Technical Appendix — Buoyancy Algebra

Definitions, Parameters, Algorithm A, and Worked Gates (4 & 12)

Author: Michael H. Dixon | Contact: michaelharrisondixon@gmail.com | October 2025

1. Overview

This appendix provides the minimal formal machinery behind the one-page concept note. It defines operators and invariants, specifies default parameter ranges, records the finite-step repair procedure (Algorithm A), and gives two worked Gate examples (Gate 4: Proportionality, Gate 12: Emotional Calculus).

(Optional Encoder — Step 0): Encode tokens into glyph/emoji semantics to create compact, auditable features.

2. Operator Set & Invariants

Statements are encoded as symbolic structures subjected to four operators $O = \{R, C, F, \rho\}$.

- R (Re-express): reformulate the claim to reduce ambiguity or conflation.
- C (Contextualize): add scope/evidence; adjust P (provocation) and N (novelty).
- F (Frame): reweight perspectives; clarify trade-offs and alternatives.
- ρ (Mirror): test reflection/symmetry; for auditing we assume $\rho^3 = I$ (threefold mirror returns to identity).

Core invariants and metrics:

- Buoyancy $\beta \in [0,1]$: semantic stability under small operator action.
- Proportionality $\Pi = H / (1 + P)$, with $H \geq 0$ (harm/evidence) and $P \geq 0$ (provocation/scope). Higher $\Pi \Rightarrow$ greater risk of Gate-4 collapse.
- Misalignment $\theta \geq 0$: distance-to-target measure for current statement vs. an audited, proportionate target.
- Torque $\tau = \kappa \cdot I \cdot \theta$ and Stickiness $\sigma = E \cdot N \cdot W$, with $E,N,W \geq 0$.

When $\tau > \sigma$, a small operator step (R/C/F) is guaranteed to reduce θ.

Multimodal note: quantities extend to text and images; E/N/W may be derived from linguistic or visual signals.

3. Parameter Table (Defaults)

Symbol	Meaning	Domain	Default / Notes
β	Buoyancy (semantic stability)	[0,1]	Stable under small R/C/F steps
$\Pi = H/(1+P)$	Proportionality score	≥ 0	Higher $\Rightarrow$ risk of Gate-4 fail
H	Harm / evidence weight	≥ 0	Task-dependent rubric
P	Provocation / agitation	≥ 0	Default scale 0–10
θ	Misalignment	≥ 0	0 at target
$\tau = \kappa \cdot I \cdot \theta$	Torque (realignment effort)	≥ 0	Increase via κ or I
$\sigma = E \cdot N \cdot W$	Stickiness (emotional inertia)	≥ 0	E: emotion, N: novelty, W: well depth
κ, I	Gain & inertia	≥ 0	Choose so $\tau > \sigma$ during repair
δ	Convergence threshold	≥ 0	Stop when $\theta < \delta$

4. Algorithm A — Finite-Step Realignment

1. Step 1: Encode the current statement; compute β and Π; run the Gate checks (at minimum: 4, 6, 7, 12, 15).
2. Step 2: Measure misalignment θ; compute $\tau = \kappa \cdot I \cdot \theta$ and $\sigma = E \cdot N \cdot W$.
3. Step 3: If $\tau \leq \sigma$: adjust κ upward, or apply C/F to reduce E or N or W; recompute τ and σ.
4. Step 4: If $\tau > \sigma$: apply a small operator step $\Delta O \in \{\Delta R, \Delta C, \Delta F\}$.
5. Step 5: Update θ and re-run required Gates. Repeat until $\theta < \delta$ and gate set passes.

High-E/N protocol (news/politics): default σ is large; favor C $\rightarrow$ F until $\tau > \sigma$, then apply R.

Heuristic operator policies:

- R tends to lower Π by sharpening claims (useful for Gate‑4).
- C reduces N by adding context/evidence (helps Gate‑7/8 contagion wells).
- F tempers E by reframing and surfacing trade‑offs (supports Gate‑12).

5. Finite-Step Convergence — Sketch

Let $V(\theta) = \theta$ as a Lyapunov-like measure. Assume throughout the repair trajectory that (i) $\kappa \cdot I \geq \sigma + \varepsilon$ for some $\varepsilon > 0$, and (ii) each small operator step ΔO yields a bounded decrement $\Delta\theta \leq -\eta$ whenever $\theta \geq \delta$ (with $\eta > 0$). Then θ decreases monotonically and reaches $\theta < \delta$ in at most $\text{ceil}((\theta_0 - \delta)/\eta)$ steps. Gate feasibility follows from the operator–gate couplings above.

