# I · V · O

## Preliminary Observer Reliability Study

**v0.1**

*"Can observers consistently read emergent state patterns using IVO notation?"*

| ~2 weeks | 8–15 observers | 25 stimulus clips | 4 observer groups |
|---|---|---|---|

Design by Authenticity

IVO Validation Research Program · Working Paper v0.3 · 2026

*This is a working paper describing a pre-registered study design. No data has been collected yet.*

## Abstract

IVO is a three-layer notation system for describing emergent field states using compact symbolic sequences. Before IVO can be used reliably in practice — across team interaction, infrastructure analysis, simulation research, and recovery environments — it must be shown that independent observers assign the same symbols to the same visual states. This study tests that assumption for the first time.

We present a 25-item stimulus set of short field recordings (5 seconds each) drawn from the IVO Asymmetry Ring — a parametric field simulation. Observers from four groups (experiential observers, mental health professionals, technical researchers, and uninstructed laypeople) independently assign I, V, and O symbols to each stimulus. We measure inter-rater agreement using Fleiss' kappa (κ) per layer, construct confusion matrices for each symbol, and compare agreement rates across observer groups and lens modes.

**Expected outcome:** symbol extremes (e.g. ! vs ·, ∞ vs ×) will show substantial agreement (κ ≥ 0.60) without training. Boundary pairs — · vs •, ~ vs ↺, O vs () — are expected to show lower agreement (κ < 0.60) and constitute the primary revision targets.

This pilot study is designed to be reproducible, small enough to complete in two weeks, and sufficient to generate the first empirical claim about IVO as an observation instrument.

**Keywords:** *inter-rater reliability · notation systems · field observation · emergent dynamics · IVO · visual state recognition · Fleiss' kappa*

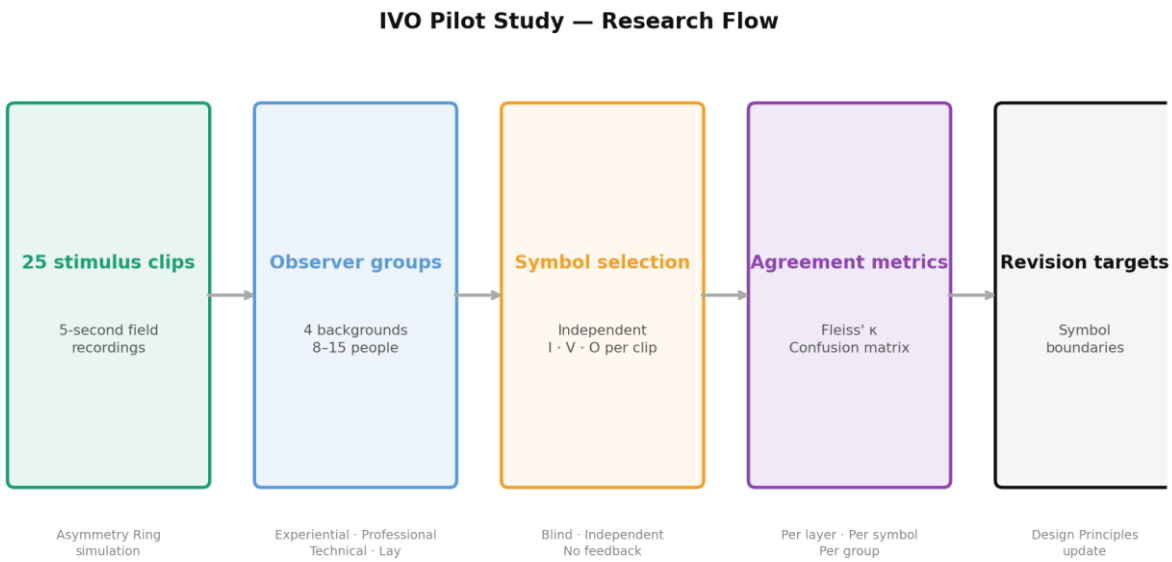


*Figure 1. Research flow — from stimulus clips to revision targets.*

# 1. Background and Motivation

## 1.1 What IVO is

IVO is a compact notation for describing the state of a dynamic system at a given moment. A state is expressed as a three-symbol sequence:

- `I` — the observation layer: intensity, density, and quality of individual presence
- `V` — the vector layer: direction, speed, and pattern of movement
- `O` — the field layer: structure, boundedness, and quality of context

A complete notation such as `I > O` means: stable observation · directed movement · enclosed field. The same notation applies across radically different domains — therapeutic sessions, team dynamics, infrastructure flows, or parametric field simulations — through lens-specific word translations that leave the symbolic layer unchanged.

