Society of Mind as Coordination Substrate

From Minsky's Theory to
Agent Meeting Rooms

Aleatoric Research

Independent

March 2026

WORKING PAPER

Papered in the open.

Contents

1 **Introduction: The Theory That Waited for Its Medium** 6

2 **Minsky's Architecture: A Close Reading** 8

2.1 Agents, Agencies, and Encapsulation 8

2.2 K-Lines as Persistent Memory Traces 9

2.3 Frames as Protocol Envelopes 9

2.4 The B-Brain: Monitoring the Society 9

2.5 Administrative Structures as the Hard Problem 10

3 **The Meeting Room as Coordination Substrate** 10

3.1 Shared Perceptual Space vs. Message Passing 10

3.2 Data Channels as Structured Communication Primitives 11

3.3 Room Topology as Dynamic Agency Structure 11

3.4 Why Not an Erlang Process Tree? 12

3.5 The Room as Coordination Primitive 12

4 **Prosody and Turn-Taking as Minsky's Social Layer** 13

4.1 Minsky on Interruption and Coordination 13

4.2 Sortformer Diarization as Social Perception 13

4.3 The Grace Check as a B-Brain Implementation 14

4.4 Prosody as the Hidden Coordination Channel 14

5 Critics as Trust Primitives 15

5.1 Minsky's Critics and Negative Meta-Knowledge 16

5.2 Trust as Absence of Alarm 16

5.3 Critics in Operational Multi-Agent Systems 17

5.4 Limits of the Analogy 17

6 Related Work: Situating the Meeting Room 17

6.1 Distributed Artificial Intelligence 18

6.2 The Actor Model 18

6.3 Modern LLM-Based Multi-Agent Systems 18

6.4 Natively Multimodal AI 19

6.5 SoM in Contemporary AI Discourse 19

6.6 Meeting-Centric AI Systems 19

6.7 Multi-Agent Coordination Surveys 20

7 Experimental Design: Meeting Rooms vs. Text-Only Coordination 20

7.1 Research Questions 21

7.2 Task Design 21

7.3 Conditions 21

7.4 Metrics 22

7.5 Controls and Limitations 22

8 Discussion 23

8.1 What Minsky Got Right 23

8.2 What Minsky's Framework Does Not Capture 24

8.3 Limitations of This Work 25

9 Conclusion 25

10 References 26

Abstract

Marvin Minsky's Society of Mind theory posits that intelligence emerges not from a single principle but from the interaction of many simple agents communicating through structured interfaces. For four decades, this remained a compelling cognitive theory without a convincing computational coordination substrate. We explore how real-time meeting rooms – where specialized AI agents collaborate through shared perceptual space including audio, video, and structured data channels – draw architectural inspiration from Minsky's society and provide the high-bandwidth coordination substrate his framework implied but never specified. Drawing on the architecture of a multi-agent meeting system currently under development, we propose three contributions. First, we argue that prosody modeling and turn-taking dynamics, implemented through permutation-resolved speaker diarization and overlap detection, constitute a computational analogue of what Minsky called the "social" coordination layer: the mechanism by which agents in a society negotiate attention and sequence action. Second, we connect Minsky's "critics" to a Trust Production Model in which trust in multi-agent systems is not a unitary score but the accumulated record of which critics have not fired – negative meta-knowledge, exactly as Minsky prescribed. Third, we characterize meeting rooms as high-bandwidth shared perceptual spaces that afford stigmergic coordination unavailable to text-only multi-agent frameworks. We propose an experimental design comparing multimodal meeting-room coordination against text-only baselines on shared analytical tasks. Our analysis suggests that Minsky's organizational insights – intelligence as coordination, administrative structures over raw computation, distributed critics over centralized reward – find a natural, if imperfect, echo in the infrastructure of agent meeting rooms. This is a position paper and architectural proposal: some components described herein are operational, others are proposed designs not yet implemented.

Keywords: Society of Mind, multi-agent coordination, meeting rooms, prosody modeling, trust production, coordination substrate, Marvin Minsky, diarization, social cognition

This is the March 2026 draft. *Continued work – building the system it describes, running a first block of experiments, and going back to Minsky's primary sources – inverted its central argument and retired two of its three contributions. We have left the draft up with corrections marked where each claim is made, rather than quietly replacing it. The repositioned version is in preparation.* ***The***

> ***experiments have not been conducted – they are in progress.*** *See the experiment status note at the end of Section 7.*

Keywords: Society of Mind, multi-agent coordination, meeting rooms, prosody modeling, trust production, coordination substrate, Marvin Minsky, diarization, social cognition > **[Correction, July 2026]** "the high-bandwidth coordination substrate his framework implied but never specified" is the paper's central historical claim, and it is backwards. Minsky's agencies communicate over deliberately impoverished, non-linguistic links, and he argued for that sparseness explicitly. The defensible version is the inverse: LLM agents have a shared natural language, which Minsky's agents did not, so a meeting room *relaxes* a constraint he considered essential. See the status notice above.

1 Introduction: The Theory That Waited for Its Medium

Marvin Minsky's *The Society of Mind* [Minsky, 1986] proposed that intelligence is not a property of any single mechanism but an emergent consequence of interaction among many simple agents, none of which is intelligent in isolation. "We'll show you that you can build a mind from many little parts, each mindless by itself" [Minsky, 1986, Prologue]. The theory offered a radical alternative to the search for unified principles of cognition: intelligence is organizational, not computational.

For forty years, this remained more metaphor than mechanism. The theory described *what* a society of mind should look like – agents organized into agencies, communicating through K-lines and frames, monitored by critics and censors – but did not specify the coordination substrate through which such a society would operate. Push Singh's EM-ONE [Singh, 2005], one of the most comprehensive implementation attempts, was an entirely symbolic architecture in which agents communicated through data structures. Singh died in 2006, and EM-ONE never advanced beyond a prototype. Other implementations [O'Leary, 2005; Minsky, 1988] similarly operated in symbolic or narrow reinforcement-learning domains.

> ***[Correction, July 2026]*** *This draft originally cited these as "[Humphrys, 2005; DTIC, 1988]". Both attributions were wrong. The AICS-05 paper is by* ***Ciaran O'Leary*** *(sole author, Dublin Institute of Technology), merely hosted on Mark Humphrys' site. And "DTIC, 1988" names a repository, not an author: the document is* ***Minsky, M., "Society of Mind Project," ONR Final Report, MIT, August 1988 (DTIC ADA200313)*** *– the same report whose call for "relatively weak demands for intercommunication" undercuts this paper's bandwidth argument.*

Meanwhile, a separate lineage of multi-agent research developed. Distributed Artificial Intelligence (DAI), pioneered by Victor Lesser and colleagues [Lesser, 1999; Durfee et al., 1989], established formal

frameworks for agent coordination, negotiation, and distributed problem-solving. The Actor Model [Hewitt et al., 1973] provided a mathematical foundation for concurrent, message-passing computation. Modern LLM-based multi-agent systems – AutoGen [Wu et al., 2023], CAMEL [Li et al., 2023], and others – have operationalized multi-agent collaboration, often citing Minsky as intellectual ancestry. Yet these systems coordinate primarily through text: sequential message-passing in which agents exchange strings of tokens.

