

DIAGNOSTIC – STS PRINCIPLES AT WORK IN THE PANEL

Teaming with Robots and AI · MWW – STS Roundtable Panel discussion in collaboration with TIHR · Zoom, 25 June 2026

A reading of the session as a temporary sociotechnical system – what was espoused, and what was enacted

Version control

Version	Date	Authors	Notes
1.0	29/06/2026	Annja Neumann, Otter.ai Notetaker, Claude Opus 4.8	Designed, prompted and edited by AN

The session set out to discuss principles for teaming with AI and, in the doing, became the thing it was discussing: a temporary sociotechnical system of fifty-one people, a Zoom platform, a polling tool and – by the room’s own consent – an AI Notetaker. This diagnostic reads the system back to itself. It maps which STS principles were at work, explicitly and implicitly, and sets what the room voted as most important against what it actually did. The distinction is the old one between espoused theory and theory-in-use (Argyris and Schön, 1974): the gap between the two is usually where the learning is.

A note on the record

The AI Notetaker joined late, so the transcript opens midway through Shankar Sankaran’s presentation. The welcome, the naming of panellists, and the consent poll – in which the room voted unanimously to admit the Notetaker – are reconstructed from the chair’s notes. The omission is itself data: the AI team-member holds no record of the moment the team decided to admit it, nor of its own arrival. The humans hold that boundary-memory; the machine does not. A panel on teaming with AI begins with the AI already having missed the part where it was made welcome.

What the room said mattered – the poll

Forty-five of fifty-one participants each chose three of twelve principles. Collaborative sensemaking was the clear winner; trustworthiness came second; and a cluster of adaptive autonomy, ethical considerations and participation / co-design followed. Cognitive workload drew the least support – which Shankar immediately flagged as “an important one”. As Annja framed it at the time, this was “a mood reading, not a verdict” – a fast picture of what the room believed it cared about.

What the room actually did

Read against the transcript, the live system enacted a wider and partly different set of principles than it voted for. Three groups stand out.

Lived and named – prized and practised

Participation and co-design was the strongest current of all. The poll itself, the consent vote, the breakouts and the heavy use of the floor made the audience part of the system rather than its spectators – Carly Riehle’s “most engaged audience we’ve had” was the design working as intended. **Collaborative sensemaking** and **adaptive autonomy** – both treated below – were likewise voted high and visibly enacted. **Trustworthiness** ran throughout, but the room kept relocating it: a participant’s statment “I might trust the machine, but I don’t trust the person who owns the machine” moved trust off the device and onto the political and economic system above it (ownership, the manpower companies, national sovereignty, Musk, the NSA). Trust was enacted as a question about the wider system, not the tool.

Lived but not prized – the tacit spine

Boundary controls scored low in the poll yet was arguably the most deliberately enacted principle in the room: the consent vote drew a line around a human-machine system, the “main room only” rule was named and held (“make sure the notetaker stays in the main room”), and the breakout prompt asked everyone to find a boundary in their own work. **Multidisciplinarity** was enacted before it was named – engineers, an ethicist, organisational psychologists, a performance researcher, OD consultants, an industrial engineer, a criminologist and, late on, a Japanese cultural reading – and only acknowledged at the close (Bert Painter: “you applied the principle of multidisciplinarity”). **Agile and responsive** showed in the running order being reshaped to the vote in real time; **unpredictability management** in how the system absorbed the late Notetaker, the stalling breakout algorithm and the constant mutings – and in a participant’s proposal for adversarial, red-team / blue-team AI as a way to make machine judgement trustworthy.