## 1.2 Why reliability must be measured first

IVO has been developed as a working instrument across multiple domains: team and group dynamics, infrastructure and logistics, technical system monitoring, and human recovery environments. The notation system, the symbol definitions, and the Configuration Matrix have been refined over multiple versions. What has not yet been tested is the most basic assumption underlying any observation language: that independent observers, looking at the same event, arrive at the same description.

> **Working assumption** This study tests whether IVO symbols can be read consistently across training backgrounds and domains — and precisely where they cannot.

Without this test, IVO remains a designed system with unknown inter-rater reliability. With it — even with a small pilot — IVO becomes a measured instrument. The difference is not cosmetic. It is the difference between "I think this works" and "we measured something."

## 1.3 Position in the validation roadmap

The IVO Validation Roadmap (Design by Authenticity, 2026) specifies three validation streams in order of priority:

```
III → II → I
```

Notation reliability (III) is addressed first because a notation that cannot be read consistently should not be the target of actuator calibration (I) or field behaviour prediction (II). This study is the first step in stream III.

## 2. Research Questions

**RQ1** *(Primary)* Can independent observers assign consistent IVO symbols (I, V, O) to short field recordings without domain training?

**RQ2** *(Boundary)* Which specific symbol pairs produce the lowest agreement, and are these the predicted boundary zones?

**RQ3** *(Group)* Does observer background (lived experience / professional / technical / lay) systematically affect agreement rates?

**RQ4** *(Lens)* Does lens mode (Simple / Reflective / Technical framing) affect which symbols observers select?

**Scope note** This pilot is not designed to validate IVO comprehensively. It is designed to produce the first reproducible measurement of symbol-level readability. Results will directly inform which symbols require redefinition, which are ready for operational use, and which boundary pairs need additional specification in the Design Principles sheet.

# 3. Participants

## 3.1 Sample

Total target sample: 11–15 observers. Minimum for meaningful Fleiss' kappa calculation at this item count: 8 observers. Recruitment stops at 15 to keep the study scope manageable for a pilot.

| Observer group | Inclusion criteria | n target | Lens framing |
|---|---|---|---|
| **Experiential observers** | *Personal or professional experience with recovery, group process, or complex social environments. Recruited via community or professional networks.* | **3–4** | Receives Simple lens framing only. |
| **Professional observers** | *Practitioner background in psychology, coaching, systems facilitation, social work, or organisational practice. At least 2 years of applied experience.* | **3–4** | Receives Reflective lens framing. |
| **Technical researchers** | *Background in systems, data, UX research, or AI. Able to interpret parametric output.* | **3–4** | Receives Technical lens framing. |
| **Uninstructed laypeople** | *No relevant professional background. No prior exposure to IVO.* | **2–3** | Receives only the symbol reference card. No lens framing. |

## 3.2 Recruitment

Participants are recruited via the existing networks of Design by Authenticity: community and practitioner contacts, design research networks, and university connections. Participation is voluntary and unpaid. Estimated time commitment: 25–35 minutes.

## 3.3 Blinding

Observers in different groups receive different framing text (lens mode). They are not informed of the other lens conditions or of the specific hypotheses about boundary pairs. The stimulus set is identical across all groups. Observers are blind to each other's responses throughout.

## 3.4 Exclusion criteria

- Prior participation in IVO development or design sessions.
- Inability to view short video clips on the study device.
- Incomplete responses (< 20 of 25 items completed).

# 4. Materials

## 4.1 Stimulus set — 25 clips

Each stimulus is a 5-second screen recording of the IVO Asymmetry Ring field simulation, captured after actuator convergence (≥ 300 frames post-activation). Clips are presented without sound, without HUD overlay, and without notation labels. Observers see only the moving field.

