PROTOCOL INSTITUTE SYMPOSIUM · 23 SEPTEMBER 2026

Protocols for Co-Cognition

The New Nature of Collaboration in the AI Era

Robert Peake

Is AI a Tool or a Relationship?

The Accountability Lens

TOOL

No accountability

It sits with the user.

RELATIONSHIP

Shared, or hierarchical accountability

Both incur consequences.

COUNTERPARTY

Accountability is asymmetric, at cross-purposes, or absent

Wired into the protocol.

Old Nature of Collaboration

Language invites a social read	only minds produced language	Nass & Moon · Epley
The burden of grounding is human	uptake was costly and meant something	Clark
Aversion and appreciation both real	reliance is miscalibrated in both directions	Dietvorst et al. · Logg et al.
Reliance follows cues not data	cues were proxies for competence	Lee & See · Parasuraman
No accountability when offloading	the other party had a stake	Risko & Gilbert · Wegner

What Is Effectiveness in Co-Working with AI?

Working with AI to get **more** of **the right things done right**,

in less time and/or **at lower cost** than doing it alone,

while staying **relaxed** and **in control**.

Working with AI to get **more** of **the right things done right**,

in less time and/or **at lower cost** than doing it alone,

while staying **relaxed** and **in control**.



Working with AI to get **more** of **the right things done right**,

in less time and/or **at lower cost** than doing it alone,

while staying **relaxed** and **in control**.

$$\frac{\sum_{tasks} \frac{(fit \times correctness)}{(load + debt)}}{time/\,time_0 \times cost\,/\,cost_0}$$

$time_0$, $cost_0$ are the time and cost of doing the same work alone
load is always > 0 and debt is never negative

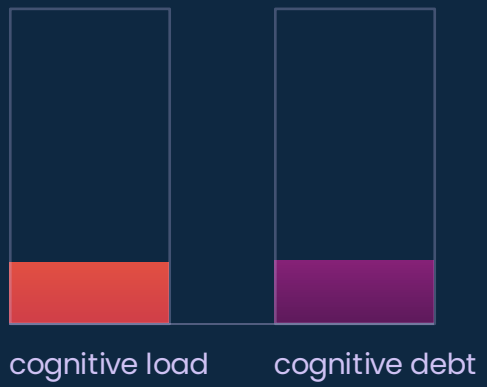
more … in less time | the right things | done right | relaxed | in control | at lower cost

What Is Ineffectiveness in Co-Working with AI?

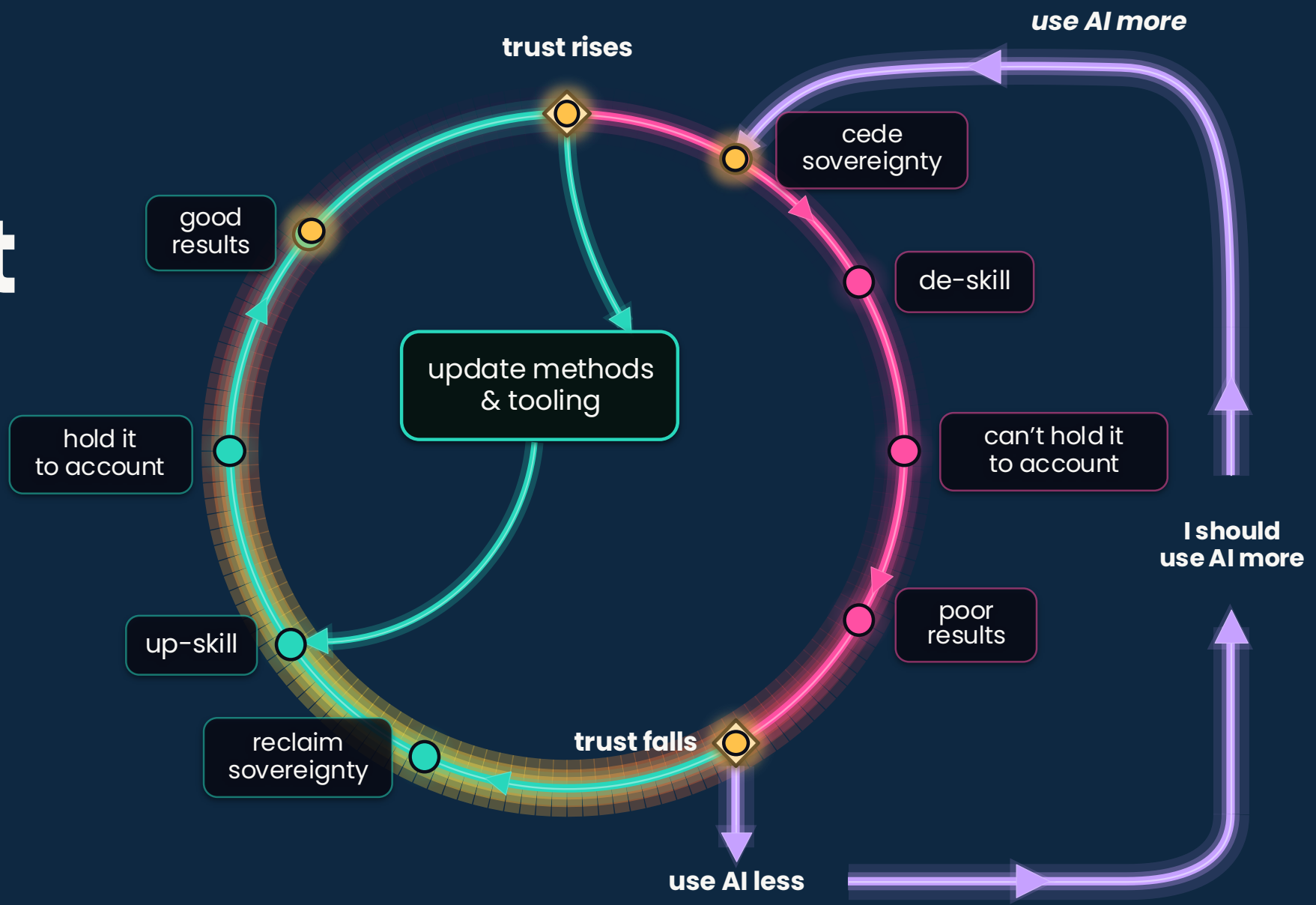
AI CO-WORKING EFFECTIVENESS

The trust trap.

