What stabilises as the market's 'collective probability' increasingly reflects differential exposure rather than differential information.

Because epistemic closure in markets is capital-weighted and reflexive, governance-mediated failure in belief-aggregation systems tends to be inflationary rather than symmetric, with optimistic narratives persisting longer than warranted by underlying information and sceptical positions exiting earlier under liquidity and risk constraints. Error correction therefore occurs late and discontinuously, not because countervailing information is absent, but because contesting an established price signal becomes institutionally expensive as salience and capital commitment reinforce one another.

This dynamic is not confined to strategic manipulation or thin markets. Experimental evidence shows that even well-designed decentralised information markets systematically misaggregate under correlated public signals. Chen, Fine and Huberman demonstrate that when participants share public information, decentralised price formation amplifies correlation rather than cancelling it, producing persistent distortions away from underlying private information (Chen, Fine and Huberman 2004, pp. 985–987; 991–992).

Crucially, their results show that governance-layer interventions in mechanism design—such as two-stage aggregation procedures and signal-separation protocols—are required to restore epistemic reliability. Decentralisation alone does not self-correct correlated-belief distortion; governance design determines whether cascades remain episodic or become durable epistemic pathologies.

This is not a claim of universal failure. It is a claim of structurally recurrent risk. Where markets operate as epistemic instruments without governance mechanisms sufficient to secure participation diversity, liquidity, independence, and signal separation, narrative capture becomes a predictable distortion. The failure mode arises not because cascades exist, but because capital-weighted closure and payoff asymmetry convert cascade dynamics into stable epistemic control over salience.

5.2 Open Source: Maintainer Oligarchy and Epistemic Closure

In problem-solving production systems, the epistemic function is the coordination of distributed cognitive labour through shared artefacts. The closure locus is maintainer integration: merge decisions, release schedules, and architectural direction. A characteristic governance-mediated failure mode is maintainer oligarchy accompanied by epistemic closure.

Open-source systems require a closure mechanism. Absent some authority to integrate contributions, projects fragment into incompatible forks. Closure through maintainership is therefore an architectural necessity rather than a moral defect. The governance sensitivity arises from how this closure is structured and sustained over time.

Because maintainer roles often emerge informally and are reinforced by reputation, contextual knowledge, and continuity, epistemic control can become concentrated and weakly contestable. In mature projects with high downstream dependency, this concentration can generate path-dependent architectures and systematic filtering of alternative design proposals. Benkler explicitly notes vulnerabilities of peer production to leadership domination, reputational consolidation, and governance capture (Benkler 2006, esp. pp. 60–63; 90–94; 375–379). Kelty highlights how governance in open-source communities is frequently performed through tacit norms and informal authority structures that are difficult to render transparent or contestable (Kelty 2008, esp. pp. 196–204; 217–224).

The epistemic failure mode is not simple gatekeeping. It is the systematic filtering of technical alternatives and minority design signals through informal closure structures that become insulated from dissent. When integration authority is concentrated and procedural contestation is weak, error correction becomes institutionally costly. The system's capacity to exploit distributed expertise is reduced not because contributions cease, but because closure mechanisms narrow the admissible epistemic space.

Where architectural revision costs rise and dependency deepens, informal closure operates as a selection bias on epistemic inputs: design alternatives that would require disruptive refactoring or destabilise existing governance equilibria are filtered out independently of their technical merit. In this way, path dependence converts ordinary coordination necessities into a durable signal-suppression mechanism, structurally misaligning closure from epistemic optimisation.

Again, the claim is not determinism. It is that where closure authority becomes entrenched and contestation channels attenuated, oligarchic closure becomes a structurally recurrent tendency. The governance mediation lies in the way discretionary integration converts ordinary coordination necessities into stable epistemic distortion.

5.3 Deliberation: Conformity Cascades and Group Polarisation

In deliberative synthesis systems, the epistemic function is the convergence of distributed perspectives into collective judgement. The closure locus is the verdict, recommendation, or acceptance decision. A characteristic governance-mediated failure mode is conformity cascades leading to group polarisation.