The 25 stimuli are structured as follows:

- 5 anchor items — unambiguous extremes for baseline calibration
- 6 boundary items — 3 pairs testing the predicted low-agreement zones
- 8 group items — configurations from Crisis, Pressure, and Searching groups
- 2 divergent items — configurations outside the Matrix (observer must construct)
- 4 repeated items — retests of items 1, 3, 8, 19 (presented as items 12, 25, 21, 19) for intra-rater consistency

| # | Target I | Target V | Target O | Config label | Group | Rationale |
|---|---|---|---|---|---|---|
| 1 | ! | >> | ) ( | ! >> ) ( | Anchor | *Crisis extreme — high activation, momentum, tight field* |
| 2 | · | > | ∞ | · > ∞ | Anchor | *Minimal extreme — barely present, open field* |
| 3 | I | > | O | I > O | Anchor | *Baseline stable — most common target state* |
| 4 | ! | ↓ | ∞ | ! ↓ ∞ | Anchor | *Recovery — alarm decreasing into open field* |
| 5 | # | – | [] | # × [] | Anchor | *Rare saturated state — maximum density + constraint* |
| 6 | · | > | O | · > O | Boundary | *· minimal — test vs. • present* |
| 7 | • | > | O | • > O | Boundary | *• present — adjacent to · minimal* |
| 8 | I | ~ | ∞ | I ~ ∞ | Boundary | *~ fluctuation in open field* |
| 9 | – | ↺ | ) ( | ↺ ) ( | Boundary | *↺ persistent loop — temporal pattern* |

| # | Target I | Target V | Target O | Config label | Group | Rationale |
|---|---|---|---|---|---|---|
| 10 | I | > | () | I > () | Boundary | *() protected — bounded field* |
| 11 | I | > | O | I > O | Boundary | *O enclosed — natural coherence (vs () above)* |
| 12 | ! | >> | ) ( | ! >> ) ( | Crisis | *Repeat of item 1 — delayed retest for consistency* |
| 13 | : | >> | ) ( | : >> ) ( | Crisis | *Fragmented + momentum in tense field* |

| # | Target I | Target V | Target O | Config label | Group | Rationale |
|---|---|---|---|---|---|---|
| 14 | ! | ↓ | ) ( | ! ↓ ) ( | Crisis | *Alarm decreasing but field stays tense* |
| 15 | · | ↓ | ) ( | · ↓ ) ( | Crisis | *Low activation + decreasing + tense* |
| 16 | * | >> | ) ( | * >> ) ( | Pressure | *Active observation with momentum in tense field* |
| 17 | I | — | ) ( | I ) ( | Pressure | *Stable observation + tense field, V unsteered* |
| 18 | ? | ~ | ) ( | ? ~ ) ( | Pressure | *Searching + fluctuation + tense field* |
| 19 | — | ↺ | ) ( | ↺ ) ( | Pressure | *Persistent loop (same as item 9 — different context)* |
| 20 | ? | ⟳ | O | ? ⟳ O | Searching | *Reorientation event — temporal, brief* |
| 21 | I | ~ | ∞ | I ~ ∞ | Searching | *Fluctuation in open field (same as item 8 — retest)* |
| 22 | — | ⟳ | O | ⟳ O | Searching | *Reorientation without I-target — ambiguous state* |

| # | Target I | Target V | Target O | Config label | Group | Rationale |
|---|---|---|---|---|---|---|
| **23** | ? | ⟳ | ∞ | ? ⟳ ∞ | **Divergent** | *Not in Matrix — observer must construct notation* |
| **24** | ! | ~ | ∞ | ! ~ ∞ | **Divergent** | *High activation + fluctuation + open — unusual combination* |
| **25** | I | > | [] | I > [] | **Anchor** | *Final anchor — stable, directed, within structure* |

## 4.2 Response form — per stimulus

For each clip, observers answer three independent questions:

**I symbol** Which I-symbol best describes what you see? `[Dropdown: · • : I ! * # ?]`

**V symbol** Which V-symbol best describes the movement? `[Dropdown: . > >> ↑ ↓ ↺ ⟳ ~]`

**O symbol** Which O-symbol best describes the field? `[Dropdown: O () )( [] ~ ∞ × :]`

A free-text field ("What made you choose this?") is optional per item. A confidence slider (1–5) is included and recorded but not used in primary analysis.

## 4.3 Reference card

All observers receive a one-page symbol reference card before beginning. The card shows each symbol, its label, and a one-line description. No field examples are shown on the card. The card is available during the task.

Observers in the lay group receive only the card. Observers in the other three groups additionally receive their lens-specific framing text (2–3 sentences contextualising the task to their domain).