How methods and tooling trade one-off design effort for less recurring vigilance.



HYPOTHESISED MECHANISM



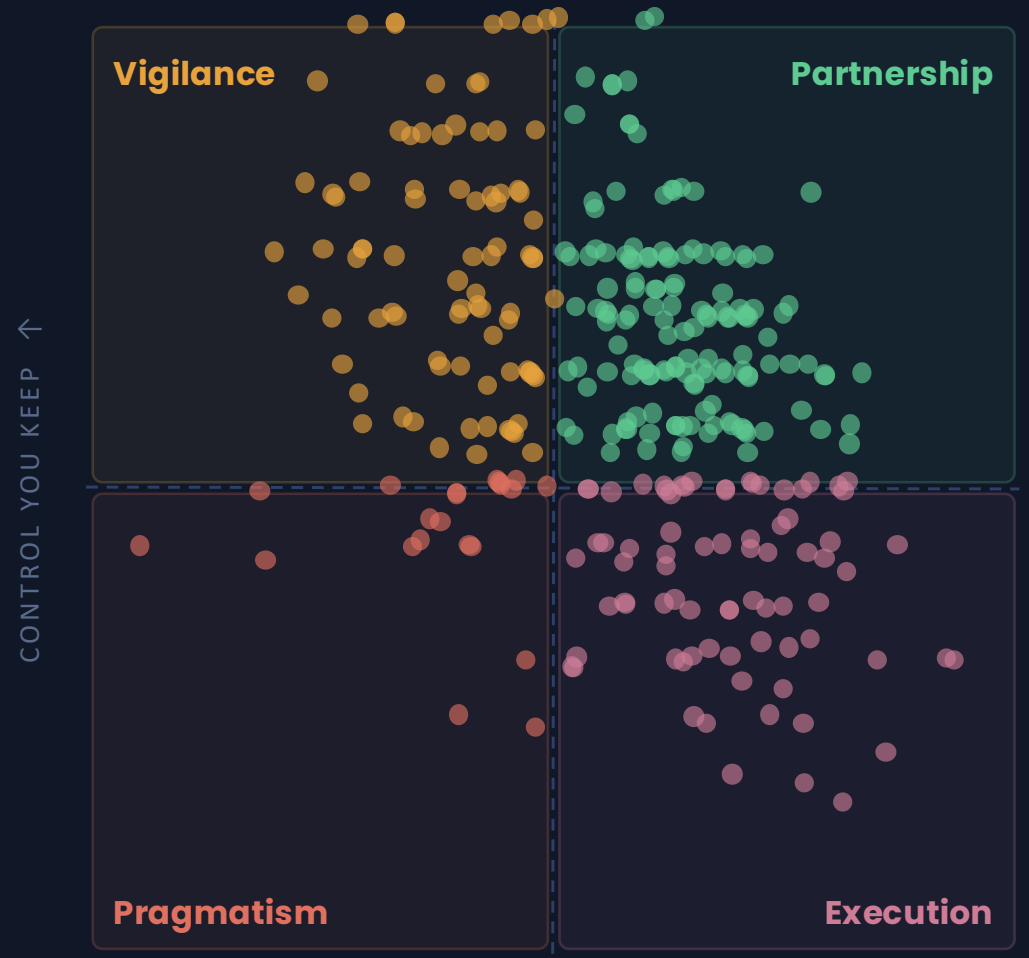
Are you sitting comfortably?

POSTURE



⇄ the same posture at its cost · dots mark the hypothesised dominant benefit or cost of each posture, not a measured outcome

Survey of 325 professionals



n = 325 RESPONDENTS · EACH MEASURED ONCE · SAMPLE FLOW IN APPENDIX VI

Posture associations

NO DETECTABLE ASSOCIATION

A technical role, or not	V = .11	**p = .24**	n = 333
English or Spanish	V = .12	**p = .18**	n = 333 · EN 271 / ES 62
Which AI you use	V = .10	**p = .63**	n = 333, five tools
Working alone, or in a team	V = .10	**p = .36**	n = 333
Your perceived improvement opportunity	V = .11	**p = .46**	n = 333
Your age band	V = .12	**p = .85**	n = 109, Prolific panel

ASSOCIATED

How deeply you have adopted	V = .19	**p = .0004 · q = .008**	n = 333, five rungs
Whether the motivation is your own	V = .19	**p = .006 · q = .036**	n = 333

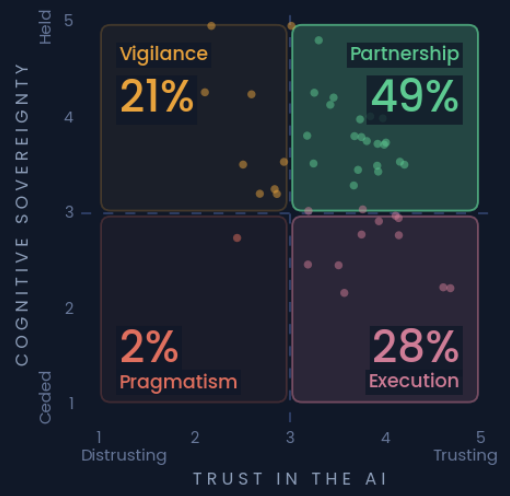
Same result without the 3.0 cutoff — continuous axes, labels permuted within recruitment source: rung → trust $\rho = +.20$ ($p = .0001$), sovereignty $\rho = -.20$ ($p < .0001$); motivation → trust $d = +.35$ ($p = .002$), sovereignty no detectable difference ($p = .79$).

