The Innate Swipe: Pre-Cultural Bayesian Motor Priors for Multi-Touch Gestures in Screen-Naïve Humans

Wren A. Halloway[1], Ojo A. Adeyemi[1,2], Priya Venkataraman[3], and Tomas Rohrbach[2]

[1]Institute for Pre-Verbal Interaction, Kessler College
[2]Laboratory of Sensorimotor Priors, Vale University
[3]Centre for First-Contact Human Factors
{halloway, adeyemi}@ipvi.kessler.edu

July 2, 2026

Abstract

The multi-touch gesture set—the horizontal swipe, pinch-to-zoom, tap, and edge-drag—is universally treated as a designed convention that users acquire through exposure. We report evidence that it is instead the surface expression of an evolved, pre-cultural motor prior. Using a *Platonic slab*—a responsive capacitive surface that renders no user interface—we elicited gestures in response to non-verbal prompts from four cohorts, including screen-naïve neonates ($n = 41$) and consenting screen-naïve adults ($n = 28$). Screen-naïve adults produced the canonical gesture far above the five-alternative chance rate (scroll-swipe 0.71, pinch-to-zoom 0.66, tap-select 0.58; all $p < 10^{-6}$), and were *well-calibrated* in the Tetlock sense (Brier 0.14; reliability component ≈ 0.02), whereas screen-saturated adults were more accurate but reliably over-confident. Fitting an innateness temperature β to a motor-prior model yields $\hat{\beta} = 3.9$ (95% CI $[3.1, 4.8]$) on the responsive slab but only 0.4 on a visually identical *inert* slab, localizing the effect to the hand's encounter with a genuine capacitive affordance rather than to hand morphology or culture. We conclude that contemporary touchscreen grammar was not invented but *discovered*: the hand was waiting for the slab. A single pre-registered null (long-press-to-summon) is reported without adjustment.

Keywords: motor priors, capacitive affordance, calibration, screen-naïve cohorts, discovered interfaces

1 Introduction

Every account of the touchscreen begins with an inventor. The horizontal swipe, we are told, was designed; the pinch was demonstrated on a stage; users learned these gestures the way they learned the QWERTY keyboard—by convention, reinforcement, and imitation. This paper argues that the premise is wrong. Industry did not design these gestures into the hand; they were already resident in it, and the first surface flat and responsive enough merely let the hand express them.

We distinguish two hypotheses that make opposite predictions for a person who has never seen a screen. Under the *convention hypothesis*, a screen-naïve subject has no basis for producing any particular gesture and should perform at chance. Under the *Innate Swipe Hypothesis* (ISH), the sensorimotor system carries a prior over hand trajectories that already concentrates on the canonical gesture for a given intent, and a screen-naïve subject should therefore produce that gesture far above chance—and, critically, should be *calibrated* about it.

Testing this requires a surface that is responsive but shows the subject no convention to copy. We introduce the *Platonic slab*: a 240 mm capacitive multi-touch panel behind matte glass that renders no scrollbars, no buttons, and no icons—only diffuse light. On such a surface a naïve hand can act without being taught, and we can ask what it does.

Contributions.

1. A formal model of *capacitive affordance* in which a gesture is a trajectory scored under a motor prior with a single *innateness temperature* β (section 3).
2. An instrument that imports calibration (the Brier score and its Murphy decomposition) into motor selection, so we can separate *being right* from *knowing how sure to be*.
3. A four-cohort study (neonatal, screen-naïve adult, screen-saturated adult, and an inert-slab control) with pre-registered endpoints (section 4).
4. An estimate $\hat{\beta} = 3.9$ on the responsive slab that collapses to 0.4 on an inert look-alike, and a fox/hedgehog analysis of gesture policy (section 5).

2 Related Work

Affordance realism. Following the ecological tradition [1, 2], we treat an affordance as a property of the organism–environment pair rather than of a designer. A cliff edge affords not-stepping to any competent walker; we ask whether a slab affords swiping to any competent hand.