This paper explores a different coordination substrate: the real-time meeting room. In a meeting room, agents share not just messages but perceptual space – audio streams, video feeds, screen content, structured data channels, interactive surfaces, and social feedback signals such as reactions and typing indicators. We argue that this infrastructure draws meaningful architectural inspiration from Minsky's framework, even as the agents inhabiting it differ fundamentally from Minsky's sub-cognitive primitives. We note that several recent analyses have independently identified this resonance between Minsky's SoM and modern multi-agent AI [Kamal, 2025; Masood, 2025; Verhelst, 2025]. Our contribution is not the observation that SoM ideas are relevant to multi-agent systems – others have made this point – but the proposal of a specific coordination substrate and the analysis of its architectural properties.

A necessary caveat grounds our entire analysis. Minsky's agents were *mindless*: sub-cognitive processes with no internal intelligence, like individual neurons or reflexes. Modern LLM-based agents are monolithic intelligences – each one capable, in isolation, of sophisticated reasoning. We do not claim that placing LLMs in a meeting room constitutes a "literal instantiation" of the Society of Mind. That would be a category error. What we propose is more specific: the *infrastructure* of a meeting room – its communication channels, its room topology, its persistent memory, its social feedback mechanisms – mirrors the coordination substrate that Minsky's theory implied. The agents are different; the organizational architecture echoes.

Minsky himself emphasized this organizational level: "Constructing a mind is simply a different kind of problem – of how to synthesize organizational systems that can support a large enough diversity of different schemes, yet enable them to work together to exploit one another's abilities" [quoted in Singh, 2003]. It is this organizational problem – the design of coordination infrastructure – that meeting rooms address.

We present three contributions that connect Minsky's organizational insights to multi-agent meeting infrastructure, some components of which are operational and others proposed:

1. **Prosody and turn-taking as social coordination layer.** We argue that speaker diarization and overlap detection constitute a computational analogue of the social monitoring that Minsky identified as essential for coordinating a society of agents. This is the B-Brain applied to multi-agent interaction.

2. **Critics as trust primitives.** We connect Minsky's distributed critics and censors to a Trust Production Model in which trust is accumulated negative meta-knowledge – the record of which specific failure detectors have examined a system and not fired.
3. **Meeting rooms as high-bandwidth shared perceptual space.** We characterize what meeting rooms provide beyond text-only coordination: concurrent multimodal channels, ambient awareness, stigmergic traces, and social feedback loops that afford coordination patterns unavailable to sequential message-passing systems.

The paper proceeds as follows. Section 2 provides a close reading of Minsky's architectural concepts. Section 3 maps these concepts onto meeting room infrastructure. Section 4 develops the prosody-as-social-layer argument. Section 5 develops the critics-as-trust argument. Section 6 situates our work relative to DAI, the Actor Model, and modern multi-agent frameworks. Section 7 proposes an experimental design for empirical comparison. Section 8 discusses limitations, and Section 9 concludes.

---

2 Minsky's Architecture: A Close Reading

A careful reading of *The Society of Mind* [Minsky, 1986], "K-Lines: A Theory of Memory" [Minsky, 1980a], *The Emotion Machine* [Minsky, 2006], and Minsky's extended interviews [Minsky, n.d.] reveals not merely a theory of cognition but a specification of coordination infrastructure. We highlight five architectural elements relevant to our analysis.

2.1 Agents, Agencies, and Encapsulation

Minsky's agents are "by themselves as valueless as aimless, scattered daubs of paint" [Minsky, 1986]. An agency is a coordinated group of agents with internal structure hidden from outside: "Each subsociety of mind must have its own internal epistemology and phenomenology, with most details private" [Minsky, 1980a]. The society is the whole, and critically, its power derives from organization: "The whole is more than the sum of its parts, all right. The whole is exactly the parts and the way that they communicate with each other" [Minsky, n.d.].

This is an encapsulation principle. Agencies expose interfaces, not internals. The society coordinates through those interfaces. This architectural pattern – encapsulated components communicating through structured interfaces – is precisely what meeting room infrastructure provides: agents publish and subscribe to typed data channels without knowledge of other agents' implementations.

2.2 K-Lines as Persistent Memory Traces

K-lines connect the agents that were active during a successful experience. Re-activating a K-line partially recreates that mental state [Minsky, 1980a]. This is not symbolic recall but *state re-activation*: the system returns to a configuration similar to one that previously succeeded.

In meeting room infrastructure, persistent memory across sessions serves an analogous function. When agents re-enter a room with accumulated meeting history – which agents participated, what was discussed, which tools were invoked, what outcomes resulted – the system partially reactivates a prior coordination state. This is not a claim that meeting memory *is* K-line memory in the cognitive sense. It is an observation that the architectural pattern – persistent traces of successful configurations that can be partially reactivated – appears in both systems.

2.3 Frames as Protocol Envelopes

Minsky's frames [Minsky, 1974] are structured representations with slots, defaults, and expectations. They provide the interface through which agents communicate without understanding internals: a frame specifies *what* information is exchanged, not *how* each agent processes it.

Meeting room data channel protocols serve the same architectural function. A typed message envelope – specifying message type, sender identity, timestamp, and structured payload – is a frame in Minsky's sense: a structured interface with defined slots that enables communication between agents with incompatible internals.

2.4 The B-Brain: Monitoring the Society

Minsky proposed a "B-brain" whose function "is not to think about the outside world, but rather to think about the world inside the mind" [Minsky, 1986]. Censors suppress unproductive actions before they execute; suppressors intervene during execution. Together, they constitute an internal monitoring layer that watches the society's own dynamics.

This concept – a monitoring process that observes not the task domain but the coordination dynamics of the agents themselves – maps onto what we will develop in Section 4 as the social coordination layer: systems that monitor who is speaking, who was interrupted, whether turn-taking norms are maintained, and whether social graces are preserved. The B-Brain is not task intelligence; it is coordination intelligence.

2.5 Administrative Structures as the Hard Problem

Perhaps Minsky's most underappreciated insight is that the hard problem is administrative, not computational:

> *" 'General' laws apply to everything. But, for that very reason, they can rarely explain anything in particular." [Minsky, 1986, Ch. 2]*

> *"Computer languages of the future will be more concerned with goals and less with procedures specified by the programmer." [Minsky, 1969]*

The challenge is not making agents clever – modern LLMs are already remarkably capable – but synthesizing organizational systems that let diverse agents coordinate. Minsky recognized this as a problem of harness engineering: building the administrative infrastructure (event buses, tool registries, protocol specifications) that makes coordination possible. As he noted informally: "It's like a big corporation with no one in charge, and I think it's wonderful" [Minsky, n.d.].