Permutation chi-square against the four-way posture distribution. q is Benjamini–Hochberg over the 19-test sweep as run (Appendix III lists the family; on the 17 distinct variables motivation is q = .048). Both remain after stratifying by recruitment source (source-stratified permutation). Nulls are a failure to reject, not equivalence.

What about us?

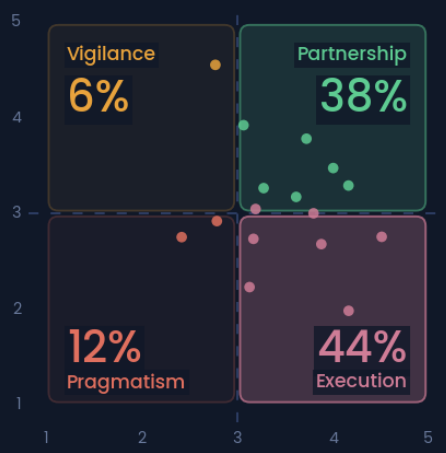
01 POSTURE

Protocolized newsletter



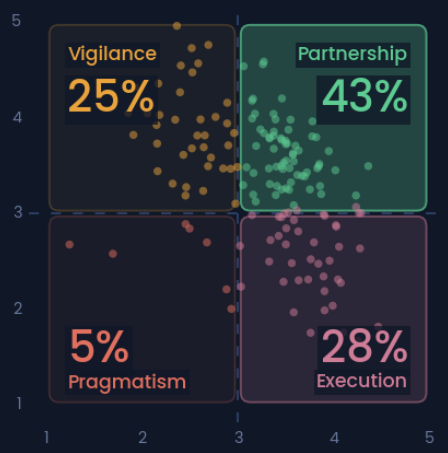
n = 43

Protocol Institute Discord



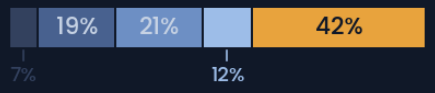
n = 16

Paid panel benchmark



n = 155

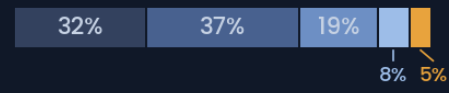
02 ADOPTION



n = 43



n = 16



n = 155

03 DIFFERENCE

posture Partnership **+6 pts**

ladder rung 4 vs 2 **+2 rungs**

n = 43 vs 155

posture Vigilance **-18 pts**

ladder rung 5 vs 2 **+3 rungs**

n = 16 vs 155

compared against

155 panel participant responses

n = 155

one respondent is worth

2 points at n = 43

6 points at n = 16

1 chatbot questions · 2 drafting · 3 running agents · 4 orchestrating · 5 building with AI

Axes 1–5, split at 3.0 (exactly 3.0 falls left / below). These are counts of everyone in each group who answered, not estimates from a sample: posture names the quadrant where a cohort and the comparison set differ most and by how many percentage points, adoption gives each one's median rung. The comparison set is the paid research panel, recruited in 3 separate waves, and those waves differ from each other. For Protocolized newsletter, measured against each of the 3 sources separately, the Partnership gap runs from +3 to +8 points and the ladder gap from +2 to +2 rungs. For Protocol Institute Discord, measured against each of the 3 sources separately, the Vigilance gap runs from -30 to -12 points and the ladder gap from +3 to +3 rungs. A gap that holds against every source is a fact about the cohort; one that swings is a fact about who else has taken the survey. Unflagged responses since 2026-07-16.

