Monitoring Under Scarcity

From information to causal use in resource-bounded intelligent systems

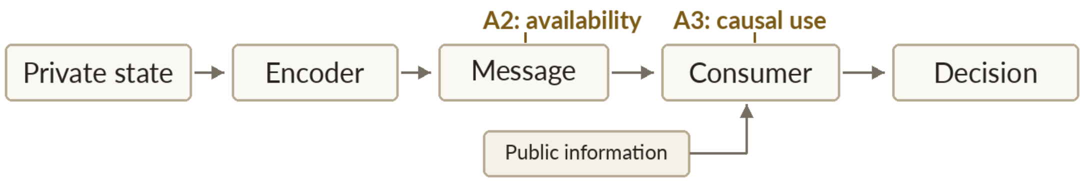
BEN AMUWO · ben.amuwo@gmail.com · Vancouver, Canada
Independent computational research + proposed Master's-scale extension · sidus.quest

STATUS · E0-A: completed measurement-validation study · E1: proposed Master's study · broader Sīdus programme: outside thesis scope · Version 1.0 · September 2026

RESEARCH QUESTION

How can we distinguish information that is merely *present* in a monitoring channel from information a system actually *uses* — and what happens to that distinction when useful monitors must compete for limited processing capacity?

Claims about self-monitoring and metacognition can slide between two different kinds of evidence: a signal may carry information about an internal state, yet never change downstream behaviour. E0-A was built as a measurement stress test for that distinction in a deliberately small system where the correct answer is known. Its purpose was not to establish metarepresentation; it was to verify the instrument before using it where ground truth is unavailable.



A2 asks whether the message carries hidden-state information. A3 asks whether replacing the message changes behaviour. The distinction is the load-bearing result.

COMPLETED STUDY — E0-A

DESIGN

- **Ground truth.** Finite task with analytically known optimal behaviour; every reported quantity is computed by exact enumeration.
- **Four gates.** Numeric thresholds fixed in advance for hiddenness, transmission, causal use, and decline of added benefit as the answer becomes public.
- **Eight adversarial controls.** Each is designed to pass some criteria and fail others.
- **500 training initializations.** All outcomes retained; sample size fixed before the registered run.
- **Frozen provenance.** Configuration, pre-execution commitment record, and a public archive of 19,502 artifacts.

RESULT

All 500 initializations met all four acceptance criteria, transmitting the full 0.5 bits hidden from the public channel and raising expected reward from 0.75 to 1.0.

The decisive control was informative-but-unused. It carried the same 0.5 bits as a successful system while producing zero behavioural benefit and zero effect under message replacement. It passed A2 and failed A3, as designed.

What the 500 seeds bought: an estimate of how reliably the training procedure reaches the analytically known solution despite optimization failure modes. They are an optimization-reliability claim, not extra statistical evidence for an exactly enumerated monitoring effect.

WHAT DID NOT SUCCEED

An initial four-way comparison of training procedures was inconclusive: every procedure recorded zero collapses across all twenty condition-runs (the independent units were five seeds per procedure, not twenty trials). The procedure carried forward was chosen by a declared tie-break, explicitly recorded as not evidence of superiority. A later paired stress test over 100 seeds separated the arms (0/100 versus 20/100, 53/100 and 9/100), but zero observed failures still does not establish zero failure probability (95% Wilson upper bound ≈3.7%).

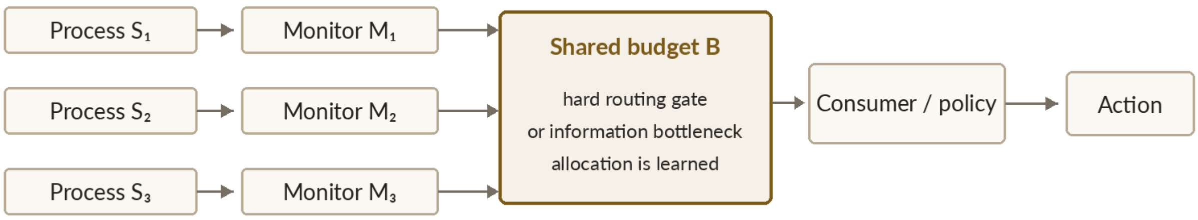
SCOPE OF THE RESULT

ESTABLISHED	NOT ESTABLISHED	NEXT
The protocol separates information availability from causal use across known ground-truth cases, reliably across training initializations.	Metarepresentation; generality to realistic architectures; an ecology of interacting channels; or any claim about biological cognition.	Introduce genuine competition among useful monitoring channels under a shared capacity budget, while preserving the A2/A3 distinction.