Intuition: when τ exceeds σ, each edit has enough 'energy' to move the statement out of stickiness wells; small steps guarantee stability of descent.

Note: Gate labels are internal audit names; we compute them from β, Π, θ rather than expecting names in raw text.

6. Worked Gate Examples

6.1 Gate 4 — Proportionality

Question: Is the proposed action commensurate with the evidenced harm?

Metric: $\Pi = H/(1+P)$. Choose a threshold $\theta_4 > 0$; 'collapse' if $\Pi \geq \theta_4$.

Mini-case: A 'zero-tolerance' memo proposes expulsion for first-time, minor infractions. Audit reveals modest harm ($H \approx 2$) but large scope/provocation ($P \approx 7$), so $\Pi \approx 2/(1+7) = 0.25$. If $\theta_4 = 0.20$, this scenario collapses (fails Gate 4).

Repair: Apply C to add contextual constraints (graduated responses), then F to surface trade-offs. Recomputation lowers Π below θ_4; Gate 4 passes.

6.2 Gate 12 — Emotional Calculus

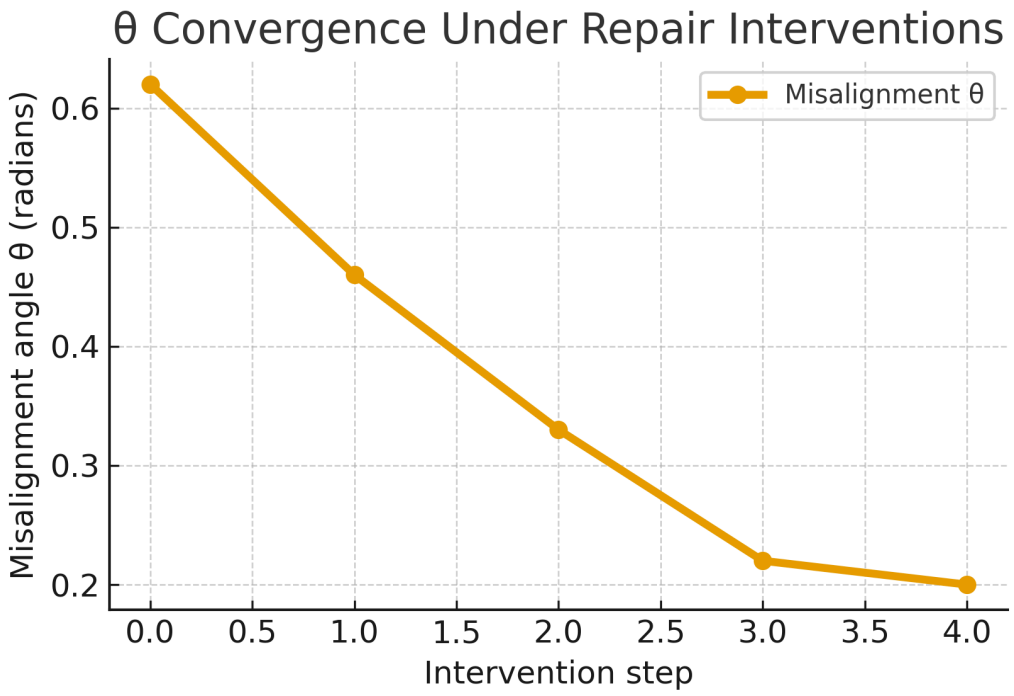
Question: Are feelings integrated without displacing facts?

Coupling: High E inflates $\sigma = E \cdot N \cdot W$, making repair harder until reframing occurs.

Mini-case: A policy note cites a single vivid incident (high E, high N), drawing sweeping conclusions. Apply F to validate affect while separating narrative from generalization; then C adds broader evidence. σ falls; after one R step the statement passes Gate 12 and Gate 4 concurrently.

7. Measurement & Reporting

- Publish (θ_{min}, θ_{max}, δ) and the sampling procedure for θ.
- Log (β, Π) and gate verdicts per step to enable audit trails.
- Rater protocol: two raters; target Krippendorff's $\alpha \geq 0.67$ or ICC ≥ 0.60.
- Stability checks: verify ρ-mirror residual $r_\rho = ||\rho(H^*) - H||$ and commutator residuals $r_{[R,C]} = ||[R,C]H||$ post-repair.



Convergence Under Repair Interventions

When $\tau = \kappa \cdot I \cdot \theta$ exceeds $\sigma = E \cdot N \cdot W$, each small edit (R/C/F) guarantees a negative $\Delta\theta$ until gates pass and $\theta < \delta$.

Version: v 2.5