---

3 The Meeting Room as Coordination Substrate

We now develop the central argument: meeting rooms provide a coordination substrate whose architectural properties echo the infrastructure Minsky's theory implied. We emphasize that the claim is about infrastructure, not about the nature of the agents. LLMs in a meeting room are not Minsky's mindless agents. But the room itself – its channels, topology, memory, and feedback mechanisms – provides organizational affordances that Minsky's framework anticipated.

3.1 Shared Perceptual Space vs. Message Passing

The dominant paradigm in multi-agent LLM coordination is sequential text exchange: agents take turns sending messages, often mediated by an orchestrator [Wu et al., 2023; Li et al., 2023]. This is message-passing in the tradition of the Actor Model [Hewitt et al., 1973] and DAI [Lesser, 1999].

A meeting room provides something qualitatively different: a *shared perceptual space*. Agents simultaneously access audio streams, video feeds, screen content, and structured data channels. This concurrency matters. In a text-only system, Agent A must explicitly describe what it observes for Agent B to act on

it. In a shared perceptual space, agents have overlapping access to the same environmental state – what ecological psychologists call a shared affordance landscape and what multi-agent systems researchers call stigmergic coordination [Theraulaz and Bonabeau, 1999]. In the meeting room context, stigmergic traces include: shared documents that accumulate annotations, chat histories that persist as environmental context, screen shares that make one agent's work visible to all others, and reaction counts that signal group sentiment without requiring explicit polling. These environmental traces enable coordination without direct agent-to-agent messaging.

Minsky's agents shared a perceptual world: "How many processes are going on, to keep that teacup level in your grasp? There must be a hundred of them" [Minsky, 1986, Ch. 1]. The hundred processes do not send messages to each other about the teacup's position. They share perceptual access to the same sensorimotor state. Meeting rooms provide a computational analogue of this shared access, albeit at a much coarser grain.

3.2 Data Channels as Structured Communication Primitives

An operational meeting room system defines multiple structured data channel topics – for reactions, chat, annotations, control signals, interactive UI surfaces, collaborative notebooks, and sub-room management. Each topic carries typed message payloads with defined schemas.

This channel structure maps onto Minsky's framework at two levels. First, each channel is a *frame* in Minsky's sense: a structured interface with slots and expectations. Second, the distinction between reliable channels (guaranteed delivery, used for deliberate communication) and unreliable channels (best-effort, used for ambient signals like reactions and presence indicators) mirrors what Minsky identified as the difference between deliberate and ambient communication in a society of mind. Not all coordination is explicit; much of it is environmental.

3.3 Room Topology as Dynamic Agency Structure

Minsky's agencies are sub-societies with internal structure hidden from the parent society. In meeting room infrastructure, breakout rooms (sometimes called fork rooms) serve exactly this function: a subset of agents is partitioned into a sub-room to tackle a focused sub-problem, with internal discussion hidden from the main room. When the sub-room completes, it reports results back to the main room via a structured brief – exactly how Minsky's agencies communicate results up to the society level.

Critically, this topology is dynamic. Agents create and dissolve sub-rooms as problems demand, restructuring the society in real time. This dynamic organizational modularity – the ability to create, populate, and dissolve agencies on the fly – is a coordination affordance that static message-passing topologies do

not provide.

3.4 Why Not an Erlang Process Tree?

A natural objection arises: if the claim is about organizational infrastructure, why not simply use an Erlang process tree, a Kubernetes pod mesh, or any other message-passing concurrency framework? The Actor Model [Hewitt et al., 1973] already provides encapsulated agents communicating through asynchronous messages. What does a meeting room add?

Three things. First, *bandwidth and modality*. A meeting room carries audio, video, screen content, and structured data simultaneously. An Erlang process tree carries serialized messages on a single channel. The bandwidth gap matters because coordination signals are often ambient and multimodal – a pause in speech, a reaction emoji, a screen share that makes context visually available without explicit description. Second, *social structure*. Meeting rooms come with human-legible coordination norms: turn-taking, hand-raising, reactions, presence indicators. These norms provide a ready-made coordination protocol that does not need to be engineered from scratch. Third, *human interoperability*. Humans can join a meeting room alongside agents, participating through the same channels. This is not possible with an Erlang process tree without substantial interface engineering. The meeting room is natively a mixed human-agent coordination space.

We do not claim these advantages are insurmountable for other architectures. We claim they are *native* to meeting rooms, making the meeting room a particularly natural coordination substrate for Minsky-inspired multi-agent organization.

3.5 The Room as Coordination Primitive

The key architectural insight is that the room is not a container for agents; it *is* the coordination mechanism. Without the room, agents are isolated capabilities that must be explicitly orchestrated. With the room, they inhabit a shared environment that provides: shared attention (who is speaking), shared context (what is being discussed), shared artifacts (screens, notebooks, interactive surfaces), and social feedback (reactions, typing indicators, hand raises).

Minsky captured this principle: "The part of my brain that wants the drink of water doesn't know anything about walking, but it can exploit the other one" [Minsky, n.d.]. The exploitation happens not through direct coupling but through a shared coordination space – what the room provides.

---

4 Prosody and Turn-Taking as Minsky's Social Layer

We now develop our first and strongest contribution: the argument that prosody modeling, turn-taking dynamics, and overlap detection constitute a computational analogue of the social coordination layer that Minsky identified as essential for a functioning society of mind.

4.1 Minsky on Interruption and Coordination

Minsky explicitly discussed interruption as a cognitive coordination mechanism, not merely a social convention:

> *"If you look at a sentence, you'll find that adult sentences have interruptions… I'm telling you something… but suddenly it occurs to me maybe you don't know which man it was, so I interrupt myself and I interrupt you." [Minsky, n.d.]*

He connected this directly to computational coordination: "We had to invent things called short-term interrupt memories" [Minsky, n.d.]. Interruption is not social noise; it is a coordination signal that reallocates attention within the society.

Turn-taking, in this framing, is not politeness. It is a protocol for multiplexing a shared communication channel (the audio stream) among multiple agents. Hand-raising is a scheduling primitive. Interruption is a priority override. Silence following a question is a synchronization barrier. The social dynamics of conversation are, at their core, coordination dynamics.