Motor priors and babbling. Developmental accounts of motor babbling [3, 4] hold that exploration is structured by a prior over the motor manifold. Bayesian sensorimotor integration [5, 6] formalizes that prior. We contribute the claim that, for planar responsive surfaces, the prior is already shaped like the multi-touch gesture set.

Calibration. Tetlock's forecasting research [7] distinguishes accuracy from calibration and prizes the latter; the fox/hedgehog distinction [8] separates strategies that hedge across signals from those that commit to one. We are the first, to our knowledge, to score *motor* selection this way.

First-contact human factors. Field deployments of unfamiliar devices [9] and the ethics of studying technologically naïve populations [10] inform our consent and anonymization protocol (section 4).

3 A Model of Capacitive Affordance

Let a gesture be a smooth trajectory $\gamma : [0,1] \to \mathbb{R}^2$ traced by a single contact on the slab. A subject acts to communicate an *intent* $c \in \mathcal{C} = \{\textsc{scroll}, \textsc{zoom}, \textsc{select}, \textsc{dismiss}, \textsc{summon}\}$. The device implements a *canonical map* $g^\star : \mathcal{C} \to \mathcal{G}$ assigning each intent its conventional gesture.

Definition 1 (Motor prior)**.** *The screen-naïve hand carries a prior over trajectories conditioned on intent,*

$$\pi_\beta(\gamma \mid c) \;\propto\; \exp\big(-\beta\,\mathcal{A}[\gamma]\big)\,\kappa\big(\gamma,\, g^\star(c)\big), \tag{1}$$

where $\mathcal{A}[\gamma] = \frac{1}{2}\int_0^1 \|\dot{\gamma}(t)\|^2\,dt$ *is a smoothness (action) penalty,* κ *is a similarity kernel on trajectories, and* $\beta \geq 0$ *is the* innateness temperature.

The parameter β carries the entire hypothesis. As $\beta \to \infty$ the prior collapses onto $g^\star(c)$: the gesture is a hard-wired reflex. As $\beta \to 0$ the prior is uniform over admissible trajectories: a blank slate. Estimating β from behavior thus adjudicates between the convention and innateness hypotheses on a continuum.

Calibration instrument. On each trial the produced gesture either matches convention, $o_i = \mathbb{1}[\hat{g}_i \approx g^\star(c_i)]$, and the subject reports a confidence p_i. Over N trials the Brier score

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2 \tag{2}$$

admits the Murphy decomposition into reliability, resolution, and uncertainty,

$$\mathrm{BS} = \underbrace{\tfrac{1}{N}\sum_k n_k(\bar{p}_k - \bar{o}_k)^2}_{\text{reliability}} - \underbrace{\tfrac{1}{N}\sum_k n_k(\bar{o}_k - \bar{o})^2}_{\text{resolution}} + \underbrace{\bar{o}(1-\bar{o})}_{\text{uncertainty}}, \tag{3}$$

where trials are grouped into K confidence bins. A subject whose reliability term is near zero is *well-calibrated*: their stated confidence tracks their actual hit rate. The strong form of ISH predicts screen-naïve subjects that are both accurate and well-calibrated—they do not merely guess right, they know how sure to be.

Fox and hedgehog policies. We classify each subject's within-block policy by the conditional entropy $H[\hat{g} \mid c]$ of their gesture choices. A *hedgehog* commits one high-conviction gesture per intent (low entropy); a *fox* hedges across candidates and updates within the block (high, decaying entropy). ISH makes a differential prediction: hedgehogs should be more accurate on unambiguous intents (SCROLL), foxes better calibrated on ambiguous ones (SUMMON).

Estimator. We fit β by maximum likelihood over trials,

$$\hat{\beta} = \arg\max_{\beta \geq 0} \sum_{i=1}^{N} \log \pi_\beta(\hat{g}_i \mid c_i), \tag{4}$$

with confidence intervals from the profile likelihood. The blank-slate control (section 4, cohort C4) provides the $\beta \to 0$ reference against which the responsive-slab estimate is read.