Sunstein's analysis of group polarisation demonstrates that deliberation can amplify initial tendencies and suppress minority signals even in groups composed of reasonable participants (Sunstein 2002, esp. pp. 176–184; 185–188). The governance relevance is not that procedures cause polarisation. It is that procedural design mediates whether social pressure is dampened or amplified into epistemic closure.

The governance mechanism is the cost of dissent. Where reputational incentives, rhetorical dominance, and procedural norms raise the cost of expressing minority views, and where decision rules encourage premature convergence, deliberation becomes epistemically brittle. Closure is stabilised through normative pressure rather than through disciplined integration of heterogeneous evidence.

This diagnosis does not replace psychological explanations of conformity. It identifies how governance design determines whether conformity remains an episodic bias or becomes a stable epistemic failure mode. Where deliberative architectures insufficiently protect dissent and minority-signal uptake, conformity cascades become a systematically recurrent distortion rather than a contingent malfunction.

5.4 Platforms: Systematic Epistemic Harm and Algorithmic Distortion

In signal-extraction platforms, the epistemic function is the ranking and surfacing of micro-contributions at scale. The closure locus is algorithmic visibility: what is made salient through ranking, recommendation, and moderation. A characteristic governance-mediated failure mode is systematic epistemic harm arising from durable salience distortion.

Noble's analysis of search and ranking systems shows how algorithmic outputs can reproduce and amplify structural inequalities, misrepresent marginalised groups, and systematically distort what becomes visible and credible (Noble 2018, Introduction, esp. pp. 1–7; chs 1–2). Gillespie's account of platform governance demonstrates that moderation and curation are not merely technical operations but discretionary governance practices that structure public knowledge environments (Gillespie 2018, ch 6). O'Neil similarly shows how optimisation-driven systems can entrench patterned distortions under claims of technical neutrality (O'Neil 2016, ch 1, esp. pp. 3–12).

The governance mechanism is the combination of three features: discretionary control over ranking and moderation, opacity that blocks contestability, and optimisation objectives (such as engagement and commercial stability) that misalign salience from epistemic merit. The resulting harm is not occasional bias. It is a systematically recurrent distortion of epistemic salience and credibility allocation. Platform architectures stabilise particular narratives, suppress minority signals, and reshape interpretive horizons in patterned ways.

Algorithmic ranking systems convert engagement feedback into self-reinforcing salience equilibria: content that initially gains prominence is preferentially amplified, generating cumulative visibility advantages that progressively decouple salience from epistemic merit. In this way, closure through ranking does not merely reflect user preferences; it dynamically reshapes the epistemic environment by locking in early signal advantages as durable output hierarchies.

The term 'systematic epistemic harm' is used descriptively. The claim is not that ranking is intrinsically harmful. It is that governance architectures that combine opaque closure, misaligned optimisation, and weak accountability convert technical mediation into durable distortions of epistemic standing.

5.5 Institutions: Epistemic Closure and Interpretive Inertia

In institutional epistemic governance systems, the epistemic function is authoritative synthesis: the transformation of evidence, testimony, and normative considerations into binding interpretations and settled norms. The closure locus is the formal decision: judgment, ruling, regulation, or standard. A characteristic governance-mediated failure mode is epistemic closure accompanied by interpretive inertia.

Scott's analysis of legibility shows how institutions simplify complex realities in order to render them administratively usable (Scott 1998, ch 1, esp. pp. 11–24). These simplifications stabilise governance, but they also generate blind spots. Power's account of audit regimes shows how institutional control systems can become self-referential, prioritising procedural compliance and reputational management over substantive learning (Power 1997, ch 2, esp. pp. 38–55). Bovens, Schillemans and Goodin show how accountability regimes can become perverse when they prioritise reputational self-protection, procedural compliance, and symbolic account-giving over substantive learning and correction, and when fragmented or overloaded accountability arrangements diffuse responsibility rather than concentrating answerability (Bovens et al. 2014, pp. 10–12; 14–16). Such incentive-misaligned accountability structures weaken genuine error correction by rendering epistemic disruption institutionally costly.