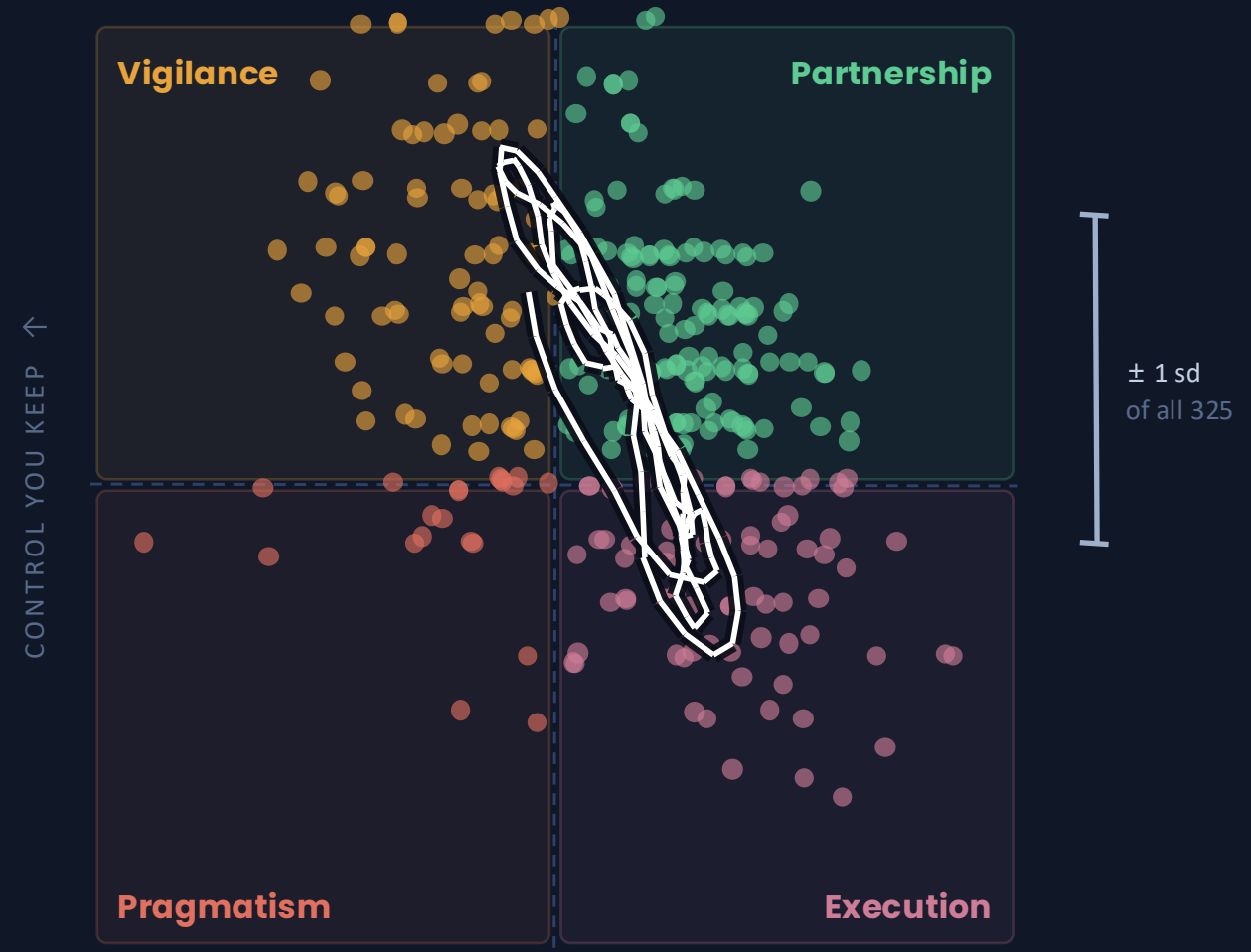
Does this change? How?