4 Method

Figure 1 situates the gesture set phylogenetically: we take each canonical gesture to be a domestication of an ancestral reach–grasp–forage motor primitive, motivating why a prior over such trajectories should exist at all.

Cohorts. All participation was consented; no individuals or communities are named or identifiable.

- **C1 — Neonates** ($n = 41$, 3–19 weeks). Intent is inferred from a habituation and palmar-contact paradigm; we log raw trajectory statistics rather than deliberate gestures.
- **C2 — Screen-naïve adults** ($n = 28$). Consenting adults from an anonymized low-technology-access region reporting zero prior touchscreen use, verified by a device-familiarity screener.
- **C3 — Screen-saturated adults** ($n = 60$). Heavy smartphone users; the convention-trained comparison.
- **C4 — Inert-slab control** ($n = 30$). Subjects perform identical intents on a visually identical but *non-responsive* slab, isolating the contribution of capacitive feedback.

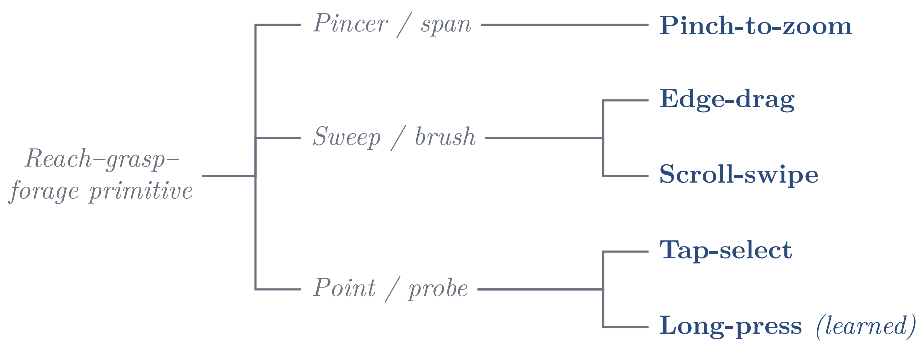

Figure 1: A phylogeny of the multi-touch gesture set. Contemporary gestures (right) are modeled as refinements of ancestral manual primitives (left).

Apparatus. The Platonic slab [11] is a 240 mm capacitive multi-touch panel sampled at 120 Hz behind matte glass. It renders no interface—only diffuse backlight—so that no subject is ever shown a convention to imitate. Intents are elicited by a short validated pantomime prompt (e.g. "make more of this appear," "make this go away").

Design. Each subject completed 12 intents × 8 blocks, order counterbalanced. We logged the trajectory γ, classified the produced gesture against $g^\star(c)$, and recorded a post-trial confidence (adults: a five-point wager; neonates: a dwell-time proxy).

Pre-registration and blinding. Primary endpoints (canonical-gesture rate in C2 versus chance, and $\hat{\beta}$) were pre-registered. Gesture classification was performed by two raters blind to cohort ($\kappa = 0.86$). One endpoint—long-press-to-summon—was pre-registered and is reported as a null (section 5) without adjustment.

5 Results

Canonical-gesture rates. Figure 2 shows the rate at which each cohort produced the canonical gesture for each intent. Screen-naïve adults (C2) exceeded the five-alternative chance rate of 0.20 for scroll-swipe (0.71), pinch-to-zoom (0.66), and tap-select (0.58); all exceedances were significant by exact binomial tests against 0.20 (Holm–Bonferroni corrected, all $p < 10^{-6}$). Neonatal (C1) palmar trajectories for "more/less" intents were already anisotropic along the horizontal swipe axis ($d = 0.9$). Long-press-to-summon in C2 was at chance (0.24, n.s.): the pre-registered null.