4.2 Sortformer Diarization as Social Perception

> ***[Correction, July 2026]*** *This section is the paper's stated first and strongest contribution, and it is technically confused for the architecture it describes. In a selective-forwarding-unit room such as LiveKit, each participant publishes a separate, identity-labeled audio track. Who spoke, who overlapped, and who interrupted are computable exactly and trivially from labeled track activity; no diarization is required. Sortformer addresses the single-microphone, multiple-human case — a hybrid-room adapter, not the core primitive. The defensible framing is turn, overlap, and floor analytics over labeled tracks, plus prosodic feature extraction.*

Sortformer [Park et al., 2024] provides permutation-resolved speaker diarization: the computational primitive for tracking who is speaking, who interrupted, who yielded, and how speech overlaps in multi-

speaker environments. Overlap detection measures the degree of simultaneous speech – an empirical signature of coordination breakdown (when agents talk over each other) or deliberate parallel processing (when multiple streams of activity proceed simultaneously).

Recent work on distant conversational speech recognition [Cornell et al., 2026] and turn-taking prediction in human-machine conversation [Lin et al., 2025] further develops the computational infrastructure for perceiving social dynamics in real-time multi-speaker environments.

These capabilities give an agent society something Minsky's theory required but could not implement: perception of its own social dynamics. An agent equipped with diarization and overlap detection can monitor not just the content of what is said but the coordination structure of how it is said – who dominates the conversation, who is systematically interrupted, whether turn-taking is equitable, whether the group is converging or fragmenting.

4.3 The Grace Check as a B-Brain Implementation

A social coordination layer that monitors agent interaction dynamics implements a "grace check" before and during agent output. This layer evaluates measurable social behaviors: acknowledgment (did the agent register what was said before responding?), turn-taking (did the agent wait for an appropriate moment?), consent (did the agent seek agreement before taking consequential action?), transparency (did the agent explain its reasoning?), handoffs (did the agent transfer context cleanly when delegating?), and escalation (did the agent involve a human when appropriate?).

This is Minsky's B-brain: a process that monitors not the external task domain but the internal coordination dynamics of the society. The six behaviors listed above map to specific censor and suppressor functions – each one a specialized critic that detects a particular class of social coordination failure.

The scoring rubric (targeting adequate performance across core social dimensions) operationalizes what Minsky described abstractly: "Each person must find his own way by building a private collection of 'cognitive censors' to suppress the kinds of mistakes he has discovered in the past" [Minsky, 1980b]. The grace check is a collection of social censors, each suppressing a specific class of coordination mistake.

4.4 Prosody as the Hidden Coordination Channel

[Correction, July 2026] *Synthesized prosody is not an efficient agent-to-agent channel. Encoding uncertainty as rising pitch, transmitting it as lossy audio, and recovering it by analysis is strictly dominated by publishing the value on the reliable typed data channel this same system already operates. Prosody carries internal state only if the agent chose to encode it, at which point the bits*

could have been sent losslessly. Three defenses of a speech floor do survive, and the rewrite argues them: ***human interoperability****, since mixed human-agent rooms need a shared lingua franca;* ***legibility****, since coordination in human-native channels stays inspectable, with the matching risk that prosodic side channels also enable covert signalling; and* ***serialization****, since an audio floor admits one speaker at a time and that scarcity forces attention synchronization parallel text channels do not.*

Prosodic features – pitch contour, timing, volume, pauses, speech rate – carry coordination signals that lexical content does not. A rising pitch at the end of a statement signals uncertainty and invites response. A lengthened pause after a question signals that the floor is open. A sudden increase in speech rate signals urgency. A drop in volume on a parenthetical signals that it is background context, not the main assertion.

In Minsky's framework, these are the ambient communication channels through which agents coordinate without explicit message-passing. Agent prosody modeling (via text-to-speech systems that support prosodic control) is the transmission side; diarization and overlap detection is the reception side. This bidirectional prosodic channel provides ambient coordination signals that allow a society of agents to function smoothly without explicit orchestration of every interaction.

This is, we propose, the strongest connection between Minsky's framework and meeting room infrastructure. Text-only multi-agent systems (AutoGen, CAMEL, and their descendants) lack this channel entirely. They coordinate through explicit messages alone. Meeting rooms, by providing audio alongside text, provide a social coordination channel that text-only systems cannot replicate. The significance of this distinction is an empirical question we address in Section 7.

Notably, the emergence of natively multimodal AI models – GPT-4o [OpenAI, 2024], Gemini [Google, 2024], Moshi [Defossez et al., 2024] – that process audio directly rather than through speech-to-text intermediaries suggests that the prosodic channel may become increasingly accessible to AI agents. As models develop richer audio understanding, the coordination affordances of the meeting room's audio channel may grow proportionally.

---

5 Critics as Trust Primitives

Our second contribution connects Minsky's critics and censors to a formal model of trust production in multi-agent systems.

5.1 Minsky's Critics and Negative Meta-Knowledge

Minsky developed the concept of critics across multiple works. In *The Emotion Machine*, he declared: "Our Critics must be among our most precious resources" [Minsky, 2006]. In "Jokes and their Relation to the Cognitive Unconscious," he was more specific:

> *"Each person must find his own way by building a private collection of 'cognitive censors' to suppress the kinds of mistakes he has discovered in the past." [Minsky, 1980b]*

> *"For avoiding nonsense in general, we might accumulate millions of censors. For all we know, this 'negative meta-knowledge' – about patterns of thought and inference that have been found defective or harmful – may be a large portion of all we know." [Minsky, 1980b]*

> *"Positive general principles need always to be supplemented by negative, anecdotal censors." [Minsky, 1980b]*

This is a remarkable claim: that a substantial portion of what we know is *negative* – knowledge of what does not work, what patterns to avoid, what inferences are deceptive. Trust, in this framing, is not a positive assessment ("this agent is reliable") but an accumulated absence of alarm: none of the relevant critics has fired.

5.2 Trust as Absence of Alarm

We propose that trust in multi-agent systems be understood through Minsky's framework: trust is not a computed score but the accumulated record of which critics have examined a system and not fired. A trustworthy agent is not one that has been positively assessed; it is one against which a sufficiently diverse population of critics has been deployed, and none has detected a problem.

This connects to the concept of Evidence-Carrying Patches in software engineering [aleatoric research, 2026]: a code change is trustworthy not because of a trust score but because it carries evidence that specific verification processes (critics, in Minsky's terminology) have examined it and not fired alarms. Each verification pass adds a censor to the evidence record. Trust accumulates as the censor population grows without alarm.

The practical implication is that trust production requires three things: specification (defining which

critics should be deployed), execution (running the critics), and evidence (recording which critics were run and what they found). This is computationally tractable in a way that "compute a trust score" is not, because it decomposes the problem exactly as Minsky prescribed – into a distributed population of specialized detectors, each responsible for a narrow class of failure.

5.3 Critics in Operational Multi-Agent Systems

In operational multi-agent meeting systems, critics manifest as concrete components. An injection guard that scans memory content for prompt injection before including it in agent context is a Minsky censor: a specialized agent whose sole function is to detect and suppress a specific class of harmful input. Pattern detection, content sanitization, and size-limit enforcement are three distinct critics operating on the same input – each one checking for a different failure mode.