The posture machine





One hypothesised trajectory, on the survey axes

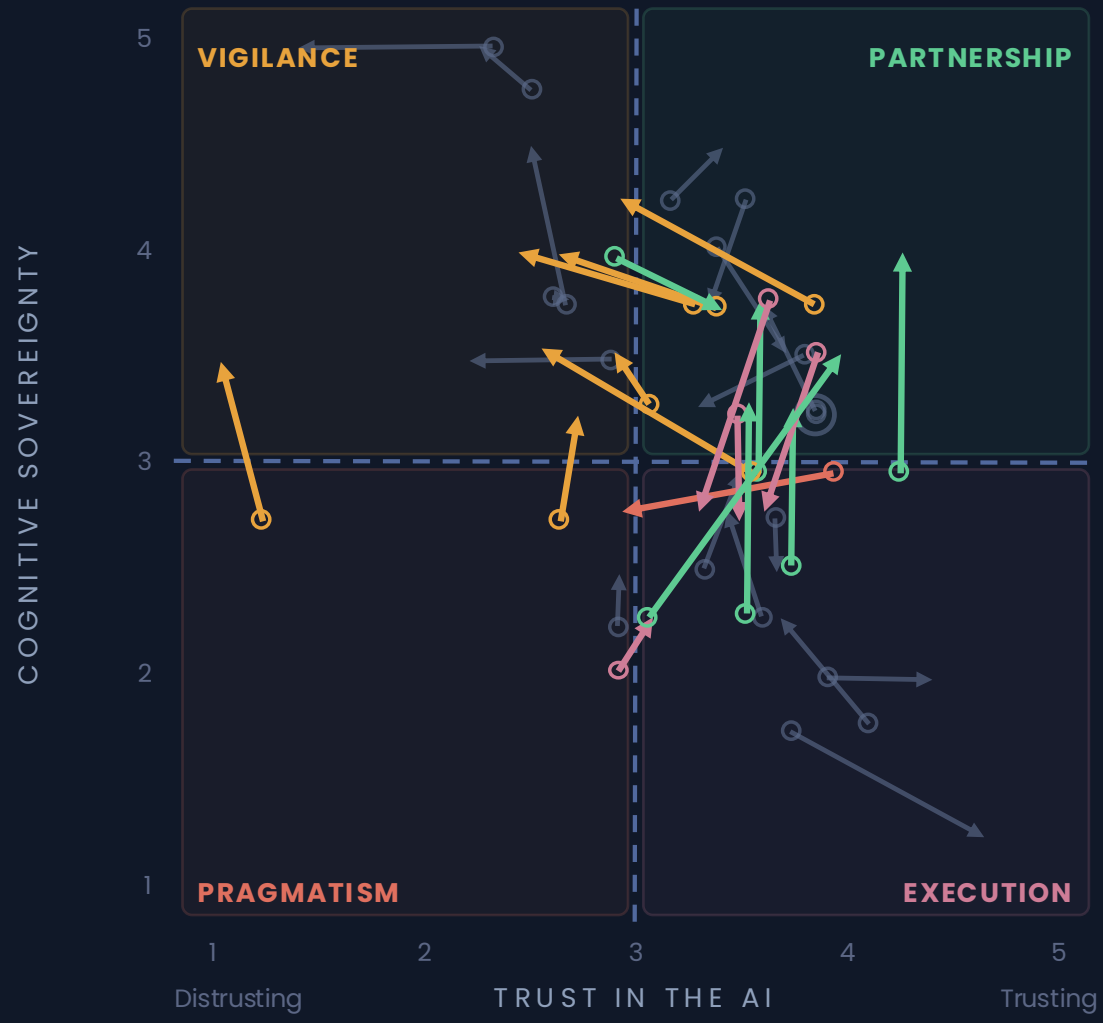


ONE SIMULATED WORKER, 300 TASKS · SIMULATED, NOT FITTED TO THE DATA · DRAWN TO SCALE

n = 325 RESPONDENTS

FOUR-WEEK STABILITY OF THE BADGE

Single axis semi-stability; quadrant instability



CHANGED POSTURE — 18 of 36

arrow coloured by the posture moved into

- Partnership
- Vigilance
- Execution
- Pragmatism

SAME POSTURE — 18 of 36

- Moved, same quadrant

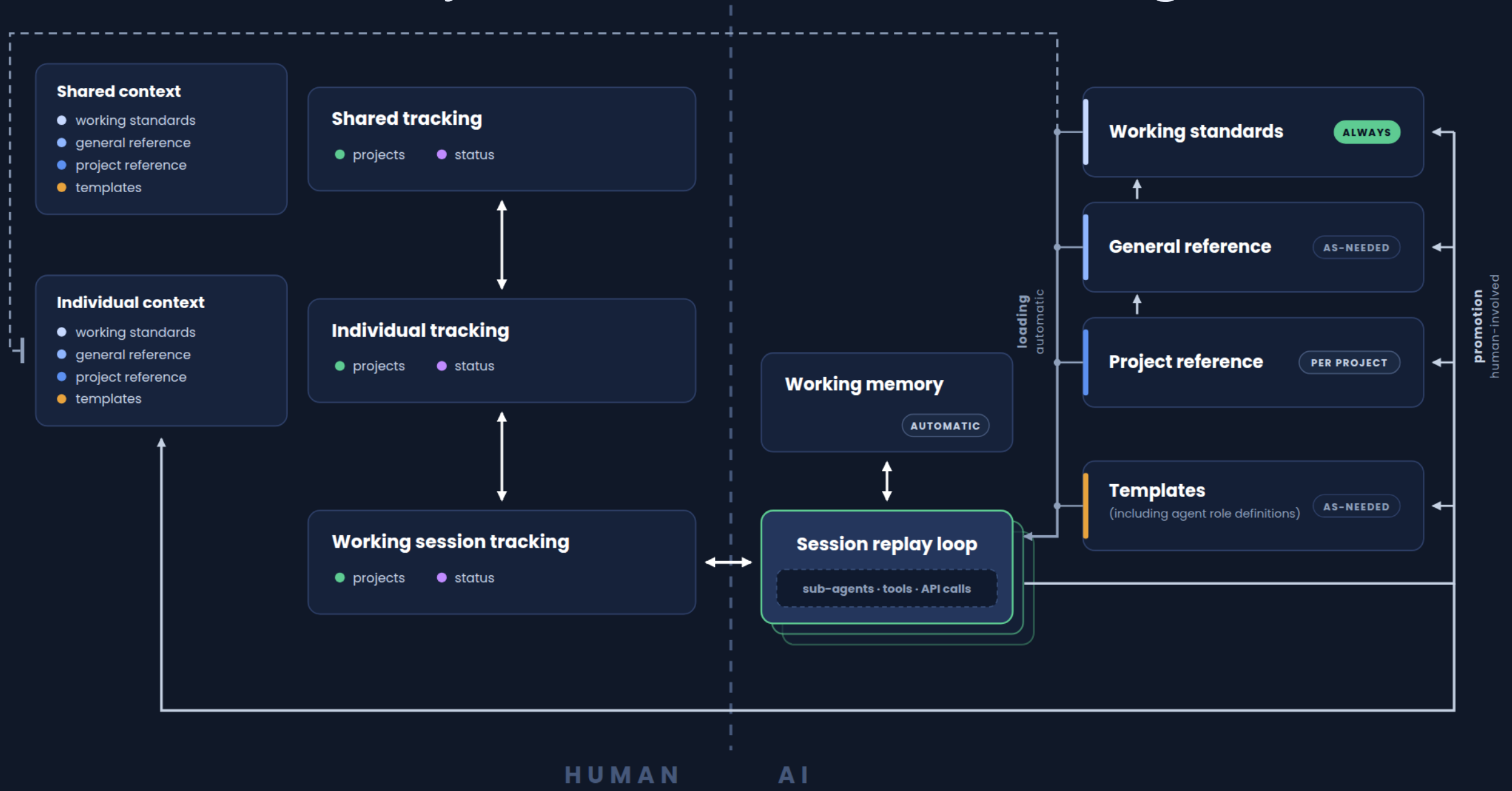
ring = first sitting · arrowhead = retest
dashed lines = the 3.0 posture threshold
n = 36 paired · 24 to 28 days apart

Retest r = .80 [.63, .89] sovereignty, .78 [.60, .88] trust; absolute-agreement ICC .78 / .75. Point estimates sit above the pre-registered .75 bar; the intervals do not. Quadrant agreement 18 of 36, κ = .30. Drift +0.19 / -0.14.
Half changing badge is some mix of state change, sensitivity to the 3.0 threshold and measurement error; attrition was selective on debt, so this sample cannot apportion them.

How do we get better?

