
Emotion as System

A Foundational Architecture for Affect, Meaning, Perception, and Action

Brendon Hawkins · Studio Lab BH

ABSTRACT

Individual-level affective science explains how a single mind constructs an emotional state. This paper proposes the next layer up. It offers a foundational systems architecture that describes emotion as structural behavior, operable at individual, collective, and computational scales. Within this architecture, affect operates as an energetic substrate. Emotion configures a global field of sensitivity and responsiveness. Meaning introduces mass that deforms the field. Perception traces trajectories through it. Action follows gradients shaped by its geometry. The vocabulary of field, mass, trajectory, and gradient is precise enough to be operationalized computationally, tested empirically, and extended through design. Recent interpretability work on large language models recovers the same primary axes of valence and arousal, and the same clustered geometry, that organize human affective experience. This provides convergent evidence that emotion has computable structure. The contribution here is architectural: it supplies a system definition that makes emotion available as an object of formal reasoning across disciplines, and it begins where individual-level frameworks end.

0. SCOPE, INTENT, AND EPISTEMIC POSITION

0.1 Purpose of This Paper

This paper defines a foundational systems framework for understanding emotion as a generative, evaluative, adaptive, and translational system. Emotion is treated here as structured system behavior that mediates between affect, meaning, perception, and action — continuous, configurable, and constitutive of experience rather than incidental to it.

The framework developed in proximity to a specific set of works and practices: Olafur Eliasson's perceptual installations, Baudrillard's account of simulacra, Lisa Feldman Barrett's construction of emotion as active inference, Andy Clark's situated agents framework, Lakoff and Johnson's structural account of metaphor, Von Bertalanffy's general systems theory, the NRC VAD lexicon, and Gabriel's and Keeling's work on AI values and alignment. These are named as the intellectual environment the framework was built inside.

The framework is operational. Metaphors here function as models — they encode system behavior with structural precision. Diagrams are formal reasoning tools that carry argumentative weight equal to the prose. Spatial descriptions are representational mappings, reflecting how the system behaves rather than where it physically resides.

By this we mean: where an individual-level framework asks what is this person feeling and why, this framework asks what is the shape of the field they are feeling within, and how is that field configured across people, time, and media.

0.2 Interdisciplinary Positioning

This work sits at the intersection of several domains that have independently addressed aspects of emotion but rarely share a common structural language. Cognitive linguistics, affective neuroscience, systems theory and cybernetics, computational and geometric modeling, and design research and media theory each contribute a partial account. Together they outline the terrain this framework crosses.

Concepts from these domains are referenced as supporting contexts. The framework draws selectively from each to construct a system-level description that remains usable across all of them.

1. MOTIVATION FOR A SYSTEMS DEFINITION

1.1 The Spatial–Energetic Organization of Emotional Experience

Emotional experience is consistently organized and reported using spatial, energetic, and geometric terms. Across languages and cultural contexts, people describe emotional states in terms of proximity and distance, weight and load, orientation and direction, intensity and charge. These descriptions are stable patterns through which emotional experience is rendered intelligible.

The recurrence of spatial and energetic language reveals that emotional experience is not apprehended as a collection of isolated internal states. It is experienced as structured variation within a constrained space — one in which relationships, gradients, and limits matter. This organization appears regardless of whether the speaker adopts scientific, colloquial, or artistic language.

Emotional experience is governed by structure, constraint, and movement. This is the operating condition of the system — and it motivates treating emotion as a form of system behavior with geometric properties that can be named, mapped, and reasoned about.

For example: a grieving person reports feeling *heavy* and says the world has gone *flat*. A person in love reports feeling *light*, *open*, *pulled toward* another. A person under threat reports the room *closing in*. This is the most precise language available for describing what the system is doing.

1.2 Limits of Discrete Emotion Frameworks

Much existing work on emotion relies on discrete emotional categories — stable classificatory tools that compress complex, continuous variation into nameable states. This approach descends from early theorizations of emotional

essence, the notion that each emotion possesses a fixed, identifiable core that can be reliably recognized across individuals and contexts. These frameworks are effective for communication and certain forms of measurement. Applied to emotional dynamics, however, they impose significant limits.

