Measuring the Glance: An Adversarial Estimate of Habituated Perceptual Entropy in entviz and SSH Randomart*

Daniel Hardman

2026-06-22

Abstract

A figure meant for a person is only as strong as the part the person checks. The entviz design paper separates a *careful* reader, who reads the figure's lossless text and so checks the whole value, from a *habituated* one, who recognizes a few landmarks and trusts the rest — and names the bits a habituated glance carries as its central open problem. We estimate that quantity from the adversary's side. A reimplementation of the render model, certified field-for-field against the reference by a differential oracle, drives a grinder that counts the hashing work to forge each casually-checkable channel. The central result is a *curve* rather than a single figure. A habituated glance is parallel: it takes in a bundle of global channels at once, so entviz's involuntary one-glance read is the low-to-mid teens of bits, and it climbs with deliberate attention toward the lossless text ceiling — the whole value. The tolerances along it are discrimination thresholds, far tighter than the recognition the design disavows. Measured the same way, SSH randomart yields the first habituated figure for any randomart. The two are comparable at the bare glance, but entviz reads down to certainty where the bishop is capped near its field, and its structure and guided spot-check afford a habit the bishop cannot. Every perceptual tolerance here is modeled, not measured on people; the human study the design paper calls for is what would turn these models into measurements.

1. Introduction

A figure meant for a person is only as strong as the part the person checks. The entviz visualization turns a high-entropy value into a small picture so two values can be told apart at a glance, and its design paper [1] is careful to separate two readers. The *careful* one reads the figure's text, which for inputs of 512 bits or fewer is lossless — the value itself — and so is beyond an attacker's reach. The *habituated* one, having seen the right figure many times, no longer reads; she recognizes a handful of landmarks and trusts the rest. Habituation is not a failure of discipline but the steady state of any verification a person performs often, and it is the reader an attacker designs for. The security of the scheme rests on a quantity the design paper estimates but does not measure: how many bits a habituated glance carries. Its Table 3 put the figure at roughly twenty to forty bits when this work began, called the range deliberately wide, and made the measurement the central open problem; the design paper has since been revised in light of what follows.

We approach it from the side that cares about the answer. An attacker does not estimate perceptual entropy; he grinds against it. Every channel a glance can check is, in the render model, a

*daniel.hardman@gmail.com

discrete field — the background's two bits, the color bar's band order, the markers, the ellipse's silhouette — and each is a deterministic function of a single SHA-512 of the input. A candidate input is therefore an independent draw of all of them at once, and the work to forge a chosen set of landmarks is the number of candidates a machine must hash before one matches. That number is computable, and measurable, and it is what an attacker spends.

The first thing the measurement teaches is that there is no single number to report. A *habituated* reader does not check everything; but neither does she read one landmark at a time. entviz's global channels — background, the field of cell colors, the color bar, the ellipse, the blank-location pattern — are pre-attentive, taken in *together* at one involuntary glance, so the everyday read is the bundle of them: the low-to-mid teens of bits, not the two of any single channel. Deliberate attention then adds the *local* channels — the cell text especially, which is lossless at $\leq$512 bits — so the cost climbs from that glance toward the value itself. Summing the whole inventory as one figure measures a diligent reader, not a habituated one; quoting that sum as *the* habituated number was the mistake the exercise exists to catch. The honest object is a curve from the involuntary gestalt up to a lossless ceiling — and the tolerances along it are *discrimination* thresholds (detecting a difference), tight when the spot-check protocol puts the reference on screen and only modestly looser when it is remembered, not the much looser *recognition* (familiarity) the design explicitly disavows.

A measured cost is worth no more than the model it grinds against, so we do not trust our own reimplementation on faith. A parser-less reconstruction of the render model, given only what the reference parser produces, reproduces the entire golden conformance corpus field-for-field; the lean projection the grinder runs is pinned to that certified model, and the check earns its place by catching a real rounding discrepancy before it could bias a measurement. And because the perceptual half of the model — how loosely a glance reads each channel — is the half a grinder cannot certify, we ran two adversarial review passes over it, one in a vision-science lens, one in security usability, before locking any number. Those passes were conducted by AI models prompted to argue from a named lens, not by human domain experts; the author directed them and adjudicated every disagreement. They are a structured way to attack one's own assumptions, not independent validation — but they changed the work materially, and §10 records how.