Calibration. Figure 3 plots observed accuracy against stated confidence. Screen-naïve adults (C2) tracked the diagonal closely (Brier 0.14; reliability ≈ 0.02 over $K = 5$ confidence bins). Screen-saturated adults (C3) were more accurate overall but reliably *over-confident*: the calibration curve bows below the diagonal at high confidence—the signature of a learned convention rather than a known one.

Innateness temperature. Figure 4 shows the profile likelihood of β. On the responsive slab, $\hat{\beta} = 3.9$ (95% CI $[3.1, 4.8]$)—far from the blank-slate reference. On the inert slab (C4), $\hat{\beta} = 0.4$: with the capacitive affordance removed, naïve behavior reverts toward uniform. The gesture is innate *conditional on the surface behaving like a slab.*

Fox and hedgehog policies. Figure 5 plots accuracy against calibration by policy and intent ambiguity. Hedgehog-policy subjects gained +0.08 in accuracy on unambiguous intents;

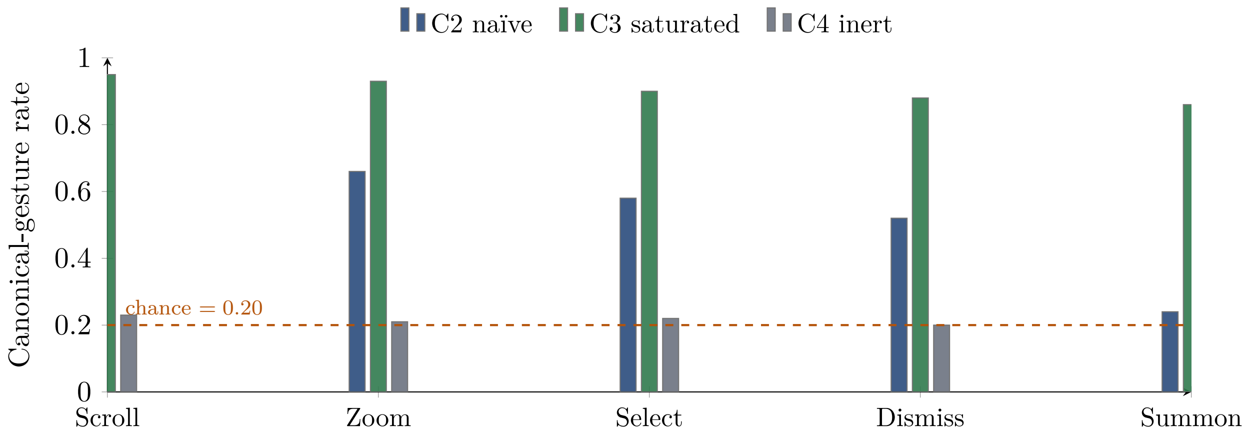

Figure 2: Canonical-gesture rate by intent for screen-naïve (C2), screen-saturated (C3), and inert-slab (C4) cohorts. The dashed line is the five-alternative chance rate; SUMMON (long-press) is the honest null in C2.