PROPOSED MODEL

Taxonomy of Promotion for Distributed Cognition



Distributed cognition: Hutchins (1995) Cognition in the Wild · Hollan, Hutchins & Kirsh (2000) ACM TOCHI 7(2)

DRIVE

D	**Decide**	What does “done” mean?
R	**Request**	intent, context, assumptions, success criteria
I	**Iterate**	stay in the loop and steer it
V	**Validate**	confirm against done and success criteria
E	**Evolve**	debrief, then update tooling, context, and practice

Example: AI Co-Coding

BEFORE: VIBE

"Let's see what we can make"

Read code diffs line by line; load rises with volume

Vigilance decays as trust (and fatigue) rise

Looks right; hope it is

Poorly documented code, hit-and-miss results

AFTER: ENGINEERED

Decide: done = tests pass, rationale is in commit line

Request: intent, constraints, precedents and acceptance criteria

Iterate: continually realign with done state as requested

Validate: test new functionality and against regression

Evolve: new functionality becomes regression tests; retrospective

It starts with one thing...



www.robertpeake.com/ai/one-thing?gid=004e15667b27

APPENDICES

APPENDIX I

Sources

THE MEASURED RESULT

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: a systematic review and meta-analysis. Nature Human Behaviour, 8, 2293–2303. 106 experiments, 370 effect sizes; studies published 1 Jan 2020 – 30 Jun 2023. Augmentation g = +0.64; synergy g = -0.23 (95% CI -0.39 to -0.07); human-better g = +0.46, AI-better g = -0.54; decision tasks g = -0.27, creation tasks g = +0.19 (its 95% CI crosses zero; the decision-vs-creation difference is what is significant). Authors note possible publication bias.

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. (2023). Navigating the Jagged Technological Frontier. Harvard Business School Working Paper 24-013. 758 BCG consultants, preregistered, three arms. Inside the frontier, across 18 tasks: +12.2% tasks completed, 25.1% faster, >40% quality. Outside: -19 percentage points on correctness, on the one task selected to sit outside.

THE FIVE LIMITS

Social read	Nass, C., & Moon, Y. (2000). Machines and mindlessness. Journal of Social Issues 56(1) · Epley, N., Waytz, A., & Cacioppo, J. (2007). On seeing human. Psychological Review 114(4)
Grounding	Clark, H. H. (1996). Using Language. Cambridge University Press
Miscalibrated reliance	Dietvorst, B., Simmons, J., & Massey, C. (2015). Algorithm aversion. J. Experimental Psychology: General 144(1) · Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation. Org. Behavior and Human Decision Processes 151
Reliance	Lee, J. D., & See, K. A. (2004). Trust in automation. Human Factors 46(1) · Parasuraman, R., & Riley, V. (1997). Humans and automation. Human Factors 39(2)
Offloading	Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences 20(9) · Wegner, D. M. (1987). Transactive memory

ACCOUNTABILITY

Elish, M. C. (2019). Moral crumple zones. Engaging Science, Technology, and Society 5 · Matthias, A. (2004). The responsibility gap. Ethics and Information Technology 6(3) · Nissenbaum, H. (1996). Accountability in a computerized society. Science and Engineering Ethics 2(1) · Bovens, M. (2007). Analysing and assessing accountability. European Law Journal 13(4) · Spence, M. (1973). Job market signaling. Quarterly Journal of Economics 87(3)

APPENDIX II

Trust Trap Loop Provenance

PREDECESSORS

Core irony, 1983	Bainbridge, L. (1983). Ironies of automation. Automatica 19(6), 775–779. The operator loses the skill the automation was meant to free them for, and is left supervising it.
Modelled in 2023	Rinta-Kahila, T., Penttinen, E., Salovaara, A., Soliman, W., & Ruissalo, J. (2023). The vicious circles of skill erosion. JAIS 24(5), 1378–1412. Peer-reviewed causal-loop model; the erosion is latent and goes unquestioned. Modelled on accounting automation, not AI. Supports the erosion arm of the loop, not the whole cycle.

RELATIVES

Sweller (1988)	Sweller, J. (1988). Cognitive load during problem solving: effects on learning. Cognitive Science 12(2), 257–285.
Kahneman (2011)	Kahneman, D. (2011). Thinking, Fast and Slow (dual-process theory). Farrar, Straus and Giroux.
Klein & Klein (2025)	Klein, C. R., & Klein, R. (2025). The extended hollowed mind: why foundational knowledge is indispensable in the age of AI. Frontiers in Artificial Intelligence 8, 1719019.

APPENDIX III

Correlation tests

WHAT WAS TESTED	TEST	V	p	SAMPLE
Rung of adoption, five rungs	permutation	.187	**.0004**	n = 333, pooled
Motivation, low vs high	permutation	.193	**.0056**	n = 333, pooled
Sex, male vs female	permutation	.250	**.070**	n = 108, Prolific
Concurrency, four bands	permutation	.201	**.080**	n = 127, one cell of 1
Job role, collapsed to five	permutation	.184	**.176**	n = 195, since 11 Aug
Language, English vs Spanish	permutation	.122	**.176**	EN 271 vs ES 62
Technical role, or not	permutation	.112	**.241**	n = 333, pooled
Working alone vs in a team	permutation	.098	**.363**	n = 333, pooled
Job role, all 17 categories	permutation	.291	**.411**	n = 195, k too large
Headroom seen in AI, 1-5	permutation	.109	**.460**	n = 333, pooled
Collaboration mode, three	permutation	.091	**.482**	n = 333, pooled
Which AI, Claude vs rest	permutation	.085	**.483**	n = 333, pooled
Student, or not	permutation	.147	**.527**	n = 109, Prolific
Employment status	permutation	.133	**.609**	n = 109, Prolific
Which AI, all five tools	permutation	.099	**.635**	n = 333, pooled
Country, UK vs elsewhere	permutation	.110	**.760**	n = 109, Prolific
Age band, 3 groups	permutation	.115	**.853**	n = 109, Prolific