The same machinery measures a competitor. SSH randomart — the drunken bishop [2] — has a single published perceptual-entropy figure, about twenty-two bits [3], and it is a whole-image estimate; no habituated number has ever been measured for it or for any randomart. (entviz is new, so there is no prior comparison between the two; the only reference point is randomart's own published figure, and reading it against a habituated number would be a regime error — careful against casual.) We measure a habituated number for both, the same way, and the comparison is honest in three parts. At the involuntary glance the two are comparable: each is a low-to-mid-teens-of-bits gestalt, the bishop's a single correlated blob, entviz's a modestly richer bundle of more independent channels. At the careful ceiling entviz is far more robust, and structurally so: its text is lossless, so it can be read down to the whole value, while the bishop is capped near its field. And along the habituation path entviz is more robust by design — its structure, its read-aloud text, and the guided spot-check protocol cultivate a habit the bishop's bare blob cannot. So entviz is the more robust artifact — not by winning the glance, which is a near-tie, but by holding every other part of the curve.

We are candid about what a grinder cannot settle. The perceptual tolerances and the salience weights are modeled — drawn from the just-noticeable-difference and visual-search literatures,

not measured on people — so the numbers should be read as "grind cost under this model of the glance," and the robust claims are the qualitative ones: that the operative cost is a curve, that it is low at a casual glance and climbs with attention, that entviz and randomart are comparable habituated. The human study that would turn the model's tolerances into measured thresholds is the same study the design paper calls for, and we end where it does, pointing at it.

2. What an attacker faces

The render model [4] is the attacker's feature vector. Each channel a glance can check — `bg`, the color bar, the markers, the ellipse, the blank positions and map, the quartile marks — is a discrete field, and every one is a function of `SHA-512` of the (normalized) input. Two consequences follow. First, a random candidate input is an independent draw of all channels at once, so the work to match a chosen *checked set* is the inverse of the per-candidate match probability, in bits. Second — the methodological rule, learned from an earlier misstep that compared rendered pixels — we operate only on the render-model fields, never on rasterized output, in any hot loop. Two SHA-512s and a few dozen integer operations per candidate; rendering happens only to build the final illustrative gallery.

A distinction that runs through everything: a channel is **un-steerable** if it is driven by the fingerprint (the hash) and **steerable** if it is driven by the input content the attacker chooses. The background, color bar, markers, ellipse, the fingerprint-edge cell colors, and the layout-driven CRC marks are all hash-driven; the attacker cannot *set* them, only grind, so the random-grinder cost is their *true* cost. The cell text and nucleus colors are input-driven; the attacker can set them by choosing the input, so for those the grinder reports an upper bound. The landmarks a habituated glance actually checks are, fortunately for the measurement, almost all in the first group.

3. A certified instrument

[The differential oracle.] A parser-less Rust port of the render model is fed, for each conformance vector, the output of the reference Python parser — the normalized core, its alphabet, whether a suffix or note adds a bottom strip, the raw byte length — and reproduces all forty-six golden vectors (forty-four short, two large) field-for-field. The lean `channels()` projection the grinder runs in its hot loop is pinned to that certified model by a consistency test. The check is not ceremonial: it caught a real port bug, a text-size rounding rule that used round-half-away where the reference uses round-half-to-even, on the one vector (`fs-6`) whose arithmetic lands on a tie. A measured attack cost is only as trustworthy as the model it grinds, and we certify the model before trusting a bit.

4. The difficulty curve, exact field

We first measure *exact-field* cost — a candidate matches a channel only when its value is bit-identical — across a spread of grid sizes, by Monte-Carlo over millions of candidates. The per-channel costs are stable and interpretable: `bg` is exactly two bits (a uniform choice among four palette colors); the color bar is two bits for the dominant stripe and about 4.6 for the full band order ($\log_2 4!$, on every grid); the markers cost about $2 \cdot \log_2 K$ for K slots; the ellipse, the load-bearing analog channel, costs about $12 + \log_2(\text{pool})$, fourteen to seventeen bits. The positional CRC channels are expensive exactly: the four quartile ticks pin four cells, $\approx 4 \cdot \log_2 N$, up to eighteen bits on the largest grid. Crucially, the empirical joint matches the independence-sum: the

channels are nearly independent, so their costs add. The full exact-field gestalt runs about thirty to forty-eight bits — a conservative ceiling, since exact-field demands distinctions a glance discards.