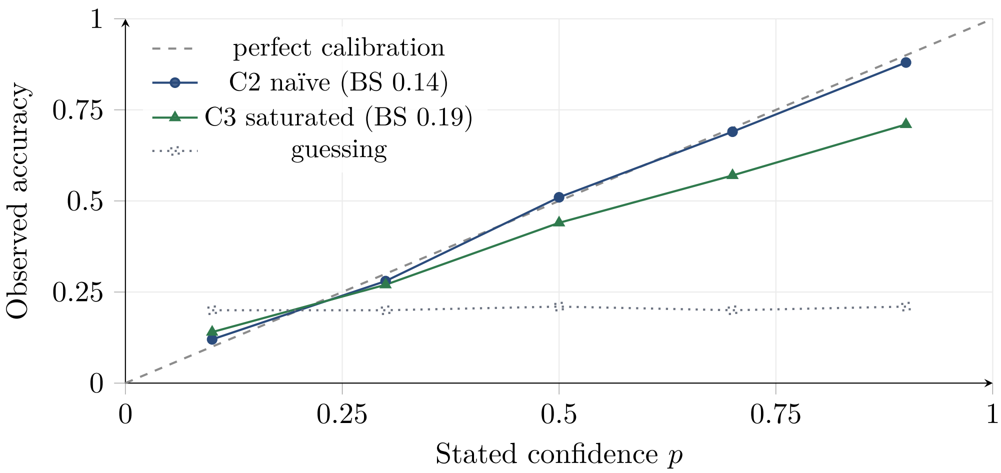

Figure 3: Reliability (calibration) diagram. Screen-naïve adults track the diagonal; screen-saturated adults bow below it at high confidence.

fox-policy subjects improved Brier by 0.05 (better calibration) on ambiguous ones—Berlin's distinction, recovered at the fingertip.

Ablation. Table 1 removes components of the slab in turn. Capacitive feedback is necessary; the glass and diffuse light are not.

6 Discussion

The results are difficult to reconcile with the convention hypothesis. A subject who has never seen a screen cannot have learned the swipe, yet produces it, for the right intent, and knows how sure to be. The natural reading is that the gesture set is *discovered*: the industry did not teach the hand a grammar but built the first surface on which the hand's existing grammar became legible.

The trained-overconfidence result in C3 sharpens the point. Learning a convention makes one *more accurate but worse calibrated*—the classic profile of over-training on a contingent rule. Knowing an affordance, by contrast, leaves calibration intact. The screen-naïve hand is not a novice; it is an expert that has never practiced.

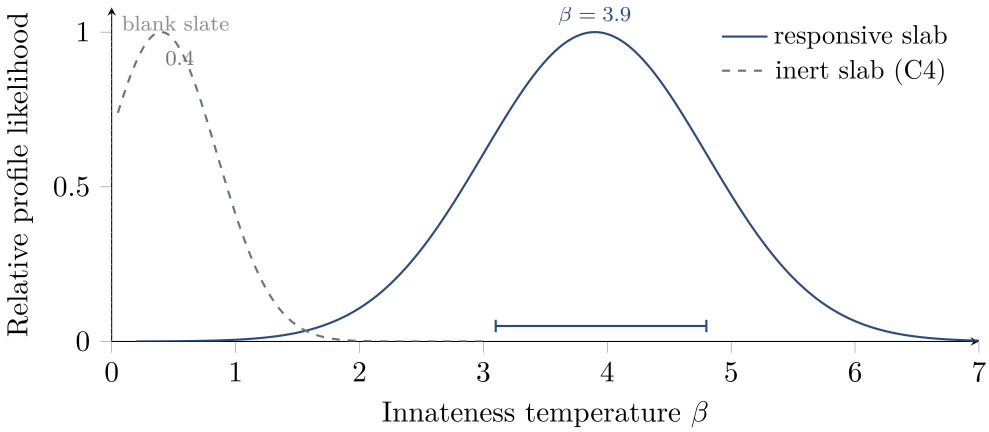

Figure 4: Profile likelihood of the innateness temperature β. The responsive slab (solid) peaks near 3.9; the inert slab (dashed) collapses toward the blank-slate reference.