PROPOSED MASTER'S STUDY — E1: MONITORING UNDER GENUINE SCARCITY

When several useful monitoring channels share a limited budget, what determines which information retains causal influence?

E0-A studied one affordable channel at a time. E1 changes the scientific problem: multiple useful monitors are trained together under a bottleneck, so information can no longer be retained and routed without trade-offs. The central object is not "how much information exists?" but how availability, marginal behavioural value, redundancy and causal use interact when capacity is scarce.



Two scarcity regimes are intentionally separable: an encoding bottleneck can remove information itself; a routing bottleneck can leave information present while preventing it from controlling behaviour.

CORE MEASUREMENTS

AVAILABILITY	CAUSAL USE	SCARCITY
$I(S_i; M_i)$	**$\Delta R_i = R - R(\text{replace } M_i)$**	**budget B**
How much target-state information remains in monitor i?	How much does behaviour depend on that monitor?	How much encoding or routing capacity can the system allocate across monitors?

PRIMARY TEST — DOES THE DISSOCIATION TRACK THE LOCUS OF SCARCITY?

H3 — Bottleneck location leaves a measurable signature. Encoding scarcity should reduce information availability and causal use together. Routing scarcity should permit informative-but-unused states: A2 can remain high while A3 falls. The non-trivial test is whether this measured dissociation tracks the known structural location of scarcity across independently trained systems. Because learned systems can compensate, reroute or redundantly encode information, the signature is not guaranteed simply by imposing a constraint.

SUPPORTING ALLOCATION PREDICTIONS

- **H1 — Value over raw information.** When channels are matched for information rate but differ in intervention-derived behavioural benefit, the higher-benefit channel should retain more capacity or causal influence.
- **H2 — Redundancy is discounted.** A channel's conditional contribution given the other channels should predict retention better than its standalone mutual information. Redundant monitors should be lost before complementary monitors of comparable standalone value.
- **H4 — Scarcity should produce selective degradation.** As the budget tightens, low marginal-value or redundant channels should lose influence before high marginal-value channels. Uniform degradation would count against the proposed selection principle.

H1, H2 and H4 constrain the allocation model; none is the novelty claim alone. Their role is to test whether value and redundancy explain when H3 appears better than generic compression or arbitrary gating.

EXPERIMENTAL PLAN

- **1. Ground-truth scarcity.** Begin with a small multi-channel environment where the optimal allocation can still be computed or tightly bounded. Cross information rate, behavioural value, redundancy and budget.
- **2. Learned representations.** Replace constructed private variables with readouts from internal states learned by a trained network. This is the point at which the toy assumptions are expected to break, and those failures are scientifically useful.
- **3. Separate encoding from routing.** Use at least one explicit information bottleneck and one hard/soft routing constraint rather than treating "scarcity" as a single mechanism.
- **4. Preserve optimization diagnostics.** Train separate systems at each budget condition; retain all initializations; keep optimization failure distinct from a negative scientific result; preregister analysis and failure criteria before the registered run.

WHAT WOULD CHANGE MY MIND?

The account is weakened if generic compression or random gating predicts allocation equally well; if value and redundancy add no predictive power; if the H3 dissociation fails; or if optimization failure carries the effect. A null selective-allocation result is therefore informative, not a failed thesis.

COGNITIVE AND NEUROSCIENTIFIC BRIDGE

A bridge by computational structure — not an anatomical identification.

E0-A establishes nothing about biological cognition. E1 nevertheless lands on a question that already exists in cognitive neuroscience: limited systems must decide which information to preserve, amplify, route or act on. The strongest bridge is therefore not "this channel is dACC" or "this is metacognition in the brain." It is a shared resource-allocation problem with measurable dissociations between representation and control.

E1 quantity	Cognitive analogue	Relevant literature	Bridge question
Finite budget	Limited attentional / processing capacity	Selective routing of behaviourally relevant neural representations [2]; rational inattention [4]	Does scarcity selectively preserve information with higher task value rather than simply reducing all signals?
Availability vs use	Uncertainty/confidence signal vs metacognitive control	Confidence can be represented yet its downstream use is a separate problem [1,5,6]	Is a monitor merely decodable, or does it regulate later sampling, effort or action?
Marginal benefit and cost	Control allocation	Expected-value-of-control and metacognitive-control models [3,6]	Does control go where expected benefit justifies cognitive cost?
Context and redundancy	Goal-dependent information sampling	Narrative context shifts gaze toward semantic salience [7]	Can contextual value and redundancy predict which information is sampled under a tighter budget?