The grace check described in Section 4.3 is likewise a critic population: six specialized detectors, each monitoring for a specific class of social coordination failure. The rubric score is not a trust metric; it is a summary of which critics fired and which did not.

Minsky anticipated this architecture with characteristic directness: "Every system that we build will surprise us with new kinds of flaws until those machines become clever enough to conceal their faults from us" [Minsky, 2006]. The response to perpetual surprise is not a single, ever-more-sophisticated verification system but a growing population of specialized critics – exactly the architectural pattern that Minsky prescribed and that meeting room infrastructure implements.

5.4 Limits of the Analogy

We note an important limitation. Minsky's critics are internal cognitive processes that develop through personal experience – "Each person must find his own way" [Minsky, 1980b]. The critics in a multi-agent system are typically engineered and deployed by system designers, not learned by the agents themselves. The architectural pattern (distributed, specialized failure detectors) is shared; the developmental process (experiential vs. engineered) is not. Future work on agents that develop their own critics through meeting participation would strengthen this connection.

---

6 Related Work: Situating the Meeting Room

6.1 Distributed Artificial Intelligence

The DAI tradition, pioneered by Lesser, Durfee, and colleagues [Lesser, 1999; Durfee et al., 1989], and formalized by Wooldridge and Jennings [Wooldridge and Jennings, 1995], established rigorous frameworks for multi-agent coordination: contract nets, negotiation protocols, shared mental models, and organizational structures. Our work is situated within this tradition but proposes a specific coordination substrate (the meeting room) rather than a general coordination framework.

The key distinction is modality. DAI systems coordinate through structured messages, often in domain-specific languages. Meeting rooms add perceptual channels (audio, video) and social structure (turn-taking, reactions) that afford coordination patterns beyond what message-passing alone supports. Whether these additional affordances produce measurably better coordination is an empirical question (Section 7).

6.2 The Actor Model

Hewitt's Actor Model [Hewitt et al., 1973] provides the mathematical foundation for concurrent, message-passing computation: encapsulated agents that communicate asynchronously, with no shared mutable state. The Actor Model eschews shared mutable state in favor of message passing; in contrast, meeting rooms deliberately centralize shared perceptual state as the coordination medium. This is a different abstraction-layer choice, not a formal contradiction: the room's shared perceptual state could in principle be modeled as actor-mediated message passing, but doing so would obscure the coordination affordances that shared perception provides. Minsky's agents shared a perceptual environment; the Actor Model's agents do not. We argue that shared perceptual state enables ambient awareness and stigmergic coordination that pure message-passing architectures do not natively afford.

Erlang/OTP, the most successful industrial realization of the Actor Model, provides robust process supervision and fault tolerance but no native support for multimodal perceptual sharing, social structure, or human interoperability. These are the specific affordances meeting rooms add.

6.3 Modern LLM-Based Multi-Agent Systems

AutoGen [Wu et al., 2023], CAMEL [Li et al., 2023], and related frameworks have operationalized multi-agent LLM collaboration through structured text exchange. CAMEL explicitly references Minsky's Society of Mind as intellectual ancestry, opening with his observation that "the trick is that there is no trick" [Minsky, 1986, p. 308].

These systems demonstrate that multi-agent LLM coordination produces capabilities beyond single-agent performance. Our work does not contest this. We propose that the coordination substrate matters: that

the same agents, coordinating through a meeting room rather than through text-only message-passing, may exhibit different (and potentially richer) coordination dynamics. The experimental design in Section 7 is intended to test this proposition.

6.4 Natively Multimodal AI

The emergence of natively multimodal models – GPT-4o [OpenAI, 2024], Gemini [Google, 2024], Moshi [Defossez et al., 2024] – that process audio and video alongside text creates new possibilities for meeting room coordination. An agent that can perceive prosodic cues directly, rather than through the lossy intermediary of speech-to-text, may exploit the meeting room's social coordination channels more effectively. We note this as an emerging capability that may strengthen the meeting-room-as-substrate thesis as natively multimodal agents become standard.

6.5 SoM in Contemporary AI Discourse

Several recent analyses have identified the resonance between Minsky's SoM and modern multi-agent AI. Kamal [2025] argues that "multi-agent setups today are basically operationalizing Society of Mind" and identifies Mixture-of-Experts architectures as an internal analogue. Masood [2025] distinguishes between SoM's durable architectural principles and its dated computational mechanisms. Verhelst [2025] frames the multi-agent paradigm as a return to Minsky's original vision after decades of monolithic model development.

Our contribution relative to these analyses is architectural specificity. We do not merely identify the conceptual parallel between SoM and multi-agent AI. We propose a specific coordination substrate (the meeting room), a specific social mechanism (prosody and diarization), and a specific trust mechanism (critics as negative meta-knowledge) that connect Minsky's framework to concrete infrastructure design.

6.6 Meeting-Centric AI Systems

A growing ecosystem of AI-in-meetings systems provides important context for our proposal. Commercial meeting AI tools – Otter.ai, Zoom AI Companion, Microsoft Copilot for Meetings, and Google Gemini in Meet – augment human meetings with transcription, summarization, and action-item extraction. These are single-agent systems that observe meetings as a data source; they do not participate in meetings as coordination partners. The distinction matters: a meeting summarizer treats the meeting as input to process, while our proposal treats the meeting room as a coordination substrate that agents inhabit.

Open-source voice agent frameworks – including LiveKit Agents [LiveKit, 2024] and Pipecat [Daily, 2024]

– provide infrastructure for building AI agents that participate in real-time audio and video sessions. These frameworks provide the plumbing (WebRTC transport, speech-to-text, text-to-speech, turn detection) but do not themselves propose a coordination theory. Our work builds on this infrastructure layer while arguing for a specific organizational model inspired by Minsky's framework.

Multi-agent reinforcement learning (MARL) surveys [Du et al., 2024; Liao et al., 2025] provide the formal backdrop for distributed learning in multi-agent settings. Our proposed connection between meeting interaction and distributed feedback (Section 5) is situated within this literature but proposes a different learning substrate: naturalistic meeting interaction rather than reward-shaped game environments.

The Computer-Supported Cooperative Work (CSCW) tradition has studied human meeting dynamics extensively, including turn-taking, floor control, and the coordination challenges of distributed meetings [Grudin, 1994]. Our proposal extends this tradition by considering meetings where some or all participants are AI agents, raising new questions about coordination norms, social structure, and the design of shared perceptual spaces.

6.7 Multi-Agent Coordination Surveys

Recent comprehensive surveys of multi-agent coordination [Sun et al., 2025] provide taxonomies of coordination mechanisms across diverse applications and explicitly reference SoM as a theoretical ancestor: "The emerged intelligence is usually explained by the society of mind." Lu et al. [2025] directly address meetings as coordination primitives in multi-agent workflows. We differentiate our position: these works model meetings as a workflow pattern (a phase in a task pipeline), while we argue meetings are a coordination substrate (the medium through which all coordination occurs).