5. Perception, and what it discards

Exact-field is not how a glance reads. We replace exact equality with a model of the glance's tolerance, channel by channel. The ellipse is matched when its clipped silhouette overlaps the target's by enough area — intersection-over-union above a threshold τ; markers when each sits within one slot of the target's; the positional CRC marks when each sits within tolerance of its counterpart. Under this model the ellipse's cost roughly halves ($\tau = 0.9$) to thirds ($\tau = 0.8$), the markers halve, and the positional channels — expensive to match exactly because position is most of their bits — collapse toward the design paper's habituated estimates once a mark need only land in roughly the right place.

We bracket the position tolerance with two models: a **±1-cell** match (the two-dimensional analogue of the marker's ±1-slot tolerance) and a **quadrant** match (each mark localized only to which quarter of the grid it sits in). The ±1-cell model is the more defensible — translation-equivariant, with no boundary artifact — and lands the CRC channels near the paper's habituated figures; the quadrant model runs a few bits higher, and the gap, widest on odd grids where a 2×2 split yields a one-cell quadrant, is itself the boundary artifact we warned of. Reported under either, the full *diligent* gestalt — every channel, comparison JNDs — sits in the low-to-mid twenties; §6 separates that diligent sum from the *habituated* read, which is the parallel one-glance gestalt (a subset of these channels) and so lower.

The second review pressed hardest here, and four notes follow from it. First, area-overlap is the wrong ellipse metric: a viewer parses an ellipse by orientation, aspect, and size as *separable* quantities, and a single τ cannot be right for both a round and a thin ellipse — visible in the IoU cost swinging seven to twelve bits at fixed τ. The faithful model is a factored predicate (an orientation tolerance that loosens as the ellipse rounds, with aspect and size as Weber fractions and the anchor as a position JND), which lands the ellipse near seven bits under simultaneous discrimination and modestly lower from memory — below the IoU≥0.9 figure we first used. We report the factored estimate and flag its full implementation as the open refinement. (The ellipse is, to be clear, a *salient* feature: not the faint fill an earlier note called it, but a high-opacity stroked oriented contour, so its high salience rank stands.) Second, the modeled color-vision collapse was wrong: a Machado [5] simulation shows the four backgrounds remain mutually supra-threshold (so matching `bg` exactly is safe), but the genuinely confusable palette pair under deuteranopia/protanopia is **gold≡red**, not the red≡blue our placeholder luminance merge used — which models no real dichromat; a proper CVD simulation plus an acuity blur is the correct treatment. Third, treating the channels as perceptually independent (costs add) is an *upper bound*: the ellipse tint shifts the perceived color of the cells beneath it (simultaneous contrast), the quartile ticks crowd, and the markers can be masked by a busy bar — all of which make the joint *less* than the sum. Fourth, every tolerance here (τ, ±1 slot/cell, the from-memory widening) is a modeled just-noticeable difference [6], not one measured on people.

6. The operative number is a curve, and the glance is parallel

The habituated number is not a point; it is a curve, and three things shape it.

The first glance is parallel, not serial. A habituated reader does not check one landmark, then a second, then a third. entviz's *global* channels — the background, the broad field of cell colors, the color bar, the ellipse silhouette, the blank-location pattern — are pre-attentive [7]: a single involuntary glance takes them in *together*, the way the same glance takes in randomart's blob. So entviz's first "landmark" is not one channel but a bundle, and crediting it serially (background first, two bits, then the ellipse, ...) understates the involuntary read. The involuntary gestalt is the sum of those global channels — about seventeen bits as an independence-sum, lower once the ellipse's tint shifts the perceived cell colors beneath it and crowding bites, so call it the low-to-mid teens. Only the *local* channels — the cell text, the quartile ticks, the exact marker and blank positions — are serial: they require deliberate attention, and they are the cultivable headroom the next section turns to.

