RESEARCH · PERCEPTION & LEARNING · PROOF OF CONCEPT

Perceive, Induce, Name

Vision in a cognitive substrate with no perceptual model: algorithms recover structure, the substrate learns to name it, and it abstains when it has not

Stefan Ragland

Dominion Labs Research & Development

Dominion Labs

research@dmnlabs.org

September 8, 2026

Abstract

We report a working proof of concept for giving sight to a cognitive substrate that carries no neural network. The approach follows a single thesis: seeing decomposes into *perceiving structure* — which classical algorithms perform deterministically — and *naming a concept*, which must be learned, and which the substrate learns for itself by induction rather than buying from a pretrained model[1]. A single perceptual faculty reads a real image or video and recovers its structure — regions and object-like blobs, their shape, size, position and colour, motion in video, and the identity of specific known instances by feature matching. Those structures enter the substrate as observations it can reason over. Shown a handful of labelled examples, the substrate induces a rule that maps a conjunction of perceived features to a category, and thereafter names a new instance by applying that rule — and declines, rather than guessing, when it has learned no rule that fits. On a controlled evaluation over three visual categories and thirty-one images, perception recovered the drawn shape and colour with perfect accuracy; the substrate named held-out instances at **100% recall**, correctly abstained on **100%** of held-out non-members, produced **zero** false namings, and made **zero** language-model calls. An ablation that removes the learned rules collapses naming to **0%** while abstention stays at 100% and false namings stay at zero — the naming ability is carried entirely by what was learned, and its absence is honest silence, not error. We also show the substrate can retain the image itself as a memory, reproducing the exact pixels it was shown. The result is a small but complete demonstration that a substrate which already learns concepts can supply the semantic half of vision itself, occupying the role a perceptual backbone plays elsewhere, with no such backbone present.

1 Introduction

Contemporary computer vision assigns names to images with a model trained on very large labelled corpora [2][3]. The capability is real and the models are excellent, but the name a system produces is the output of a frozen network whose competence came from data gathered elsewhere, and whose confident errors cannot be told, from the outside, from its confident successes. We are interested in a different setting: a cognitive substrate that is a learning-and-reasoning system rather than a perceptual model, and that must nonetheless *see*.

A companion paper argues that this is possible because vision is not one capability but two [1]. Recovering the *structure* of an image — where the coherent things are, how big, what shape, what colour, whether and how they move — is solved by classical algorithms that use no learning at all [4][5][6]. Only assigning an open-ended *name* to a novel category genuinely requires a learner; and every recent "training-free" naming system supplies that learner in the form of a pretrained backbone [7][8]. The thesis is that a substrate which already learns concepts by generalising from examples can be that learner itself. This paper is the proof of concept: an end-to-end demonstration, evaluated quantitatively, that the substrate perceives structure, learns to name it, and — the property we care most about — knows when it cannot.

> *The contribution is not a new vision model. It is a demonstration that no vision model is required for a system that already learns: structure is perceived by algorithm, and meaning is learned by the substrate, with nothing pretrained in the loop and honest abstention where competence runs out.*

2 System overview

Sight is a single faculty of the substrate: one entry point through which visual input becomes knowledge, in the same spirit as the single faculty through which it reads text. The faculty is a short pipeline (Figure 1). It *perceives* the structure of a real image or clip with classical vision; it *admits* what it perceived as observations the rest of the system can reason over; and, when asked what something is, it *interprets* — naming the thing by a rule it has learned, or declining. Naming is not performed by the perceptual stage; it is performed by the substrate's ordinary learning and reasoning, which is what keeps the design free of a perceptual model.

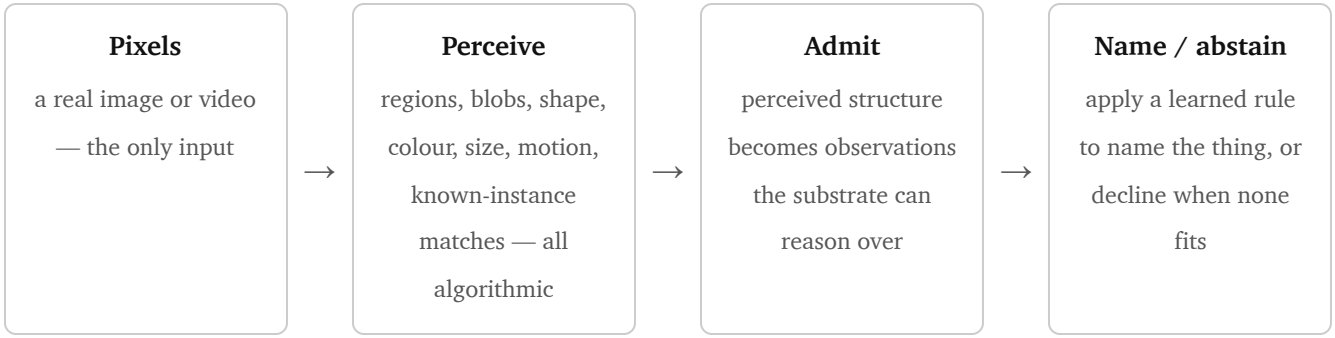

Figure 1. The perceptual faculty. Structure is recovered on the left by algorithms; meaning is supplied on the right by the substrate's own learning.