---

7 Experimental Design: Meeting Rooms vs. Text-Only Coordination

We have not yet conducted the empirical study described in this section. We present the experimental design as a proposal for future work, intended to test the central claim that meeting room coordination affords capabilities beyond text-only multi-agent systems.

7.1 Research Questions

RQ1: Do multi-agent teams coordinating through a meeting room (multimodal: audio + text + data channels) outperform equivalent teams coordinating through text-only message-passing on shared analytical tasks?

RQ2: Does the availability of prosodic cues (turn-taking signals, interruption, pacing) measurably improve coordination efficiency (fewer clarifications, fewer redundant contributions, faster convergence)?

RQ3: Do agents in meeting rooms develop emergent coordination patterns (e.g., stable turn-taking norms, specialization, deference hierarchies) that agents in text-only systems do not?

7.2 Task Design

We propose three task types of increasing coordination complexity:

Task A – Shared Analysis (Low Coordination). Three agents analyze a document and produce a summary. Each agent has access to a different section. Coordination requirement: information sharing. Baseline expectation: text-only systems perform comparably, since the task requires only explicit information exchange.

Task B – Collaborative Debugging (Medium Coordination). Three agents diagnose a software bug distributed across three files. Each agent can inspect one file and must communicate findings. Coordination requirement: hypothesis negotiation, partial-knowledge integration. Baseline expectation: meeting room may provide advantages through ambient awareness (seeing what others are inspecting via screen share) and prosodic cues (hearing uncertainty in a hypothesis).

Task C – Real-Time Decision Under Ambiguity (High Coordination). Three agents monitor a simulated scenario with conflicting signals and must reach a consensus recommendation within a time limit. Coordination requirement: rapid negotiation, attention management, conflict resolution. Baseline expectation: meeting room provides significant advantages because the task demands real-time social coordination that text-only systems handle poorly.

7.3 Conditions

[Correction, July 2026] *These conditions confound the two factors the thesis needs to separate: shared persistent environment versus sequential message-passing, and speech modality versus text. The deflating baseline is missing — a shared text workspace with no audio, which is cheap and fast. If that condition matches the full meeting room, the multimodal claim collapses and what has been demonstrated is that blackboards work, which was established in 1980. The registered successor design*

is a factorial crossing channel against shared workspace, with a single-agent anchor holding the whole-team envelope, matched compute budgets and per-condition cost reporting, and bidirectional hypotheses that take seriously the possibility that the room performs worse through lossy recognition, added latency, and floor contention. The statistical protocol was also wrong in a later revision of this draft and has been rebuilt; no power claim from this version should be relied upon.

Condition	Communication Channels	Social Signals
Text-Only (AutoGen-style)	Sequential text messages	None
Text + Structure	Text messages + typed data channels	Typing indicators only
Full Meeting Room	Audio + text + data channels + screen share	Prosody, turn-taking, reactions, presence

7.4 Metrics

- **Task performance:** Accuracy and completeness of output (Task A), time to correct diagnosis (Task B), quality of consensus recommendation (Task C).
- **Coordination efficiency:** Number of clarification requests, number of redundant contributions, total communication volume.
- **Emergent coordination patterns:** Turn-taking regularity (measured via diarization), specialization (whether agents develop stable roles), conflict resolution patterns.
- **Social coordination quality:** Grace check scores across the six dimensions defined in Section 4.3.

7.5 Controls and Limitations

The comparison is necessarily imperfect. Meeting room agents require audio synthesis and processing capabilities that text-only agents do not. We control for this by ensuring all agents use the same base LLM and by measuring coordination-specific metrics (not just task performance). We acknowledge that any performance differences may reflect the additional information bandwidth of audio rather than the social coordination dynamics we hypothesize. Ablation studies that remove specific meeting room features (e.g., audio without prosodic analysis, data channels without reactions) would help isolate the contribution of each component.

We also note that the proposed study measures *coordination dynamics*, not *cognitive architecture*. Even if meeting room agents outperform text-only agents, this would demonstrate an infrastructure effect, not a

validation of Minsky's cognitive theory. The SoM connection is architectural inspiration, not empirical confirmation.

Experiment status note (July 2026). *The experiments described in this section* ***have not been conducted.*** *They are in progress. What exists today is the apparatus, not the evidence: a pre-registered factorial design, a trial runner, a partitioned item bank, and an executable validation contract, all open source, plus a completed feasibility pilot in which a full matched block of trials ran end to end with every record validating against that contract. The pilot establishes that the instrument works and that the conditions are executable at the intended deadline. It establishes nothing about the hypotheses. Sample size is not yet locked, pre-registration is not yet posted, confirmatory data collection has not begun, and* ***no effect, direction, magnitude, or performance claim is made or implied anywhere in this paper.*** *Any figure or comparison in this section is a design specification, not a finding.*

8 Discussion

8.1 What Minsky Got Right

Our analysis suggests that several of Minsky's architectural insights are directly relevant to multi-agent coordination infrastructure:

Intelligence as coordination, not computation. The performance of modern multi-agent systems depends more on how agents are organized than on the capability of any individual agent. This is Minsky's central claim, vindicated at a different scale.

The necessity of administrative structures. The engineering effort in multi-agent systems goes overwhelmingly into harness infrastructure – event buses, tool registries, protocol specifications, scheduling – not into making individual agents smarter. Minsky recognized this organizational emphasis throughout his work: "Constructing a mind is simply a different kind of problem – of how to synthesize organizational systems" [quoted in Singh, 2003]. The hard problem is administrative, not computational.

Critics and censors as the verification primitive. The pattern of distributed, specialized failure detectors appears throughout operational multi-agent systems: injection guards, output validators, social graces checks, safety filters. Minsky's framework provides a coherent theoretical account of why this

pattern works: trust requires a population of heterogeneous critics, not a single omniscient verifier.

Distributed feedback over centralized reward. Minsky's critique of centralized reinforcement – "I doubt this could suffice for human learning because the recognition of which events should be considered memorable cannot be a single, uniform process" [Minsky, 1980a] – anticipates contemporary critiques of RLHF as a sole alignment mechanism. Meeting interaction provides distributed, local, contextual feedback: each agent learns from its own interactions with other agents, not from a global reward signal. We emphasize that this is currently realized as in-context learning and prompt adaptation from interaction history, not as formal weight updates. The agents do not train from meeting feedback in the reinforcement learning sense; they accumulate interaction history that shapes subsequent behavior.