The tolerance is discrimination, not recognition. The security task is *same/different discrimination* — detecting that an imposter value differs from the legitimate one — not *familiarity recognition*, which the design paper explicitly rejects as a goal. This matters because recognition (a gist judgment, capped near a handful of categories per dimension [8]) is far looser than discrimination, and an earlier draft of ours wrongly imported the recognition framing and cratered the number. Discrimination splits by whether the reference is present: under the guided spot-check protocol it is, so the read is *simultaneous* discrimination at comparison JNDs (tight, the figures above); unguided, the reference is remembered, so it is *successive* discrimination — modestly looser than simultaneous, but nowhere near the recognition gap. Familiarity is the loose degenerate corner the design fights, not its target.

The operating point is artifact-dependent, so we do not match k. Comparing entviz-at-k against randomart-at-k begs the question: the distribution over what a habituated user attends is itself set by the artifact. Each visualization sits at its own operating point on its own curve, and where that point sits is the human-study unknown — from the artifact we can argue only the direction. So the honest object is the curve and the trajectory, not a single matched-k number.

For a representative 3×4 grid, then: entviz's involuntary one-glance gestalt is in the low-to-mid teens of bits (the bundle of its global channels); deliberate attention adds the serial channels, and because the cell text is lossless at ≤512 bits, a reader who attends it climbs without bound, toward the value itself. The everyday number is the involuntary gestalt — low-to-mid teens — and it rises with attention rather than being a fixed figure.

7. What the glance is assumed to check, and the two attackers

The curve assumes the landmarks a glance attends are the gestalt channels above. Two things it does *not* assume are worth stating. It assumes the reader reads no cell text: the forgeries share no text with the target, which matches the design paper's habituated model (text ≈ 0–3 bits) and is the attacker-favorable extreme. It is also defensible from the other side — the text *is* the input, so an attacker can largely *set* a cell's text rather than grind for it, which makes text-checking a weak casual defense rather than a strong one. And it assumes the reader does not tally the full field of cell colors; we model that field by the per-cell surround color anchored at the first-fixation cell, whose robust (un-steerable) contribution is a single ~2-bit anchor — which is exactly where the v10 redesign placed its un-grindable color.

The attacker, too, comes in two forms, and the reviews were right that we had foregrounded the wrong one. The **unconstrained** attacker grinds free random inputs; this is the vanity- or

birthday-style forgery, and it is the worst case for the *defender's* per-target difficulty. The **constrained** attacker must impersonate a *specific published identifier* — the usual real case — and his cost is set by the un-steerable channels alone, measured against the *easiest* target in the population rather than a typical one. Because the gestalt landmarks are all un-steerable, the two attackers face nearly the same gestalt cost here; the constrained attacker's advantage is target selection (he picks the victim whose figure is cheapest to forge), which shifts the operating point toward the floor.

8. The drunken bishop, measured habituated

SSH randomart renders a hash as a 17×9 field of visit counts: a bishop starts at center and steps diagonally per two bits of the hash, and the visit counts render through an ink ramp from `.` to `#`. We measure its habituated entropy with the same grinder, and the modeling choice that matters most — and that the reviews pressed hardest — is that **a glance does not read the characters**. The ramp encodes density as ink weight; adjacent levels are indistinguishable at speed; only `S` (fixed at center) and `E` (the endpoint) are legible letters, and only `E` carries information. So the read is of the *blob*, not the glyphs.

Following the vision-science review, the primary model is an ensemble read — salience-ordered landmarks: the silhouette of which regions are inked (the blob's shape and extent), the ink-weighted centroid, the densest region away from the structurally-always-dense center, and the endpoint's region — reported as the same cost(k) curve. (A finer density map is kept only as a diligent ceiling.) Density is banded against the scene *mean*, not its maximum, so a single dense cell cannot relabel the whole field; and a diagnostic confirms the start-at-center, edge-clamped walk leaves the center region densest far more often than chance on the classic 64-move walk, which is why the hot-region landmark excludes it.