The governance mechanism is the alignment of closure authority with reputational self-protection and proceduralisation. Where institutions face asymmetric reputational costs from admitting error or revising settled interpretations, epistemic disruption becomes institutionally expensive. Closure stabilises not because interpretations are maximally reliable, but because governance architectures reward continuity and penalise epistemic volatility.

The failure mode is therefore not ignorance. It is interpretive inertia: the systematic resistance to external critique and delayed uptake of countervailing evidence. Where external contestability is weak and internal incentives punish epistemic disruption, closure becomes self-reinforcing.

5.6 Systematicity, Interaction Effects, and Non-Monocausal Diagnosis

The mapping above is offered as a disciplined account of structurally characteristic failure modes under specified governance conditions. It does not deny multi-causality. It does not claim that governance explains everything. It claims that governance mediates whether known psychological, social, and strategic tendencies become stable epistemic distortions.

Moreover, these failure modes interact across systems. Markets feed narratives into platforms; platforms reshape deliberative agendas; deliberation legitimates institutional closure; institutions entrench contested interpretations into binding norms; and open-source artefacts encode values into the infrastructures that platforms and institutions rely on. The epistemic ecology is coupled. Governance failures propagate across system boundaries because epistemic closure is exercised in interlocking ways across system types.

The common structural denominator is discretionary epistemic control exercised under conditions of dependency without governance mechanisms sufficient to secure contestability of closure, transparency of power asymmetries, and protection of minority signals. Where that structure obtains, epistemic failures cease to be merely episodic. They become systematically recurrent risks.

5.7 Conclusion: From Epistemic Error to Governance-Mediated Failure

The central claim of this section can now be stated precisely. The characteristic epistemic failures of distributed systems are governance-mediated failure modes. They arise when governance architectures translate predictable human and organisational tendencies into durable distortions of salience, credibility, and closure. Narrative capture under capital asymmetry, maintainer oligarchy, conformity cascades, algorithmic salience distortion causing systematic epistemic harm, and institutional interpretive inertia are not merely isolated dysfunctions. They are structurally intelligible patterns that recur where contestability is weak, closure authority is concentrated, and error correction is institutionally costly.

Two caveats remain essential. First, this analysis does not deny empirical successes. Distributed epistemic systems often work and can outperform alternatives. Secondly, it does not offer monocausal explanations. It identifies governance as a mediating structure that determines whether known cognitive and social dynamics become stable epistemic failure modes.

The implication for the broader argument of this article is narrow but decisive. If epistemic failures are governance-mediated, then epistemic optimism without governance analysis is not merely incomplete. It is under-specified. Distributed epistemic systems do not fail simply because humans are biased or incentives are imperfect. They fail in systematically recurrent ways when closure mechanisms are misaligned with epistemic aims and weakly contestable.

6. Conclusion: Distributed Cognition as Epistemic Infrastructure

This article has argued that ‘distributed cognition’ is not a unitary epistemic good, but a family of governance-fragile epistemic infrastructures. What contemporary discourse often treats as a single phenomenon—collective intelligence generated through decentralisation and aggregation—comprises structurally distinct systems that perform different epistemic functions under different closure mechanisms, incentive structures, and governance conditions. The prevailing tendency to treat distributed cognition as an epistemic virtue in its own right therefore rests on a conceptual flattening that obscures decisive differences in how collective epistemic systems actually operate.

The first contribution of the article is the formal taxonomy developed in §3. By distinguishing belief-aggregation systems, problem-solving production systems, deliberative synthesis systems, signal-extraction platforms, and institutional epistemic governance systems, the taxonomy makes explicit what is being distributed in each case, how aggregation occurs, where epistemic closure is stabilised, and how characteristic failure modes arise. These distinctions are not classificatory refinements for their own sake. They are analytically necessary to prevent a category error that underwrites much contemporary crowd epistemology: the conflation of epistemic coordination systems with epistemic closure systems, and of heterogeneous socio-technical architectures with a single epistemic kind.