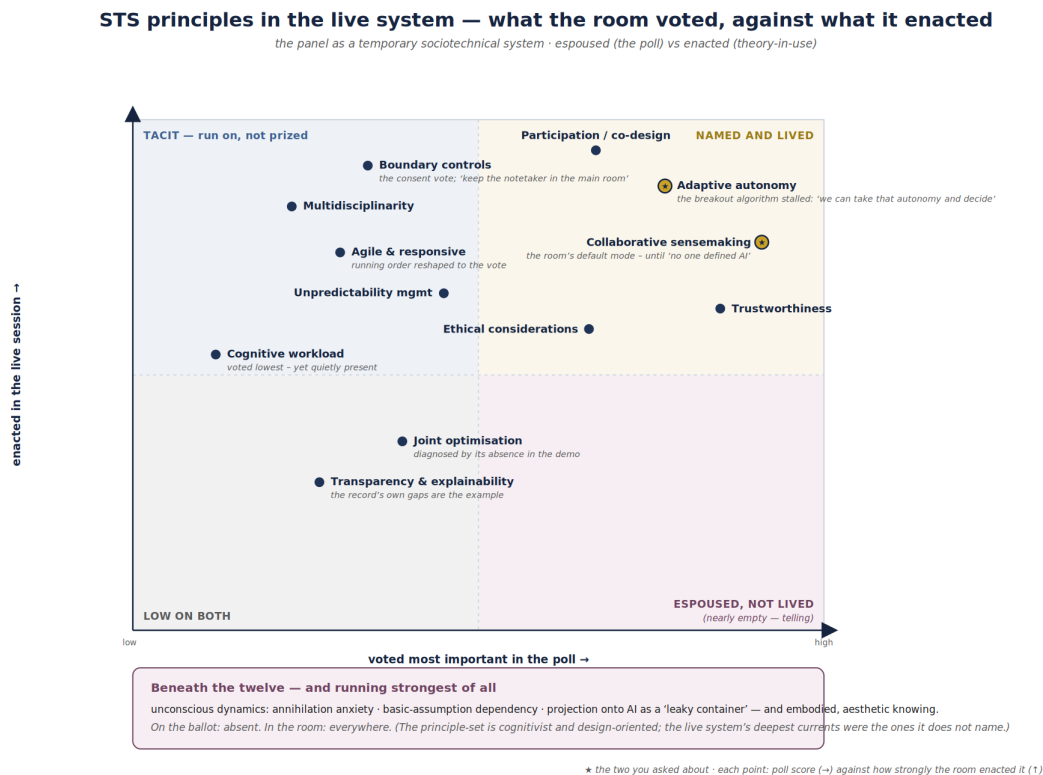
Discussed but not lived – present mostly as absence

Joint optimisation was defended in principle (Carly) but diagnosed by its absence in the demonstration: David Lawlor’s observation that “the technology was really managing the humans”, and that the early-STS commitment to autonomous work groups – workers deciding among themselves how to use the tool – was missing, located the social half of the pairing precisely where it had been left out.

Transparency and explainability surfaced as content (Francis Irving on the limits of current systems) but was enacted as a gap: the system’s own record is opaque and partial, the Notetaker having missed its own beginning. **Cognitive workload**, voted lowest, was quietly the most operative of the dismissed principles – Zoom fatigue, the “lot to hold”, and Henrietta Hale’s closing worry that we are caught in “continual learning of a system... without actually getting the job done”. The room voted out the very thing that was taxing it.

The map

Each principle is placed by how high it scored in the poll (horizontal) against how strongly the room enacted it (vertical). The two you asked about are starred.



The shape of the map is the finding. The top-right — prized and practised — holds the principles the room both valued and lived. The top-left is fuller than the room would have predicted: boundary controls, multidisciplinarity, agility and unpredictability management were the tacit spine it ran on without putting them forward. The bottom-right — valued but not lived — is almost empty, which says something good about this group: it did not espouse principles it then abandoned. And the bottom band names what no ballot of twelve cognitivist, design-oriented principles could capture: the layer that, on the evidence, ran strongest of all.

The two principles participants voted highest

Adaptive autonomy

Voted into the top cluster, and then handed its clearest demonstration by accident. When the automated breakout allocation stalled, the system performed adaptive autonomy in real time without anyone designing the moment: “It feels like we’re at the mercy of the algorithm”, then “the automated process is slower than us deciding ourselves”, and finally “I think we can take that kind of autonomy and decide what the topic is.” Control passed back to the humans the instant the machine’s performance fell below the task’s tolerance — which is exactly what the principle says should happen, shifting moment to moment with the task. Shankar’s two-hour dialogue with Copilot is the same principle on a longer arc and in the opposite

direction: distrust, then graduated handover as competence is demonstrated, until “it became a collaborative effort” rather than master-and-assistant. Between the two, the room saw autonomy negotiated in both directions – the bidirectionality the ‘Advancing STS’ paper claims – once live and once unbidden.

Collaborative sensemaking

The room’s top vote, and its default register: the entire ninety minutes was distributed sense-making across disciplines. But the transcript also stages its failure condition. A participant’s intervention – that the discussion had run on “AI” without anyone defining it, with different people “thinking it means completely different things” – names precisely what the principle’s own definition warns against: without a shared object and shared roles, people build not a shared picture but parallel ones, and talk past each other. The room had been doing collaborative sensemaking and, in Alter’s moment, sensemaking about its own sensemaking – discovering that its most-prized principle had been operating on an undefined term. That the group could metabolise the challenge rather than defend against it is itself evidence the principle was alive.