Discrete emotional categories obscure key properties of emotional experience:

- continuity and gradation across affective states
- overlap and simultaneity between emotional configurations
- transitions and trajectories over time
- body state variability — the same named category produces fundamentally different physiological responses across individuals
- cultural variability — emotional categories carry different meanings, thresholds, and expressions across cultural contexts
- coupling between perception and action

Consider what discrete categories cannot capture. The English word *anxiety* compresses anticipatory dread, social hypervigilance, and chest-tightening physiological arousal into a single label. German *Angst* carries metaphysical weight English does not. Japanese *amae* names a relational emotional configuration — the comfort of presuming on another's care — that has no English equivalent at all. Tahitian has no general word for *sadness*; instead it differentiates fatigue, social loss, and physical illness as distinct affective conditions.

These are evidence that emotional categories are tools cultures build for the configurations that matter to them, and that any single language's categories give a partial map of the underlying system.

Because discrete frameworks privilege states over processes, they are poorly suited for describing how emotional configurations emerge, change, and guide behavior. They offer names for outcomes without providing a structural account of how those outcomes are produced.

STRUCTURAL BEHAVIOR OF THE EMOTIONAL SYSTEM

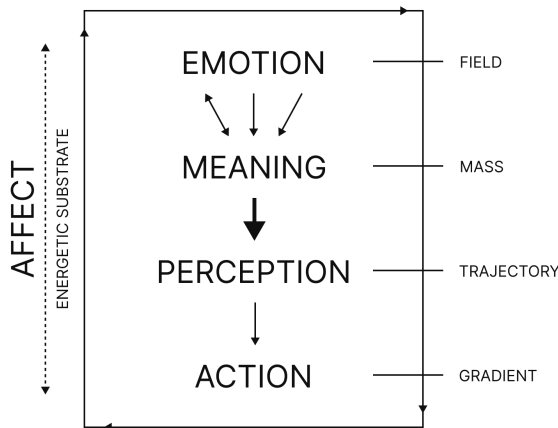


Fig. 1 — Structural Behavior of the Emotional System. Field, mass, trajectory, and gradient as structural roles; affect as energetic substrate.

2. CORE SYSTEM DEFINITION

2.1 Emotional System

An emotional system is defined as a predictive, adaptive system that structures affect into meaningful configurations that guide perception and action. It operates continuously — as an ongoing process of configuration, update, and translation.

The emotional system is:

- **Generative** — it produces states and anticipates futures, continuously constructing conditions of readiness and expectation
- **Evaluative** — it assesses affective conditions along dimensions of intensity, stability, and direction
- **Adaptive** — it updates through feedback, revising configurations in response to prediction error and changing conditions
- **Translational** — it renders emotional structure across senses and media, producing coherent expression from the same underlying configuration

- **Communicative** — it receives and transmits signals, coupling individual systems into shared affective environments

The emotional system governs how affective conditions are organized, stabilized, and transformed into patterns of sensitivity, interpretation, and action.

2.2 Affect as Energetic Substrate

Affect is defined as a continuous, low-resolution evaluative condition characterized by three primary dimensions:

- **Valence** — ranging from pleasant to unpleasant
- **Arousal** — ranging from calm to agitated
- **Dominance** — ranging from submissive to dominant

Affect is presemantic. It evaluates before it means — operating below the threshold of concept, category, and narrative. It supplies the conditions under which emotional structuring becomes possible. Emotional content is constructed downstream, by the system that sits on top of it.

This characterization draws on a well-established line of research. Russell's circumplex model (1980) established valence and arousal as the primary dimensions of core affect. Mehrabian's PAD model (1996) extended this with dominance. Barrett's theory of constructed emotion (2017) formalized affect as the continuous low-level evaluative signal from which categorical emotions are constructed by the brain in context. The NRC VAD lexicon (Mohammad, 2025) operationalized these dimensions across 55,000 English words, demonstrating that valence, arousal, and dominance are recoverable from natural language at scale.

Within the emotional system, affect functions as an energetic substrate. It is continuously available, capacity-constraining, and enabling of state transitions without determining their form. Affect conditions emotional activity while remaining beneath it — the ground the system stands on.