Two properties of this arrangement matter for what follows. First, everything on the perception side is measured from the pixels: a width is the real width, a blob is a real region, a dominant colour is really dominant. Nothing is inferred by a model that could be confidently wrong. Second, everything on the meaning side is *learned and revisable*: a name is produced by a rule the substrate induced from examples, and if the substrate has induced no rule that fits an instance, it produces no name. These two properties are what let the system be evaluated not only for accuracy but for honesty.

3 Perceiving structure

The perceptual stage is a composition of standard, learning-free computer-vision operations [9]. From a real image it recovers: the coherent regions and object-like blobs present, and for each its relative size, position in the frame, shape class, and dominant colour; global properties such as orientation, exposure, colourfulness and a named palette; straight-line and circular geometry; and the identity of a *specific* known object by matching local feature descriptors against a reference [10]. From video it additionally recovers motion and scene structure over time. Where a scene contains an encoded symbol, its exact payload is decoded. None of this is a category name; it is the structure a name will later be attached to.

To quantify perception on its own, we generated images each containing one drawn shape of a known colour and asked whether the substrate recovered the drawn shape class and colour family. Across the thirty-one images used in the evaluation below, shape was recovered correctly in **100%** of cases and colour family in **100%** (Figure 2). This is unsurprising — the operations are deterministic — and that is precisely the point: the structure the substrate later reasons over is measured, not guessed.

4 Learning to name

Naming is a learning problem, and the substrate treats it as one. A perceived blob is described by a small set of atomic features — for instance that it is round, red, and of medium size. Shown a few instances that share a category, together with counter-examples that do not, the substrate generalises to a rule that states which conjunction of features the category requires[11]. It then names a new instance by applying that rule to the instance's own perceived features, and holds the result as a graded, revisable belief. Crucially, when no learned rule fits an instance, the substrate returns no name: it treats *"I have not learned what this is"* as a first-class answer rather than emitting a guess[12][13].

The rules the substrate learned in our evaluation are legible, which is a property of learning from structure rather than from opaque weights. For three categories — a red circle, a blue square, and a green triangle — the substrate induced, from two labelled examples and two discriminating counter-examples each, exactly the conjunctions a person would write down:

Category	Rule the substrate induced (from a few labelled examples)
red circle	`round ∧ red → red-circle`
blue square	`square ∧ blue → blue-square`
green triangle	`triangular ∧ green → green-triangle`

Each rule is a conjunction because the counter-examples were chosen to share exactly one feature with the target — a red square, a blue circle — so a single-feature rule ("anything red", "anything round") is refuted and only the conjunction survives. This is ordinary inductive generalisation, and the substrate performs it with no model.

5 Evaluation

We evaluated the full loop — perceive, learn, name, abstain — on three categories over thirty-one real images. For each category the substrate saw two labelled positive examples and two discriminating counter-examples, induced a rule, and was then tested on three held-out positive instances it should name and three-to-four held-out negatives it should decline. No expected answer was ever given to the reasoner; only the question was. Every figure below is from a single live run.