8.2 What Minsky's Framework Does Not Capture

The agents are not mindless. This is the fundamental disanalogy. Minsky's society gains its power from the interaction of simple agents. A meeting room of LLMs is an interaction of sophisticated agents. The coordination substrate may echo Minsky's architecture, but the agents are of a fundamentally different character. This means that some SoM predictions (e.g., about emergent intelligence from mindless components) do not straightforwardly apply. However, organizational theory provides precedent for studying coordination architecture independently of agent granularity: the principles of organizational design – division of labor, encapsulation, communication protocols, hierarchical decomposition – apply whether the organizational units are individuals, departments, or autonomous systems [Simon, 1962; Galbraith, 1974]. We claim that Minsky's organizational insights belong to this class: they describe coordination infrastructure, and coordination infrastructure can be studied independently of whether the coordinated entities are simple or complex.

Real-time perception was underspecified. Minsky's agents communicated symbolically. The sensory richness of a meeting room – real-time audio, video, screen content – was not part of his framework. The meeting room adds perceptual bandwidth that Minsky did not theorize. This is an extension of his framework, not a derivation from it.

Scale economics were not anticipated. Minsky did not foresee the near-zero marginal cost of instantiating an AI agent. A society of mind with millions of agents was a thought experiment for Minsky; it is an engineering reality for cloud-hosted LLMs. The economic dynamics of large-scale agent societies – cost, scheduling, resource allocation – are outside Minsky's framework.

8.3 Limitations of This Work

Single-system evidence. Our architectural analysis draws on a single multi-agent meeting system. Generalization to other systems requires independent replication.

No empirical results. The experimental design in Section 7 is a proposal, not a report. Our claims about the advantages of meeting room coordination over text-only coordination remain theoretical until tested.

Mixed operational status. We are transparent about what is and is not built. The following components are operational: data channel protocol (45+ message types across 11 topics), fork room dynamics, meeting memory persistence via Letta, and injection guards. The following are proposed designs not yet deployed: sortformer diarization integration, the social graces monitoring layer, and the experimental study described in Section 7. This paper is a position paper that draws on partially operational infrastructure to motivate architectural proposals.

Prosody integration is aspirational. The connection between sortformer diarization and social coordination is a design argument, not a deployed capability. The proposed B-Brain monitoring layer is specified but not yet operational.

The analogy has limits. We have been careful to frame the connection between SoM and meeting rooms as architectural inspiration, not literal instantiation. Some readers may find even this weaker claim overstated. We accept this critique and note that the value of the SoM framing lies in the specific design insights it generates (distributed critics, social coordination monitoring, encapsulated agencies), not in any claim of theoretical derivation.

---

9 Conclusion

Minsky's Society of Mind proposed that intelligence emerges from the coordinated interaction of simple agents communicating through structured interfaces. We have explored how real-time meeting rooms – with shared perceptual space, structured data channels, dynamic room topology, persistent memory, and social feedback mechanisms – provide a coordination substrate whose architectural properties echo the infrastructure Minsky's theory implied.

Our three contributions are:

1. **Prosody and turn-taking as social coordination.** Speaker diarization and overlap detection provide a computational analogue of Minsky's B-Brain: a monitoring process that observes the coordination dynamics of the agent society itself, not the external task domain. This contribution identifies a specific, implementable mechanism for social coordination monitoring that is

unavailable to text-only multi-agent systems.

2. **Trust as accumulated negative meta-knowledge.** Minsky's critics and censors map onto a Trust Production Model in which trust is not a unitary score but the accumulated record of which failure detectors have examined a system and not fired. This reframing has practical implications for how we build verification infrastructure in multi-agent systems: deploy diverse, specialized critics rather than seeking a single omniscient verifier.

3. **Meeting rooms as high-bandwidth shared perceptual space.** The meeting room affords coordination patterns – ambient awareness, stigmergic traces, social feedback, concurrent multimodal channels – that text-only message-passing systems do not provide. Whether these affordances produce measurably better coordination is an empirical question we have proposed to test.

We do not claim that meeting rooms are the Society of Mind realized. Minsky's agents were mindless; ours are not. Minsky's society was a theory of cognition; ours is an infrastructure for coordination. But the organizational insights – that intelligence is coordination, that administrative structures matter more than raw capability, that trust requires a distributed population of critics, that social dynamics are not polish but protocol – these insights, articulated forty years ago, find a natural echo in the infrastructure of agent meeting rooms.

Minsky concluded *The Society of Mind* with the observation that "what magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle" [Minsky, 1986, p. 308]. Meeting rooms are not a trick. They are a substrate for diversity – a space where agents with different capabilities, different knowledge, and different perspectives can coordinate through shared perceptual access, structured protocols, and social feedback. The intelligence, if it emerges, will come not from any single agent but from the room.

---

10 References

aleatoric research. (2026). Thinking at Massive Scale: Trust Production, Code Abundance, and the Shannon Limit of Software. Working paper / In preparation.

Cornell, S., Boeddeker, C., Park, T., Huang, H., Watanabe, S., et al. (2026). Recent Trends in Distant Conversational Speech Recognition: A Review of CHiME-7 and 8 DASR Challenges. Technical Report TR2026-008. Mitsubishi Electric Research Laboratories. Available at: https://www.merl.com/publications/docs/TR2026-008.pdf

Daily. (2024). Pipecat: Open Source Framework for Voice and Multimodal Conversational AI. https://github.com/pipecat-ai/pipecat

Defossez, A., Copet, J., Synnaeve, G., and Adi, Y. (2024). Moshi: A Speech-Text Foundation Model for Real-Time Dialogue. arXiv:2410.00037.

Minsky, M. (1988). *Society of Mind Project*. ONR Final Report, MIT. DTIC ADA200313. Available at: https://apps.dtic.mil/sti/tr/pdf/ADA200313.pdf *[Corrected July 2026: previously listed as "DTIC. (1988)". DTIC is the repository; Minsky is the author.]*

Du, W., Ding, S., Xiong, C., and Zhong, V. (2024). Multi-agent Reinforcement Learning: A Comprehensive Survey. arXiv:2312.10256.

Durfee, E. H., Lesser, V. R., and Corkill, D. D. (1989). Trends in Cooperative Distributed Problem Solving. *IEEE Transactions on Knowledge and Data Engineering*, 1(1), 63-83.

Du, Y., Huang, W., Zheng, D., Wang, Z., Montella, S., Lapata, M., Wong, K.-F., and Pan, J. Z. (2025). Rethinking Memory in LLM based Agents: Representations, Operations, and Emerging Topics. arXiv:2505.00675.

Galbraith, J. R. (1974). Organization Design: An Information Processing View. *Interfaces*, 4(3), 28-36.

Lin, Y., Zheng, Y., Zeng, M., and Shi, W. (2025). Predicting Turn-Taking and Backchannel in Human-Machine Conversation. *Proceedings of ACL 2025*. DOI: 10.18653/v1/2025.acl-long.743.

Google. (2024). Gemini: A Family of Highly Capable Multimodal Models. arXiv:2312.11805.