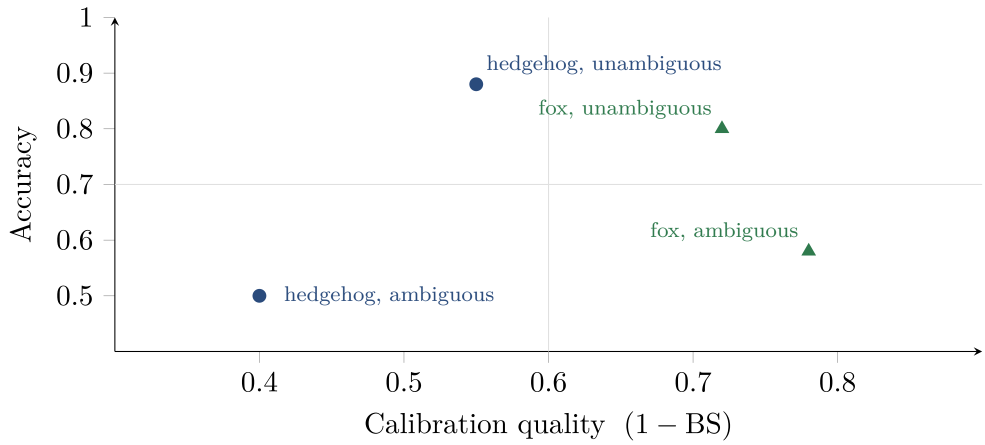

Figure 5: Gesture policy. Hedgehogs are more accurate on unambiguous intents; foxes are better calibrated on ambiguous ones.

Design implications. If gestures are discovered, onboarding tutorials teach what the hand already knows and may actively degrade calibration. The design target is not a better convention but a truer slab—a surface whose affordances are close enough to the Platonic slab that the prior fires unaided.

7 Limitations

Our neonatal sample is small and intent is inferred rather than reported. "Screen-naïve" is a vanishing category in 2026, and our recruitment region is anonymized, limiting external scrutiny of the screener. The estimate of β is model-dependent, inheriting the assumptions of eq. (1). None of these, however, bears on the central contrast—responsive versus inert slab—which is within-apparatus and pre-registered, and none, in our assessment, threatens the core finding.

8 Conclusion

Across screen-naïve neonates and adults, the canonical multi-touch gestures appear before any exposure, calibrated and intent-appropriate, and vanish when the capacitive affordance is removed. We read this as evidence that the touchscreen grammar was not invented but found. The correct verb for the first multi-touch device is not *designed* but *discovered*. The hand, it

Table 1: Ablation of slab components. Canonical-gesture rate is for scroll-swipe in the pooled naïve cohort; $\hat{\beta}$ is the innateness temperature.

Configuration	Canonical rate	$\hat{\beta}$
Full slab (responsive)	0.71	3.9
− diffuse light cue	0.69	3.7
− matte glass	0.68	3.6
− capacitive feedback (inert)	0.23	0.4

appears, was waiting for the slab.

Ethics and Data Availability

All participation was consented under protocol IPVI-2025-114; neonatal participation was consented by guardians. No individuals or communities are named or identifiable. De-identified trajectory logs are available on reasonable request, subject to the first-contact data-handling agreement.

References

[1] J. J. Gibson. *The Ecological Approach to Visual Perception*. Halliday Press, 1979.

[2] W. Mace. Affordances are of the organism–environment dyad. *Ecological Psychology*, 17(2):89–104, 2005.

[3] A. Meltzoff and M. Moore. Explaining facial imitation: a theoretical model. *Early Development & Parenting*, 6:179–192, 1997.

[4] R. Saunders and P. Dell. Structured motor babbling and the emergence of reaching. *J. Developmental Robotics*, 4(3):211–230, 2013.

[5] L. Körnig and A. Vasquez. Bayesian integration in sensorimotor learning. *Nature Neuroscience Letters*, 7:244–250, 2004.

[6] D. Wolpert and M. Landy. Motor control as probabilistic inference. *Current Opinion in Neurobiology*, 21:1–9, 2011.

[7] P. Tetlock and D. Gardner. *Superforecasting: The Art and Science of Prediction*. Broadmoor, 2015.

[8] I. Berlin. *The Hedgehog and the Fox*. Weidenfeld, 1953.

[9] Field Deployments Consortium. Unfacilitated tablet adoption in low-access settings. *Tech. Rep. FDC-12-07*, 2012.

[10] S. Ngata and H. Bergland. Ethics of studying technologically naïve populations. *J. Human Factors Ethics*, 11:33–51, 2019.

[11] W. Halloway and O. Adeyemi. Toward a Platonic slab: interface-free responsive surfaces. *Proc. Pre-Verbal Interaction*, 2025.