The second contribution is the paradigm contrast between prediction markets and open-source software development in §4. These cases are routinely cited as paradigms of decentralised collective intelligence, yet they exhibit fundamentally different epistemic architectures. Prediction markets distribute probabilistic belief and stabilise closure through capital-weighted price formation. Open-source systems distribute cognitive labour and stabilise closure through maintainer-centred integration. They aggregate different epistemic resources, operate through different mechanisms, and fail in systematically different ways. The contrast demonstrates why treating these systems as interchangeable instances of 'distributed cognition' is not a harmless simplification but a category error that misdescribes how epistemic closure is actually organised.

The third contribution is the governance mediation claim developed in §5. Across system classes, characteristic epistemic failures are not best understood as merely cognitive accidents or contingent distortions. They are governance-mediated failure modes. Narrative capture in markets, maintainer oligarchy in open source, conformity cascades in deliberation, algorithmic salience distortion in platforms, and interpretive inertia in institutions arise when closure mechanisms, incentive gradients, and contestation pathways convert ordinary human and organisational tendencies into durable epistemic distortions. In belief-aggregation systems specifically, capital-weighted and reflexive closure generates a directional bias toward inflationary epistemic distortion, with optimistic narratives persisting longer than warranted by underlying information and correction occurring late and discontinuously. Governance does not generate epistemic failure ex nihilo. It mediates whether known social and cognitive dynamics remain episodic malfunctions or become stable features of collective epistemic systems.

Taken together, these contributions support the core thesis: distributed cognition is best understood as an engineered epistemic infrastructure rather than as a natural epistemic good. Like all infrastructures, distributed epistemic systems perform reliably only under specific architectural conditions. Decentralisation is not reliability. Aggregation is not epistemic adequacy. Participation plurality is not epistemic robustness. Architecture determines epistemic performance.

Independent work in political theory and governance studies is converging on the same diagnosis: epistemic authority and closure are infrastructurally mediated through durable socio-technical arrangements that stabilise what counts as fact, salience, and admissible interpretation (Neame 2025, esp. pp. 416–426; 436–442; cf. Chastin et al. 2025, esp. pp. 3–7).

This diagnosis has direct implications for theories of collective intelligence. First, any serious theory of collective intelligence must also be a theory of epistemic governance. It must specify not merely how information is distributed or aggregated, but how closure is stabilised, how incentives are aligned, how contestation is secured, and how error correction is institutionally enabled or blocked. Absent this, appeals to decentralisation and diversity function as rhetorical placeholders rather than as explanatory principles.

Secondly, the analysis undermines the widespread assumption that decentralised architectures are epistemically self-stabilising. Markets, platforms, deliberative bodies, and peer-production systems do not converge on reliable outputs simply because they aggregate many perspectives. Their epistemic performance depends on whether their governance architectures dampen or amplify known failure dynamics: reflexivity, conformity, reputational pressure, path dependence, and incentive misalignment. Distributed cognition does not eliminate epistemic control. It redistributes it.

Thirdly, the analysis clarifies the nature of the category error at the heart of contemporary distributed-cognition discourse. The error is not merely that heterogeneous systems are grouped together. It is that descriptive claims about the functional distribution of cognition are treated as if they entailed claims about epistemic reliability or collective adequacy. The move from 'distributed' to 'epistemically superior' is an unsupported inference unless mediated by an account of how closure, incentives, and contestation are structured.

None of this entails epistemic pessimism. The article does not deny the empirical successes of collective epistemic systems. Prediction markets often outperform individual experts. Open-source development has produced robust technical infrastructures. Platforms can surface relevant information at scale. Deliberative bodies sometimes correct individual bias. The claim is not that distributed cognition fails everywhere or always. It is that its successes are contingent on governance conditions that are rarely theorised and rarely secured as systems scale and become consequential.