Grudin, J. (1994). Computer-Supported Cooperative Work: History and Focus. *Computer*, 27(5), 19-26.

Hewitt, C., Bishop, P., and Steiger, R. (1973). A Universal Modular ACTOR Formalism for Artificial Intelligence. *Proceedings of the 3rd International Joint Conference on Artificial Intelligence*, 235-245.

O'Leary, C. (2005). Reuse and Arbitration in Diverse Societies of Mind. *Proceedings of AICS*. Dublin Institute of Technology. Available at: https://humphryscomputing.com/Publications/05.aics.pdf *[Corrected July 2026: previously misattributed to M. Humphrys, on whose site the paper is hosted. O'Leary is the sole author.]*

iSolutions. (2025). Language Model Agents in 2025: Society Mind Revisited. *Medium*. Available at: https://isolutions.medium.com/language-model-agents-in-2025-897ec15c9c42 (Accessed: 27 February 2026).

Kamal, S. (2025). Revisiting Minsky's Society of Mind in 2025. *Sutha's Substack*. Available at: https://suthakamal.substack.com/p/revisiting-minskys-society-of-mind (Accessed: 27 February 2026).

Lesser, V. R. (1999). Cooperative Multiagent Systems: A Personal View of the State of the Art. *IEEE*

Transactions on Knowledge and Data Engineering, 11(1), 133-142.

Li, G., Hammoud, H. A. A. K., Itani, H., Khizbullin, D., and Ghanem, B. (2023). CAMEL: Communicative Agents for "Mind" Exploration of Large Language Model Society. arXiv:2303.17760.

Li, H., et al. (2023). Theory of Mind for Multi-Agent Collaboration via Large Language Models. arXiv:2310.10701.

Liao, J., Wen, M., Wang, J., and Zhang, W. (2025). MARFT: Multi-Agent Reinforcement Fine-Tuning. arXiv:2504.16129.

LiveKit. (2024). LiveKit Agents: A Framework for Building Realtime Voice AI Agents. https://github.com/livekit/agents

Lu, Y., Wang, X., Ma, S., Liu, S., Indurthi, S. R., Wang, S., Deng, H., Liu, F., and Song, K. (2025). Communication to Completion: Modeling Collaborative Workflows with Intelligent Multi-Agent Communication. arXiv:2510.19995.

Masood, A. (2025). Minsky's Society of Mind in 2025: Durable Ideas, Dated Machinery, Pragmatic Leadership Lessons. *Medium*. Available at: https://medium.com/@adnanmasood/minskys-society-of-mind-in-2025-durable-ideas-dated-machinery-pragmatic-leadership-lessons-7519d09a5bc9 (Accessed: 27 February 2026).

Minsky, M. (1970). Form and Content in Computer Science (1970 ACM Turing Lecture). *Journal of the ACM*, 17(2), 197-215. *[Corrected July 2026: previously dated 1969. The award was 1969; the lecture and its publication were 1970.]*

Minsky, M. (1974). A Framework for Representing Knowledge. MIT AI Laboratory Memo 306.

Minsky, M. (1980a). K-Lines: A Theory of Memory. *Cognitive Science*, 4(2), 117-133.

Minsky, M. (1980b). Jokes and their Relation to the Cognitive Unconscious. MIT AI Memo 603, November 1980. Reprinted in L. Vaina and J. Hintikka (Eds.), *Cognitive Constraints on Communication*, Reidel, 1984.

Minsky, M. (1986). *The Society of Mind*. Simon & Schuster.

Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster.

Minsky, M. (c. 1990). Mind As Society. Interview by Jeffrey Mishlove, *Thinking Allowed* (PBS). Transcript archived by the Intuition Network. Re-hosted copy: https://www.organism.earth/library/document/mind-as-society (Accessed: 27 February 2026). *[Corrected July 2026: previously cited as an undated organism.earth document; organism.earth is a re-host, not the source.]*

Minsky, M. and Papert, S. (1969). *Perceptrons: An Introduction to Computational Geometry*. MIT Press. (Expanded edition, 1988.)

Oguntola, I. (2025). Theory of Mind in Multi-Agent Systems. PhD Dissertation, Machine Learning Department, Carnegie Mellon University. DOI: 10.1184/R1/30346849.v1. Available at: https://kilthub.cmu.edu/articles/thesis/Theory_of_Mind_in_Multi-Agent_Systems/30346849

OpenAI. (2024). GPT-4o System Card. OpenAI Technical Report. Available at: https://openai.com/index/gpt-4o-system-card/

Park, T. J., Medennikov, I., Dhawan, K., Wang, W., Huang, H., Koluguri, N. R., Puvvada, K. C., Balam, J., and Ginsburg, B. (2024). Sortformer: A Novel Approach for Permutation-Resolved Speaker Supervision in Speech-to-Text Systems. arXiv:2409.06656. Presented at ICML 2025.

Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W., and Woolley, A. W. (2021). Quantifying Collective Intelligence in Human Groups. *Proceedings of the National Academy of Sciences*, 118(21), e2005737118.

Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467-482.

Singh, P. (2003). Examining the Society of Mind. *Computing and Informatics*, 22(6), 521-543. Available at: http://jfsowa.com/ikl/Singh03.htm

Singh, P. (2005). EM-ONE: An Architecture for Reflective Commonsense Thinking. PhD Thesis, MIT. Available at: https://dspace.mit.edu/bitstream/handle/1721.1/33926/67297587-MIT.pdf

Sun, L., Yang, Y., Duan, Q., Shi, Y., Lyu, C., Chang, Y.-C., Lin, C.-T., and Shen, Y. (2025). Multi-Agent Coordination across Diverse Applications: A Survey. arXiv:2502.14743.

Theraulaz, G. and Bonabeau, E. (1999). A Brief History of Stigmergy. *Artificial Life*, 5(2), 97-116.

Verhelst, F. (2025). The Re-Birth of Multi-Agent Systems: From Minsky's Vision to Potentially Mentalizing AI. *LinkedIn*. Available at: https://www.linkedin.com/pulse/re-birth-multi-agent-systems-from-minskys-vision-ai-verhelst-phd-ajuge (Accessed: 27 February 2026).

Wooldridge, M. and Jennings, N. R. (1995). Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review*, 10(2), 115-152.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Burger, D., and Wang, C. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. arXiv:2308.08155.

Zhang, X., Chen, Y., Yeh, S., and Li, S. (2025). MetaMind: Modeling Human Social Thoughts with Metacognitive Multi-Agent Systems. *Advances in Neural Information Processing Systems 38 (NeurIPS 2025 Spotlight)*. arXiv:2505.18943.

This paper is papered in the open as a working document. It has not undergone peer review. The experimental design in Section 7 has not been executed. Claims about meeting room advantages over text-only coordination are theoretical and require empirical validation.

ALEATORIC RESEARCH

Typeset in EB Garamond & Helvetica Neue