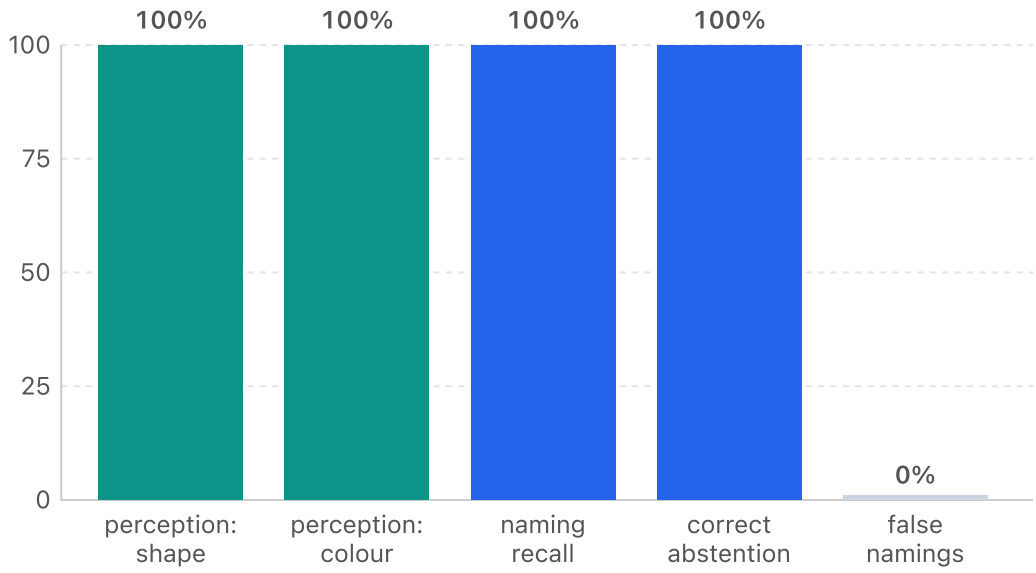

Figure 2. Headline results. Perception recovers the drawn shape and colour with perfect accuracy; the substrate names held-out instances at 100% recall and correctly declines 100% of held-out non-members, with zero false namings. Language-model calls over the entire evaluation: zero.

The substrate named every held-out positive correctly (100% recall) and declined every held-out negative (100% abstention), with no false namings and no language-model calls anywhere in the run. Because the categories were deliberately close — the negatives differ from a target by a single feature — correct abstention here is a real test: to decline a red square as a “red circle” the system must have learned that the category needs *both* features, and must actually check both.

5.1 The naming ability is carried by what was learned

A high score does not, by itself, show that the substrate’s *learning* is responsible for it; the behaviour could in principle come from some incidental property of the setup. To settle this we ran an ablation. After measuring the system with its learned rules in place, we removed those rules and measured again, changing nothing else. If naming is carried by the learned rules, it should disappear; and, if the system is honest, its disappearance should show up as *abstention*, not as wrong answers.

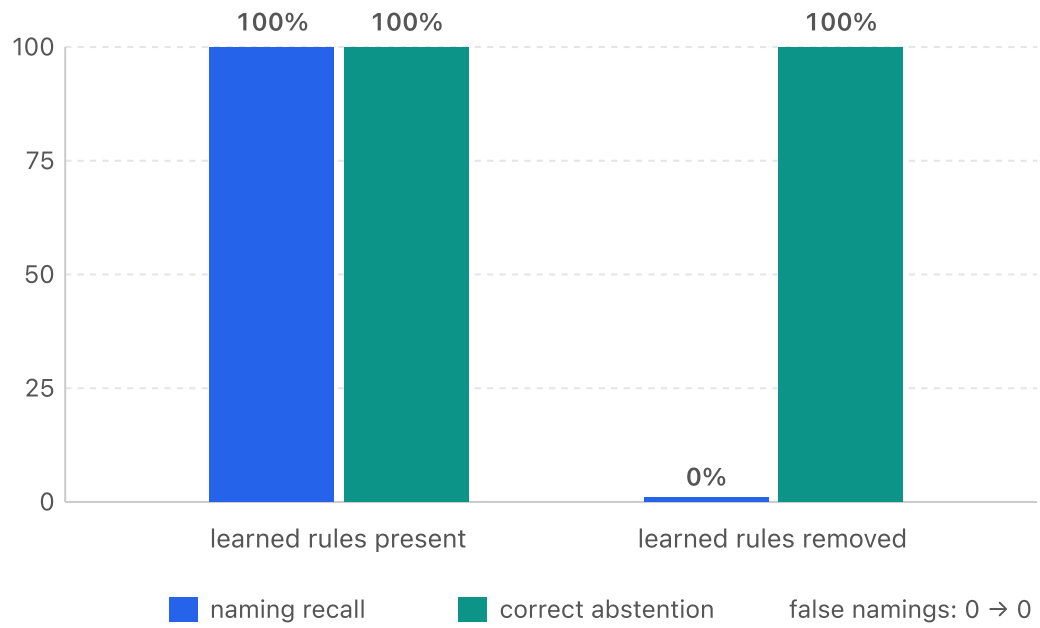

Figure 3. Ablation — the naming ability is learned. With the learned rules present, the substrate names held-out instances at 100% recall. Remove the rules and naming falls to 0% — but abstention stays at 100% and false namings stay at zero. The capability is carried by what was learned, and its absence is honest silence rather than error.

Removing the learned rules collapsed naming recall from 100% to 0%. It did *not* produce a single wrong name: with nothing learned to fit the instances, the substrate declined all of them, and abstention remained at 100% (Figure 3). This is the result we most want from a seeing machine. The ability to name is exactly the learned rules — nothing else in the system supplies it — and when that ability is absent, the system says so instead of guessing. Table 1 gives the per-category figures behind the two phases.