2.3 System Openness and Coupling

Emotional systems are open systems. They interact continuously with other systems at multiple scales — individual, social, and technological.

Coupling occurs through:

- shared affective conditions
- coordinated perception–action dynamics
- signaling and synchronization across agents

The outputs of an emotional system are actions and relational effects. Through coupling, emotional systems influence and are influenced by surrounding systems — allowing affective conditions and structured responses to propagate across individuals and environments.

2.4 Convergent Work at Adjacent Scales

This framework operates at the system level. Adjacent work operates at scales below it, and its findings converge with the architecture proposed here.

Adjacent: computational interpretability of emotion concepts. Sofroniew et al. (2026) extracted internal emotion concept representations from a frontier large language model (Claude Sonnet 4.5) and found that their geometry recovers the same primary organizing dimensions identified across decades of human affective research — valence ($r = 0.81$ with human ratings) and arousal ($r = 0.66$). Related emotions cluster intuitively. The representations are causally linked to behavior, including alignment-relevant behaviors such as reward hacking and sycophancy. This is convergent evidence, derived from a substrate (transformer activations) far removed from both the body and the cultural field, that emotion has computable geometric structure of the kind this framework describes.

Where this framework begins. This line of work describes representations within a single system: a model. The architecture proposed here opens the next scale — what happens when emotional systems couple, when fields interact, when meaning concentrates across populations, when

affective conditions propagate through shared media and infrastructure. Individual-level frameworks describe the unit. This framework describes the field that units configure together.

3. STRUCTURAL ARCHITECTURE OF THE EMOTIONAL SYSTEM

3.1 Emotion as Field

Emotion is modeled as a global field condition within the emotional system — the system's current configuration of sensitivity and responsiveness.

As a field, emotion:

- configures what signals are salient
- modulates sensitivity to internal and external variation
- establishes conditions of readiness and responsiveness

The emotional field shapes what differences matter, how strongly they matter, and which possibilities are amplified or dampened. It is the structural condition within which perception and action occur.

3.2 Meaning as Mass

Meaning is defined as accumulated relevance or value within the emotional system — built through experience, memory, learning, and social context.

Within the geometric model, meaning functions as mass:

- it introduces weight into the emotional field
- it deforms the field locally
- it exerts directional pull on perception and action

High-meaning elements stabilize certain paths and resist others. Meaning biases movement by altering the geometry of the field in which perception and action unfold.

A wedding ring weighs almost nothing. Its meaning weighs a great deal. In this framework, that second weight is literal system behavior — the ring deforms the field around it, pulling attention, organizing memory, biasing action. Mass here is the structural role that importance plays. The ring's weight in the field is system

behavior, observable in where attention goes, how memory organizes, which actions become more available.

3.3 Perception as Trajectory

Perception is defined as guided movement through an emotional field under the influence of meaning. It is directional, biased, and constrained by the current configuration of the emotional system.

As a trajectory:

- perception follows paths shaped by field curvature and mass
- it reflects predictive movement rather than objective sampling
- it is sensitive to prior configuration and accumulated meaning

Because emotional fields differ across time and context, identical sensory inputs may yield different perceptual trajectories under different configurations. Perception, in this framework, is inseparable from the structure that guides it.

The same hallway looks different to the child arriving for their first day of school and to the same person returning thirty years later. The hallway is unchanged. The trajectory through it is not.

3.4 Action as Gradient

Action is defined as movement along gradients within the emotional field. Gradients represent directions of change shaped by field configuration and meaning-induced deformation.

Action follows paths characterized by:

- predicted relief
- least resistance
- steepest change within the current field

Action reflects the instantaneous geometry of the system. It emerges from the same structural conditions that shape perception, extending perceptual trajectories into behavior — continuous with emotion and meaning rather than separate from them

4. SYSTEM DYNAMICS

4.1 Predictive and Recursive Operation

The emotional system operates as a continuous predictive loop. It anticipates sensory input, bodily demand, and potential action outcomes — updating continuously rather than reacting to discrete stimuli.

Each cycle involves:

- prediction
- simulation
- comparison
- error correction

Each cycle updates the configuration of the emotional field, adjusting sensitivity, salience, and directional bias. These updates occur continuously and typically outside of conscious awareness.