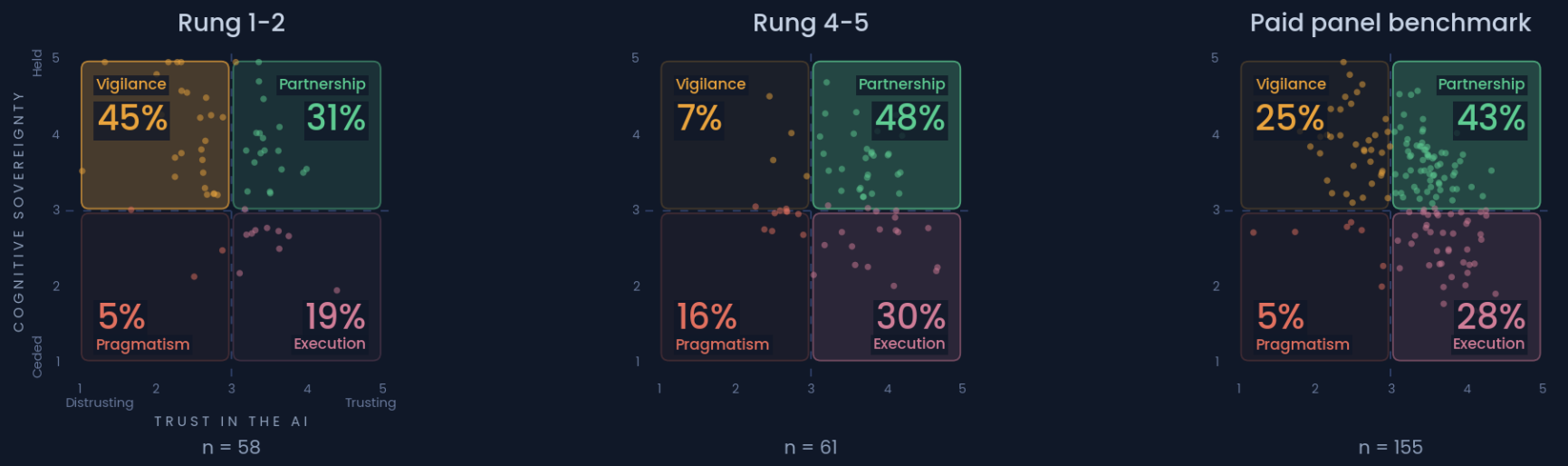
Amber rejects and survives BH. The BH family is the 17 rows above plus two re-cuts of the same variables run in the sweep: rung below 3 vs 3+ (V .22, p .0009) and motivation on five levels (V .16, p .015) — 19 tests. q adjusts each p by its rank in that family: rung .0004 → q = .008; motivation .0056 (rank 3) → q = .036; on the 17 distinct variables alone motivation is rank 2, q = .048. Both survive either way. Dropped as untestable: modality, feedback delay (n = 49), usage frequency (n = 71).

Robustness: every verdict holds on the wider eligibility set — primary plus single-miss rows, n = 355. No row changes side. Both rejections remain after stratifying by recruitment source (source-stratified permutation) and on the continuous axes (Appendix VI).

The 110 Prolific-linked respondents show no detected differences from the 45 unlinked on five observed demographics (all p ≥ .35); selection on unmeasured characteristics remains possible.

Vigilance is less common at higher adoption rungs

01 POSTURE



02 ADOPTION



03 DIFFERENCE

posture	Vigilance	+20 pts
ladder	rung 2 vs 2	0 rungs

n = 58 vs 155

posture	Vigilance	-18 pts
ladder	rung 5 vs 2	+3 rungs

n = 61 vs 155

compared against

155 panel participant responses

n = 155

one respondent is worth

2 points at n = 58

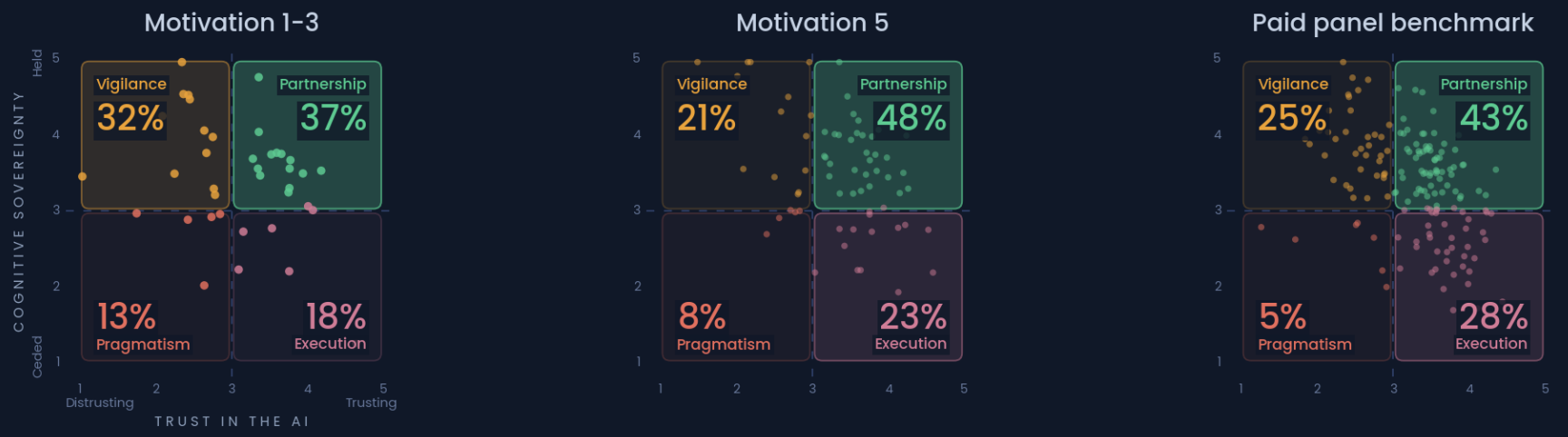
2 points at n = 61

1 chatbot questions · 2 drafting · 3 running agents · 4 orchestrating · 5 building with AI

Axes 1–5, split at 3.0 (exactly 3.0 falls left / below). These are counts of everyone in each group who answered, not estimates from a sample: posture names the quadrant where a cohort and the comparison set differ most and by how many percentage points, adoption gives each one's median rung. The comparison set is the paid research panel, recruited in 3 separate waves, and those waves differ from each other. For Rung 1–2, measured against each of the 3 sources separately, the Vigilance gap runs from +8 to +27 points and the ladder gap from +0 to +0 rungs. For Rung 4–5, measured against each of the 3 sources separately, the Vigilance gap runs from -30 to -11 points and the ladder gap from +3 to +3 rungs. A gap that holds against every source is a fact about the cohort; one that swings is a fact about who else has taken the survey. Unflagged responses since 2026-07-16.