The appropriate conclusion is therefore not to abandon collective epistemic systems, but to stop treating them as epistemically self-sufficient. Distributed cognition is not a substitute for epistemic governance. It is one of its most demanding domains. Without architecture-sensitive analysis of how closure, incentives, and contestation are structured, theories of collective intelligence will continue to mistake infrastructural design problems for cognitive phenomena and governance failures for incidental anomalies.

In short, distributed cognition is not a unitary epistemic good. It is a family of governance-fragile epistemic infrastructures. Understanding that fact is a necessary condition for any serious attempt to theorise, design, or evaluate collective epistemic systems in contemporary knowledge environments.

References

Albagli, E., Hellwig, C., and Tsyvinski, A. (2024). 'Information Aggregation with Asymmetric Asset Payoffs'. *Journal of Finance*, 79(4), pp. 2715–2758. doi:10.1111/jofi.13361.

Benkler, Y. (2006) *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.

Bovens, M., Goodin, R. E. and Schillemans, T. (2014) *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press.

Chastin, S.F.M., Smith, N., Agnello, D.M., An, Q., Altenburg, T.M., Balaskas, G., de Boer, J., Cardon, G., Chinapaw, M.J.M., Chrifou, R., Dall, P.M., Davis, A., Deforche, B., Delfmann, L.R., Giné-Garriga, M., Goh, K., Hunter, S.C., Leask, C.F., Lippke, S., Loisel, Q.E.A., Longworth, G.R., McCaffrey, L., Messiha, K., Morejon, S., Pappa, D., Papadopoulos, H., Ryde, G.C., Sandlund, M., Schreier, M., Steiner, A., Verloigne, M., Vogelsang, M. and Wadell, K. (2025). 'Principles and attributes of evidence-based co-creation: From naïve praxis toward a trustworthy methodology – A Health CASCADE study'. *Public Health*, 248, 105922. doi:10.1016/j.puhe.2025.105922.

Chen, K.-Y., Fine, L.R. and Huberman, B.A. (2004). 'Eliminating public knowledge biases in information-aggregation mechanisms'. *Management Science*, 50(7), pp. 983–994. doi:10.1287/mnsc.1040.0247.

Clark, A. (1997) *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A. and Chalmers, D. (1998) 'The Extended Mind', *Analysis*, 58(1), pp. 7–19.

Gillespie, T. (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Goldman, A. I. (1999) *Knowledge in a Social World*. Oxford: Oxford University Press.

Hanson, R. (2013). 'Idea futures: Encouraging an Honest Consensus'. In M. More & N. Vita-More (eds), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future* (pp. 357–375). Chichester: John Wiley & Sons.

Hutchins, E. (1995) *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kelty, C. (2008) *Two Bits: The Cultural Significance of Free Software*. Durham, NC: Duke University Press.

Neame, L. (2025) 'Facts, arte-facts, and fabrications: The crisis of public epistemic authority', *Political Theory*, 53(4), pp. 413–450. doi:10.1177/00905917251321274.

Noble, S. U. (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

O'Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Power, M. (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Raymond, E. S. (2001) *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. Sebastopol, CA: O'Reilly Media.

Scott, J. C. (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.

Surowiecki, J. (2004) *The Wisdom Of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. New York: Doubleday.

Sunstein, C. R. (2002) 'The Law of Group Polarization', *Journal of Political Philosophy*, 10(2), pp. 175–195. doi:10.1111/1467-9760.00148.

Sunstein, C. R. (2006) *Infotopia: How Many Minds Produce Knowledge*. Oxford: Oxford University Press.

Wilson, R. A. and Clark, A. (2009) 'How to Situate Cognition: Letting Nature Take Its Course', in Robbins, P. and Aydede, M. (eds.) *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press, pp. 55–77.

Wolfers, J. and Zitzewitz, E. (2004) 'Prediction Markets', *Journal of Economic Perspectives*, 18(2), pp. 107–126.