The right comparison is not at a matched glance budget — §6 — but each artifact at its own involuntary gestalt and its own ceiling. The bishop's measured one-glance gestalt is about eleven bits (modern 32-byte walk) to fourteen (classic 16-byte); its channels (silhouette, centroid, density, endpoint) all derive from the one walk, so they are correlated, and the joint is that cumulative, not a sum. entviz's involuntary gestalt — the parallel bundle of its global channels — is in the low-to-mid teens, modestly richer and, importantly, *more independent*: color, shape, and layout are orthogonal, where the bishop's are one blob seen three ways.

Three honest readings follow, and they do not all point the same way.

At the involuntary glance, the two are comparable. Low-to-mid teens of bits each; we do not claim a large habituated-glance gap. The only published randomart figure, ~22 bits, is a whole-image, careful-regime number, not this one — so it cannot be read as evidence that randomart is the *stronger* artifact at a glance, and our measurement shows it is not; but neither is entviz dramatically stronger at the bare glance.

At the ceiling, entviz is far more robust — and this is structural, not hypothesized. entviz's cell text is lossless at ≤512 bits, so a reader who attends it verifies the whole value; entviz degrades gracefully from a glance to a full reading, its careful ceiling the input itself. The bishop has no text and no serial headroom: the density field is its entire content, and its ceiling is that ~22-bit field whether glanced or studied. One artifact can be read all the way down to certainty; the other cannot. That is the largest genuine difference between them.

On the trajectory between glance and ceiling, entviz is more robust by design — grounded in the artifact, pending the human study. Habit is artifact-dependent. The bishop affords only the blob, so a habituated bishop-reader's attention has nowhere to climb. entviz affords rows and columns, chunked readable text, reading aloud (a verbal channel the bishop lacks), and — above all — the guided spot-check protocol, which raises both how many channels are attended and how tightly they are discriminated (toward the simultaneous, reference-present regime). That protocol is precisely the countermeasure to the habit degenerating to the cheapest subset. Whether real users actually climb is the human-subjects question we cannot settle here; the *affordance* asymmetry — entviz has the rungs, the bishop has none — is not in doubt. So the honest verdict is not "the choice barely matters." At the bare glance it nearly doesn't; but entviz is the more robust artifact, decisively at the careful ceiling and, by design, along the habituation path that a deployment would actually cultivate.

9. The defense, instantiated

The design paper's seeded comparison walk reduces to two combinatorial bounds, and we instantiate both with the measured costs. The landmark walk — survival C(J,L)/C(K,L) when the attacker matched J of K landmarks and the walk checks L — reproduces the paper's worked example (one in roughly twenty-five hundred), and because the channels are independent the clean J-of-K partition is the right model. The cell-reading walk crushes a gestalt forgery on the first cell read, because a free-core forgery shares no cell with the target; the only residual hard case is the steered near-collision that matched most cells, which is the one attack worth a dedicated future grinder. We are careful, as the HCI review insisted, not to let the clean combinatorics imply the *human* defense is established: the walk's security is conditional on the user following the seeded order, which is exactly the diligence habituation erodes, and that compliance is unmeasured.

10. What the review passes changed, and the limits that remain

We put the modeling through two adversarial review passes — vision-science and security-usability — before locking numbers, and they changed the work. The headline became a salience-ordered curve rather than a summed point (both passes' blocking finding). The bishop's primary read became an ensemble silhouette-and-centroid model rather than a nine-level density map, which the vision-science review judged a near-ceiling rather than a glance; its density banding moved from scene-max to scene-mean to kill a single-cell relabeling artifact, and the center-dense artifact of the walk was diagnosed and handled. The constrained, fixed-identifier attacker was promoted from a caveat to a co-equal case. And several claims were softened: what we report is grind cost under a modeled checked-set, not a measured perceptual entropy; agreement with the design paper's estimate is consistency between two models by the same author, not independent validation.