Intrinsic motivation is associated with more partnership

01 POSTURE



02 ADOPTION



03 DIFFERENCE

posture	Execution	-9 pts
ladder	rung 3 vs 2	+1 rungs

n = 38 vs 155

posture	Partnership	+5 pts
ladder	rung 3 vs 2	+1 rungs

n = 66 vs 155

compared against

155 panel participant responses

n = 155

one respondent is worth

3 points at n = 38

2 points at n = 66

1 chatbot questions · 2 drafting · 3 running agents · 4 orchestrating · 5 building with AI

Axes 1–5, split at 3.0 (exactly 3.0 falls left / below). These are counts of everyone in each group who answered, not estimates from a sample: posture names the quadrant where a cohort and the comparison set differ most and by how many percentage points, adoption gives each one's median rung. The comparison set is the paid research panel, recruited in 3 separate waves, and those waves differ from each other. For Motivation 1-3, measured against each of the 3 sources separately, the Execution gap runs from -13 to -2 points and the ladder gap from +1 to +1 rungs. For Motivation 5, measured against each of the 3 sources separately, the Partnership gap runs from +3 to +8 points and the ladder gap from +1 to +1 rungs. A gap that holds against every source is a fact about the cohort; one that swings is a fact about who else has taken the survey. Unflagged responses since 2026-07-16.

APPENDIX V

Cognitive Debt

WHERE THE EVIDENCE CUTS THE OTHER WAY: AI AS SKILLING

Positive meta-analysis	Wu, M., & Ouyang, F. (2026). How generative AI influences learning outcomes in programming education: a three-level Bayesian meta-analysis. Educational Psychology Review 38, Article 117. 35 studies, 131 effect sizes; small-to-medium positive effects, including unassisted performance. The authors stress a small, uneven corpus.
Adaptive sequencing	Chung, A. T.-H., Zhang, B., Kung, L.-C., Bastani, H., & Bastani, O. (2026). Effective personalized AI tutors via LLM-guided reinforcement learning. arXiv:2608.16907 (also SSRN 6423358). Adaptive vs fixed sequencing, both arms on the tutor: +0.15 SD on unaided exam performance.

DESIGN DETERMINES THE DEGREE OF DEBT

Guardrails, not the model	Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning. PNAS 122(26). Unguarded: +48% assisted, -17% later unaided. Tutor arm: +127% assisted, later unaided indistinguishable from control.
Passive vs active use	Lee, E. H., Yin, Y., Jia, N., & Wakslak, C. J. (2026). Relying on AI at work reduces self-efficacy, ownership and meaning while active collaboration mitigates the effects. Scientific Reports 16(1), 13583. The outcomes are self-efficacy, ownership and meaning — not unaided performance.
Persistence under assistance	Liu, G., Christian, B., Dumbalska, T., Bakker, M. A., & Dubey, R. (2026). AI assistance reduces persistence and hurts independent performance. arXiv:2604.04721, v4, 5 Aug 2026. Preprint; brief exposure; the authors caution against broad or longitudinal generalisation.

APPENDIX VI

Instrument and sample

WHAT THE INSTRUMENT IS

Items and scoring

Five-point items. Sovereignty = mean of the four-item core (ag1b, cede1b, fd1b, ov1b). Trust = ½ anchor + ½ mean of the trust-tail items, which include the three overtrust items the factor analysis could not place. Posture splits both at 3.0, the scale midpoint (exactly 3.0 falls low). Quadrants are a reading aid over two continuous axes; the primary associations are re-tested on the axes.

Factor structure

Two well-identified factors, trust disposition and sovereignty; a verification doublet apart from both; three overtrust items the data cannot place. 2 vs 3 factors is unsettled. n = 145 post-freeze, 1,000 bootstraps (2026-09-14).

Reliability

McDonald's ω on the four-item cores: trust .785 [.70, .84]; sovereignty .774 [.69, .84]. Retest at 24–28 days, n = 36: r .80 / .78, ICC(A,1) .78 / .75, quadrant κ = .30.

Without the cutoff

Source-stratified permutation on the continuous axes, n = 333: rung → trust ρ = +.20 (p = .0001), sovereignty ρ = -.20 (p < .0001); motivation (1–3 vs 4–5) → trust d = +.35 (p = .002), sovereignty n.s. (p = .79). Retest κ = .30 says the badge is the fragile object; the axes are the finding.

Not yet established

Measurement invariance across language or cohort (a cross-language null is not invariance). Criterion validity against observed verification or reliance behaviour — the instrument cannot see whether your checking is any good.

WHERE THE NUMBERS COME FROM

355 eligible incl. single-miss rows → 333 primary, pooled EN 271 + ES 62 (the covariate tests) → 325 scored on both axes after the 16 July sovereignty re-point (the scatter) → 109 Prolific-linked with demographics → 36 retest pairs. The "paid panel benchmark" is three Prolific first-sitting waves, n = 155, not a probability sample.