Beneath the twelve

The strongest currents in the room were not on the ballot. The panel repeatedly reached past the cognitivist, design-oriented frame of the principles towards two registers it insisted are primary rather than supplementary. The first is the psychodynamic at individual, group and systems levels: Jennie McShannon’s naming of annihilation anxiety and the fear of lost competence, and her warning that an anxious group can tip into basic-assumption dependency on the AI; Henrietta Hale’s and a participant’s reading of AI as a container for projection – a place to put blame and rage – and a participant’s sharpening of it into the abdication of responsibility (“it’s not me, it’s the AI’s fault”). Annja’s own footnote is the most pointed diagnostic in the session: that as projection lands on the AI rather than on other people, the consultant has “less data to work with” – the machine is absorbing the very material a systems-psychodynamic reading depends on. The second register is the embodied and aesthetic: the felt response to the principles that Jennie described, and the question, held in reserve from the speaker notes, of what the body knows about a system that no design principle captures. Annja closed by naming a thirteenth principle that had not been on the list – ‘socially shaped’ – to hold how the social system shapes what the technology becomes.

Espoused against enacted – a summary

Principle	Poll (espoused)	Live (enacted)	Where it showed
Collaborative sensemaking	1st	High	default mode; falters when ‘AI’ left undefined
Trustworthiness	2nd	Med-high	trust the owner, not the machine; Copilot arc
Adaptive autonomy	top cluster	High	breakout algorithm stalls; control taken back; Copilot

Participation / co-design	top cluster	High	poll, consent vote, breakouts; 'most engaged audience'
Ethical considerations	top cluster	Medium	consent; 'who pays'; projection as abdication
Boundary controls	low	High	consent vote; 'notetaker stays in the main room'
Multidisciplinarity	low	High	the mix; named only at the close
Agile & responsive	low	High	running order reshaped to the vote
Unpredictability mgmt	low	Med-high	late Notetaker; stalling algorithm; red/blue AI
Cognitive workload	lowest	Low-med	Zoom fatigue; 'a lot to hold'; continual catch-up (Hale)
Joint optimisation	low-mid	Low (absence)	'the technology was managing the humans'
Transparency & explainability	low-mid	Low (absence)	explainability gap; the record's own missing opening

What it tells us

Four things. First, the room is more sociotechnically competent than it knows: the principles it ran on most reliably – boundary controls, multidisciplinarity, agility – were ones it did not think to prize, which is the signature of a practised group operating below its own awareness. Second, the cleanest demonstrations were unplanned: the stalling algorithm taught adaptive autonomy better than any provocation could, and the question about a shared language and understanding of AI taught the limits of collaborative sensemaking from inside it. The system's incidents were its best curriculum.

Third, and most useful for the wider programme: the principle-set, for all its reach, does not name the layer that moved the room most – anxiety, dependency, projection, and the felt and embodied sense of the system (as part of the sentient system). If those are to count as STS evidence rather than as secondary, the framework has to make room for what a body and a group know and cannot measure. On this session's evidence, that is not a supplement to the principles. It is the part that was doing the work. Fourth some of the learning was embedded in the boundary management which was at first not rated highly by the group but gained visibility and significance through the contracting on the AI Notetaker and the sociotechnical conditions of it crossing the boundary ("only main room"; "anonymising all participants", "transcript available on request") which was unanimously accepted by the group to join.

Tavistock Institute of Human Relations · diagnostic prepared for review · not for circulation without consent

AI Transparency statement: I used Claude Opus 4.8 and the AI Notetaker's transcript to produce the mapping of the STS principles that were live and at work in the panel discussion session and a diagnostic (see prompt below) and then edited and expanded the diagnostic subsequently. This included anonymising all participants as contracted. Panel members, speakers and Bert continue to be named.

Prompt:

The panel discussion took place yesterday. I attached the AI Notetaker transcript here. Note that the AI Notetaker joined late. What it missed is covered in the first sections of my speakers notes 1-2 (attached here too - take into account that I did not cover all questions listed in the notes; the transcript takes priority after the naming of participants and the AI Notetaker poll). It also missed that all participants voted for the AI Notetaker to join the meeting. Map what STS principles were at work during the session, implicitly and explicitly, based on the transcript and drawing on the interaction between espoused theory and theory-in-use (Argyris and Schön, 1974). Create a diagnostic text and a visual map that names the STS principles which informed the interactions and topics of the discussion. Also compare them against the principles with the most votes, especially collaborative sensemaking and adaptive autonomy.