# 5. Procedure

## 5.1 Session structure

The study is conducted online, asynchronously, via a web-based form. Each session takes 25–35 minutes. Observers complete the study in a single sitting; the form does not save partial progress.

| Step | Time | Phase | Description |
|---|---|---|---|
| **1** | 00:00 | **Welcome + consent** | *Participant reads information sheet and gives digital consent. Domain affiliation is recorded.* |
| **2** | 00:03 | **Reference card study** | *Observer studies the IVO symbol reference card. No time limit; they proceed when ready.* |
| **3** | ~05:00 | **Practice block (3 clips)** | *Three non-scored practice clips to familiarise participants with the response format. No feedback is given on symbol choices during the practice block.* |
| **4** | ~10:00 | **Main task (25 clips)** | *Clips are presented in fixed randomised order (same for all observers). Per clip: watch (5s, plays once auto-repeat), answer I / V / O dropdowns, optional free-text, confidence slider.* |
| **5** | ~32:00 | **Exit questions** | *5 questions: which symbols felt clearest, which felt ambiguous, general comments. Observer group affiliation confirmed.* |
| **6** | ~35:00 | **Debrief** | *Brief explanation of the study purpose. Link to results summary offered (to be shared post-analysis).* |

## 5.2 Clip presentation

Each clip plays automatically. It does not loop during the task; observers see it once. A "replay" button is available for a maximum of one additional viewing. This mirrors realistic field observation conditions where events are not repeatable at will.

## 5.3 Randomisation

Clip order is randomised once and held constant across all observers (fixed seed randomisation). This controls for order effects while keeping the session structure comparable across participants. The order of the I / V / O questions within each item is fixed (always I → V → O) to prevent strategy differences.

## 5.4 Data recording

- Response per item: I symbol (categorical), V symbol (categorical), O symbol (categorical).
- Response time per item (ms from clip end to final answer).
- Confidence rating per item (1–5 Likert).
- Free-text comments (optional).
- Exit questionnaire responses.
- Observer group, lens mode, session timestamp.

# 6. Analysis Plan

> **Pre-registered** This analysis plan is specified prior to data collection. No analysis decisions will be made after viewing the data.

## 6.1 Primary measure — Fleiss' kappa per layer

*Fleiss' kappa (κ)* is calculated independently for the I-layer, V-layer, and O-layer across all observers and all items. This produces three primary κ values. A symbol-level κ is also calculated per individual symbol (e.g. κ for ! vs all other I-symbols) to identify which symbols drive disagreement.

| κ range | Interpretation | Action for IVO |
|---|---|---|
| κ < 0.40 | Poor / fair | *Symbol needs redefinition or removal from active set* |
| 0.40–0.59 | Moderate | *Boundary description required; add to Design Principles* |
| 0.60–0.79 | Substantial | *Acceptable for operational use in current version* |
| κ ≥ 0.80 | Almost perfect | *Self-evident; use as calibration anchor in training* |

## 6.2 Confusion matrices

For each layer, a symbol × symbol confusion matrix is computed showing how often each intended symbol was read as each other symbol. Confusion matrices are computed separately for each observer group. The diagonal represents correct identification; off-diagonal cells reveal systematic substitution patterns.

Particular attention to predicted confusions:

- I-layer: · ↔ •, : ↔ ?
- V-layer: ~ ↔ ↺, ↻ ↔ >
- O-layer: O ↔ (), ~ ↔ :

## 6.3 Group comparison

Mean κ per layer is compared across observer groups using a simple descriptive summary. Given the small sample, no inferential statistics are used for group comparisons. Patterns are reported as directional tendencies only. A larger replication study (n > 40) would be required for inferential group comparison. The four-group structure — Experiential, Professional, Technical, Lay — is intentionally broad, covering a range from practitioner to uninstructed observer.

## 6.4 Intra-rater consistency

Items 1, 3, 8, and 19 appear twice in the stimulus set (items 12, 25, 21, 19 respectively). Agreement between an observer's own responses on repeated items is calculated as a consistency rate. Observers with < 50% self-agreement on repeated items are flagged for quality review.

### 6.5 Free-text analysis

Free-text responses are reviewed qualitatively after quantitative analysis. Recurring themes are noted and used to generate hypotheses for symbol boundary descriptions. This is exploratory only and does not contribute to primary κ calculations.