RADBOUD/DONDERS: A CONCRETE EMPIRICAL BRIDGE

Berlot and colleagues at the Donders Institute showed that coherent narrative context shifts gaze away from purely visual salience and toward semantically salient objects: what observers sample depends on what the unfolding context makes useful for understanding [7]. Their Discussion goes further than the general claim. Beyond surprisal, they explicitly identify stimulus importance, relevance and entropy as factors that may exert additional or different influences on gaze, and ask how different forms of salience interact with task demands.

Those variables offer candidate empirical counterparts to E1 without requiring a one-to-one mapping: relevance to marginal behavioural value, entropy to information structure, and task demands to the effective budget on sampling. A later bridge could impose or exploit a measurable sampling constraint and test whether E1-derived predictors — value, redundancy, availability and budget — explain which information receives scarce gaze or inspection time. That would be a test of information allocation in human cognition, not by itself a test of metarepresentation.

WHERE NEUROSCIENCE ENTERS — AND WHERE IT DOES NOT

Neural measurement should follow a validated behavioural/computational model, not substitute for one. Attention research provides evidence that selected representations are strengthened and preferentially routed through large-scale networks [2], while control theories propose value-sensitive allocation of control [3]. Metacognition and confidence provide a second bridge because a signal may be represented or reportable without necessarily regulating later sampling, effort or action [1,5,6] — the cognitive analogue of E0-A's availability/use distinction.

If a behavioural bridge succeeds, model-derived allocation variables could then be tested against fMRI, MEG/EEG, pupil or gaze measures under supervision. Region-level mappings would remain hypotheses rather than premises: the bridge is by computational structure, not anatomical identification.

WHY THIS IS A MASTER'S PROJECT

E0-A identified a specific transition point between work I can validate by exact construction and work that requires formal supervision: stochastic estimation, learned representations I did not design, model comparison under real variability, and eventually behavioural or neural data. The proposed thesis is deliberately sized around that transition.

CORE THESIS	BRIDGE / STRETCH
– Registered E1 computational study with explicit scarcity. – Ground-truth and learned-state phases. – Open, reproducible archive with failure cases retained. – Primary outcome: which factors predict preserved causal influence under budget.	– Human behavioural pilot or secondary-data analysis if supervision, ethics and time permit. – Naturalistic attention bridge as a candidate extension, not a required deliverable. – Neural data only after the behavioural model earns the right to generate neural hypotheses.

TRAINING I AM SEEKING

The project needs machine learning and computational modelling taught beyond self-directed approximation; statistics and experimental design strong enough for preregistered inference rather than exact enumeration; and cognitive/neuroscientific grounding that constrains which analogies to biological systems are legitimate. The value of the Master's is not that it validates Sīdus. It is that it provides the methods and supervision required to let the next experiment outrun the toy system without outrunning the evidence.

BACKGROUND

B.Sc. Biology with a Psychology minor (University of British Columbia), followed by professional work in health systems and integrated youth services. Alongside that work I taught myself full-stack development and applied machine learning, then built Sīdus independently as a way to force broad questions about cognition into explicit measurements, controls and failure conditions.

SELECTED REFERENCES

[1] Fleming, S. M. (2024). Metacognition and Confidence: A Review and Synthesis. Annual Review of Psychology, 75, 241–268. doi:10.1146/annurev-psych-022423-032425.

[2] Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. Neuron, 88(1), 127–144. doi:10.1016/j.neuron.2015.09.017.

[3] Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. Neuron, 79(2), 217–240. doi:10.1016/j.neuron.2013.07.007.

[4] Gershman, S. J., & Burke, T. (2023). Mental Control of Uncertainty. Cognitive, Affective, & Behavioral Neuroscience, 23, 465–475. doi:10.3758/s13415-022-01034-8.

[5] Aguilar-Lleyda, D., & de Gardelle, V. (2021). Confidence Guides Priority Between Forthcoming Tasks. Scientific Reports, 11, 18320. doi:10.1038/s41598-021-97884-2.

[6] Bénon, J., Lee, D., Hopper, W., et al. (2024). The Online Metacognitive Control of Decisions. Communications Psychology, 2, 23. doi:10.1038/s44271-024-00071-y.

[7] Berlot, E., Schmitt, L.-M., Huber-Huber, C., Peelen, M. V., & de Lange, F. P. (2026). Narrative Context Shifts Gaze from Visual to Semantic Salience. Communications Psychology, 4, 59. doi:10.1038/s44271-026-00426-7.

WHERE TO LOOK

E0-A experiment and results
Master's scoping document
Full Sīdus programme and archive