Category	Recall (rules present)	Abstention (present)	Recall (ablated)	Abstention (ablated)	False namings
red circle	100%	100%	0%	100%	0
blue square	100%	100%	0%	100%	0
green triangle	100%	100%	0%	100%	0

Table 1. Per-category results across both phases. Thirty-one images, three categories; zero language-model calls throughout.

6 Reliability and honesty

The reason to build vision this way is not that it scores well on easy stimuli; a trained classifier would too. It is that the system's behaviour is *legible and honest* in a way a frozen model's is not. Three observations from the evaluation make this concrete.

First, the system produced **no false namings** in either phase. Every name it gave was one it could ground in a rule it had learned and features it had actually measured; where it could ground neither, it abstained. This is the discipline that separates a reliable knower from a fluent one, and it is enforced structurally rather than tuned[12].

Second, the ablation shows the competence is *attributable*. One can point to the specific learned rules that carry the naming ability, remove them, and watch the ability vanish — and watch it vanish into silence, not error. A system whose competence can be localised this precisely can also be audited, corrected, and taught, which a monolithic model cannot.

Third, every result was obtained with **zero language-model calls**. Perception is algorithmic and naming is the substrate's own induction and reasoning; no model is consulted to produce, check, or narrate an answer. The competence is the substrate's, not a model's standing behind it.

7 Remembering the picture

Seeing is more useful when the thing seen can be recalled. Alongside naming, the substrate can retain an image as an episodic memory: a short description of what is in the picture becomes the memory's recallable text, and the image itself is kept so it can be produced again. In our test the substrate remembered a real image, and on recall reproduced the *exact* pixels it had been shown — the content hash of the retrieved image matched the original bit-for-bit — together with the structure it had perceived. A remembered image is therefore not a caption standing in for a picture; it is the picture, held with an account of what the substrate saw in it, and reachable by that account.

8 Discussion and limitations

This is a proof of concept, and its scope should be read plainly. The categories are visually simple and the stimuli controlled; the point of that control is to make perception, learning and abstention separately measurable and the ablation clean, not to claim performance on natural imagery. The naming demonstrated here is learned from a few labelled examples over features the perceptual stage recovers; it is not open-vocabulary recognition of arbitrary categories, and it inherits the reach of classical perception — which recovers structure and specific known instances, but does not, by itself, put a general name to a novel natural ob-

ject. That last step is precisely the one the substrate is meant to learn, and extending the feature vocabulary and the range of learnable categories is the natural next stage of the work.

What the proof of concept establishes is the architecture: that a substrate with no perceptual model can nonetheless see, by pairing algorithmic perception with its own inductive learning, and that doing so buys a property the model-based route does not — the system knows when it does not know. The naming ability is learned, attributable, revisable and legible; its failures are abstentions rather than confident errors; and none of it depends on a pretrained backbone. Where a conventional pipeline would place a frozen network to turn structure into names, this substrate places a learner it already had.

9 Conclusion

We set out to give sight to a system that carries no neural network, by taking seriously the idea that vision is structure perceived and meaning learned. The demonstration is small but complete: real images are perceived into structure with deterministic algorithms; a few labelled examples teach the substrate a rule that names a category; new instances are named by applying that rule, and unlearned ones are declined. On a controlled evaluation the substrate named held-out instances at full recall with no false namings and no model calls, and an ablation showed the ability to be carried entirely by what it learned, its absence surfacing as honest abstention. A cognitive substrate can supply the semantic half of vision itself — and in doing so, it sees in a way you can question, correct, and trust.

References

[1] Dominion Labs. Structure Before Meaning: giving a cognitive substrate sight. Dominion Labs Research, 2025.

[2] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *NeurIPS*, 2012.

[3] A. Radford et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

[4] P. F. Felzenszwalb, D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[5] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, A. W. M. Smeulders. Selective search for object recognition. *International Journal of Computer Vision*, 104(2):154–171, 2013.

[6] J. Matas, O. Chum, M. Urban, T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[7] A. Kirillov et al. Segment Anything. In *ICCV*, 2023.

[8] Harnessing vision foundation models for high-performance, training-free open-vocabulary segmentation. *arXiv:2411.09219*, 2024.

[9] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2010.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[11] G. D. Plotkin. A note on inductive generalization. *Machine Intelligence*, 5:153–163, 1970.

[12] C. K. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41–46, 1970.

[13] R. Reiter. On closed world data bases. In H. Gallaire, J. Minker (eds.), *Logic and Data Bases*, Plenum Press, 1978.