### 6.6 Reporting threshold

Results are reported regardless of outcome. A finding of low agreement ($\kappa < 0.40$) is considered as informative as high agreement. The pilot is not designed to confirm the notation; it is designed to locate where the notation needs work.

# 7. Timeline

| Period | Phase | Activities |
|---|---|---|
| **Week 1, days 1–2** | **Preparation** | *Finalise stimulus set. Record 25 clips in Asymmetry Ring. Build response form. Prepare reference card (all lens versions).* |
| **Week 1, days 3–4** | **Piloting** | *Researcher completes full session to check timing and form logic. One colleague outside the project completes a test run.* |
| **Week 1, days 5–7** | **Recruitment** | *Invite participants via network. Target: all 11–15 observers confirmed and scheduled by end of week 1.* |
| **Week 2, days 1–5** | **Data collection** | *All observers complete sessions. Send reminders on day 3 for outstanding completions.* |
| **Week 2, days 6–7** | **Analysis** | *Calculate Fleiss' kappa per layer. Build confusion matrices. Write results summary. Draft v0.1 report.* |
| **Post-study, week 3** | **Dissemination** | *Share results summary with participants. Circulate working paper externally. Update IVO Design Principles sheet based on findings.* |

## 7.1 Resource requirements

- Researcher time: ~12–16 hours total (clip recording, form build, analysis, writing).
- Screen recording tool: any software that records the Asymmetry Ring at 1080p, 60fps.
- Form platform: Google Forms, Typeform, or equivalent. Must support video embed and dropdown response.
- Analysis: R or Python (numpy, pandas, statsmodels) for Fleiss' kappa and confusion matrix.
- No budget required. This pilot does not involve clinical intervention, deception, or sensitive personal data. Ethics review requirements depend on the institutional context of the replicating researcher.

> **Scalability note** This design is intentionally minimal. If κ values are promising (≥ 0.60 on two layers), the design scales directly to a larger replication study with n = 40–60 observers and formal statistical group comparison.

# 8. Expected Outcomes and Implications

## 8.1 If agreement is high (κ ≥ 0.60 on all three layers)

IVO notation is readable as a stand-alone observation instrument. The symbol reference card is sufficient for reliable use without domain training. Implication: IVO can be released for broader field testing without a mandatory training protocol.

## 8.2 If agreement is mixed (some layers high, some low)

This is the expected outcome. Likely pattern: O-layer (field quality) shows higher agreement than V-layer (movement pattern), because field shape is more visually salient than directional nuance. Implication: V-symbols require boundary descriptions in the Simple lens. The boundary pairs are revised in the Design Principles sheet.

## 8.3 If agreement is low across layers (κ < 0.40)

The symbol set requires structural revision before field deployment. This finding would not invalidate IVO — it would locate the revision target precisely. Implication: the Configuration Matrix is paused at current version; symbol redesign process begins.

## 8.4 Value of this study independent of outcome

Regardless of κ values, this study produces:

- The first empirical measurement of IVO observer reliability.
- A confusion matrix showing exactly which symbols are misread and as what.
- Qualitative data from four observer groups about which elements of the notation feel natural.
- A reproducible protocol that can be run again after revisions to measure improvement.
- A publishable working paper establishing IVO as a measured instrument, not just a designed one.

> **For prospective student researchers** This study is designed to be completed within a 2-week bachelor thesis sprint or as a defined sub-study within a larger research programme. The scope is deliberately bounded, the method is standard, and the results are directly applicable. Contact Design by Authenticity for collaboration inquiry.

# 9. Limitations

## 9.1 Sample size

A sample of 8–15 observers is sufficient for an exploratory pilot and for computing Fleiss' kappa at a group level. It is not sufficient for inferential group comparisons (e.g. testing whether GGZ professionals outperform laypeople on specific symbols). Findings from group comparisons should be treated as directional hypotheses for replication, not as confirmed effects.

## 9.2 Stimulus validity

Stimuli are generated from a parametric field simulation (IVO Asymmetry Ring). They are valid representations of IVO field states by construction, but they are not naturalistic recordings of real social or clinical events. It is possible that observers who read the simulation reliably cannot read naturalistic events reliably, or vice versa. A follow-up study with naturalistic stimuli (annotated video clips from team sessions or therapeutic conversations) is planned as a second validation phase.