A second, focused round reviewed the entviz tolerances specifically, and a third correction — the author's, overriding part of it — is worth recording honestly because the numbers moved both ways. That round proposed that habituation is *recognition against memory*, two to three times looser than comparison, and on that basis we briefly dropped the casual figure to about four to six bits. That was wrong: the security task is *discrimination* (detecting a difference), which the design targets and which is far tighter than recognition (familiarity), the regime the design explicitly disavows. So the recognition widening was reverted, and the casual number came back up to the involuntary-gestalt range — the low-to-mid teens. What survived from the second review,

and did lower the figure somewhat, was real: the IoU ellipse metric is too tight and should be a factored orientation/aspect/size predicate (≈7 bits, not ≈11); the peripheral CRC and quartile marks are local, not part of the parallel one-glance gestalt; the additive channel model is an upper bound because the overlay tint and crowding couple the channels; and the modeled color-vision collapse named the wrong pair — gold≡red under deuteranopia/protanopia, not red≡blue — and should come from a real CVD simulation. One factual error the review caught: the ellipse is a high-opacity stroked contour, not the faint fill an earlier note called it, so its high salience rank stands.

The limits that remain are honest ones. The tolerances and salience weights are modeled from the literature, not measured on people; the factored ellipse predicate and a real CVD simulation are implemented only as estimates here; the discrimination-from-memory widening for the unguided ceremony is asserted, not measured; the additive gestalt is an upper bound on the joint; and salience is target-dependent in a way a single synthetic target cannot capture. Above all, the claim that entviz's design *cultivates* a better-than-glance habit — the rungs the bishop lacks — is grounded in the artifact's affordances, not in observed behavior. Each is a place where a human-subjects study — the one the design paper names as its central open problem — would turn a model into a measurement.

11. Conclusion

The habituated number, estimated for years, now has a *shape*: not a single figure but a curve, rising from an involuntary one-glance gestalt of the low-to-mid teens of bits — the bundle of global channels a glance takes in at once — through deliberate attention to a lossless ceiling, the whole value, at or below 512 bits. Measured the same way — the first habituated number for any randomart, against a visualization too new to have been compared to anything — SSH randomart is comparable to entviz at the bare glance; but the bishop is capped near its field while entviz reads down to certainty, and only entviz brings the structure, the spoken text, and the guided spot-check that can pull a habituated reader up off the glance. So the honest comparison is not a tie and not a rout — a near-tie at the bare glance, and a clear entviz advantage everywhere else the curve goes. None of this is a human measurement, and we have been careful to say so; it is the most an adversary's arithmetic can establish before the study that would close the loop. What the arithmetic does establish is that the quantity the design paper called its central unknown is tractable, that it is a curve rather than a number, and that the casual end of that curve — where security is thinnest and habituation lives — is exactly where a verification protocol, not a prettier picture, has to do the work.

References

[1] Hardman, D. 2026. *Amplifying Difference: Perceptual Design and Verification of Human-Centric Entropy Visualizations.* Codecraft Papers. https://dhh1128.github.io/papers/amp-diff.html

[2] Loss, D., Limmer, T. and von Gernler, A. 2009. *The Drunken Bishop: An Analysis of the OpenSSH Fingerprint Visualization Algorithm.* Technical report (September 20, 2009). http://dirk-loss.de/sshvis/drunken_bishop.pdf

[3] Hsiao, H.-C., Lin, Y.-H., Studer, A., Studer, C., Wang, K.-H., Kikuchi, H., Perrig, A., Sun, H.-M. and Yang, B.-Y. 2009. A Study of User-Friendly Hash Comparison Schemes. In *Proceedings*

of the 2009 Annual Computer Security Applications Conference (*ACSAC '09*). IEEE, 105–114. https://doi.org/10.1109/ACSAC.2009.20

[4] Hardman, D. 2026. *entviz — Algorithm Specification.* Version 15. https://dhh1128.github.io/entviz/spec

[5] Machado, G. M., Oliveira, M. M. and Fernandes, L. A. F. 2009. A Physiologically-Based Model for Simulation of Color Vision Deficiency. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 1291–1298. https://doi.org/10.1109/TVCG.2009.113

[6] Gescheider, G. A. 1997. *Psychophysics: The Fundamentals* (3rd ed.). Lawrence Erlbaum Associates, Mahwah, NJ. ISBN 978-0805822816.

[7] Treisman, A. M. and Gelade, G. 1980. A Feature-Integration Theory of Attention. *Cognitive Psychology* 12, 1 (1980), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

[8] Miller, G. A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63, 2 (1956), 81–97. https://doi.org/10.1037/h0043158