## 9.3 Single-session design

The study is completed in a single session. Learning effects within the session (observers improving as they practice) are not controlled. The fixed clip order mitigates but does not eliminate this. A counterbalanced order in a replication study would address this.

## 9.4 Intended notation as provisional ground truth

The "correct" notation for each stimulus is defined by the researcher's intended configuration — the parameter values set in the Asymmetry Ring before recording. For this pilot, intended configuration functions as provisional ground truth. This is a principled starting point, but it carries a circularity risk: the researcher who designed the stimuli also defines the correct answer.

To mitigate this, a minimum viable annotation step is built into the pre-study protocol (see §9.4.1 below). A full resolution requires a larger follow-up study with a naturalistic stimulus set annotated by independent domain experts.

### 9.4.1 Independent annotation protocol (pre-study)

Before data collection begins, the 25 stimuli are independently annotated by three annotators who were not involved in designing the stimuli or the notation system:

- One annotator from a technical background (systems, data, or UX research).
- One annotator from a practitioner background (coaching, facilitation, or social work).
- One annotator who is unfamiliar with IVO — working only from the symbol reference card.

Annotators work independently, blind to each other and to the intended configuration. For each stimulus, the consensus label is defined as the notation on which at least 2 of 3 annotators agree per layer (I, V, O independently). Where no 2-of-3 consensus is reached on a layer, that layer is marked as contested for that item. Contested items are flagged in the analysis and excluded from the primary κ calculation; they are reported separately as items with unclear ground truth.

> **Effect on credibility** This protocol replaces single-researcher ground truth with a minimal consensus procedure. It does not require expert IVO knowledge — only that annotators apply the reference card independently. Two hours of annotator time before the study eliminates the circularity objection.

# References

Design by Authenticity (2026). IVO Matrix v3.5. Unpublished working document.

Design by Authenticity (2026). IVO Asymmetry Ring Feedback Laboratory — User Guide and Technical Reference. Unpublished working document.

Design by Authenticity (2026). IVO Validation Report — Working Paper v0.3. Unpublished working document.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378–382.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174.

# Appendix A — Symbol Reference Card (observer version)

## I — Observation layer

| Symbol | Label | Simple description |
|---|---|---|
| · | **Minimal** | *Barely present. Hardly anything is happening.* |
| • | **Present** | *Something is here. Measurably active.* |
| : | **Fragmented** | *Parts falling apart. Low coherence.* |
| I | **Stable** | *Clearly present and centred. Holding steady.* |
| ! | **Alarm** | *High activation. Intense. Difficult to ignore.* |
| * | **Active** | *Engaged, focused, doing something with the field.* |
| # | **Saturated** | *Too much. Overwhelmed. Full beyond capacity.* |
| ? | **Searching** | *Not yet clear. Looking for direction or meaning.* |

## V — Movement / vector layer

| Symbol | Label | Simple description |
|---|---|---|
| . | **Still** | *Barely moving. Almost no change.* |
| > | **Direction** | *Moving somewhere. A clear path.* |
| >> | **Momentum** | *Moving fast and focused. Strong drive.* |
| ↑ | **Rising** | *Building up. Energy or pressure increasing.* |
| ↓ | **Decreasing** | *Winding down. Intensity or speed dropping.* |
| ↺ | **Repetition** | *The same pattern returning. A loop.* |
| ⟳ | **Reorientation** | *Deliberately changing direction. A reset.* |
| ~ | **Fluctuation** | *Irregular, variable movement. Not yet settled.* |

## O — Field / context layer

| Symbol | Label | Simple description |
|---|---|---|
| o | **Enclosed** | *Coherent, internally consistent. Naturally bounded.* |

| Symbol | Label | Simple description |
|---|---|---|
| `()` | **Protected** | *Clearly bounded. Contained by active pressure.* |
| `)(` | **Tense** | *Compressed, repellent. Not much room to move.* |
| `[]` | **Framework** | *Structured. Rules or constraints define movement.* |
| `~` | **Unstable** | *Variable field. No clear form, fluctuating.* |
| `∞` | **Expansive** | *Wide open. Maximum space. Low tension.* |
| `×` | **Blockage** | *Cannot pass through. Stopped by a hard boundary.* |
| `:` | **Fragmented** | *Scattered. Loose parts without a shared field.* |