State values change — affective intensity, field curvature, meaning distribution. The structural roles remain invariant. Emotion continues to function as field, meaning as mass, perception as trajectory, and action as gradient.

4.2 Capacity, Bandwidth, and Constraint

Emotional systems operate under finite capacity. At any moment, there are limits to how much affect can be sustained, differentiated, and translated into adaptive perception and action.

These constraints arise from:

- embodiment and physiological regulation
- cognitive and attentional limits
- energetic availability and depletion

Within this framework, emotional intelligence, regulation, and maturity are reframed as forms of system-level capacity management. They describe the system's ability to maintain functional structure across a wider range of affective intensity, ambiguity, and duration. Capacity is situational and variable — a property of the system under specific conditions, not a fixed attribute of the person. What appears as emotional intelligence in one context may reflect available capacity rather than stable trait.

This reframing has consequences. It makes capacity measurable, context-dependent, and subject to design — opening questions that trait-based frameworks cannot ask.

4.3 Rendering and Translation

Emotional structure is rendered outward through expressive and interface functions — language, gesture, posture, sound, color, movement, and behavior. These renderings translate internal system configuration into externally perceivable signals.

Rendering is goal-dependent, context-sensitive, and shaped by cultural and situational conditions. Because emotional structure is defined at the level of system configuration, it is modality-agnostic — no single sensory or communicative modality is privileged as the primary carrier.

The primary function of rendering is synchronization. Through shared concepts and communicative interfaces, emotional systems coordinate affective conditions, align expectations, and influence perception–action dynamics across individuals and environments. Emotional communication operates as a mechanism of coupling and regulation — extending the system outward into shared affective space.

4.4 Failure Modes

When capacity limits are exceeded or structural constraints are violated, the emotional system enters failure modes — conditions in which the system's geometry no longer supports adaptive perception or action.

Representative failure modes include:

- **Field saturation** — excessive salience flattens differentiation
- **Meaning overload** — accumulated mass produces perceptual lock-in
- **Gradient trapping** — action repeatedly follows the same paths despite negative outcomes

These conditions are systemic. They describe breakdowns in functional organization — limits of capacity and constraint within the system, indicators of where the architecture requires support or intervention.

AFFECTIVE SYSTEMS MATRIX

The diagram is a 4x3 grid. The vertical axis is labeled 'SOCIAL SCALE' with an upward arrow, and the horizontal axis is labeled 'TEMPORAL RESOLUTION' with a rightward arrow. The grid cells contain the following text:

	AFFECT	EMOTION	SENTIMENT
INDIVIDUAL	PHYSIOLOGICAL SUBSTRATE	CONSCIOUS APPRAISAL	NARRATIVE INTERPRETATION
DYADIC	AFFECTIVE COUPLING	EXPRESSIVE SIGNALING	RELATIONAL MEANING
GROUP	COLLECTIVE AROUSAL	COORDINATED BEHAVIOR	SHARED NARRATIVES
CULTURE	HISTORICAL TONE	NORMALIZED RESPONSE	IDEOLOGY

Fig. 2 — Affective Systems Matrix. Affect, emotion, and sentiment as three registers of affective experience, distributed across social scale (individual, dyadic, group, culture) and temporal resolution.

5. IMPLICATIONS OF THE FRAMEWORK

5.1 Why Geometry Emerges

When a system operates continuously through prediction and recursive updating under constraint, its behavior stabilizes into structured patterns. Repeated successful configurations of perception–action coupling leave residual organization. That residual organization is geometry.

Geometry accumulates through repeated operation. Each cycle of prediction, simulation, and error correction reinforces existing paths or carves new ones, gradually deforming the field, redistributing mass, stabilizing trajectories, and sharpening gradients. The geometric vocabulary of this framework is a description of what predictive systems produce when they operate within bounded capacity over time.

Eliasson understood this in perceptual terms — that encounter with a structured environment leaves structural residue in the perceiver. Baudrillard understood it in semiotic terms — that repeated simulation produces effects as real as

the original. Both observations point toward the same underlying dynamic. The emotional system is always building geometry from experience, and that geometry in turn shapes every subsequent encounter.

Spatial and energetic descriptions of emotion — direction, weight, proximity, intensity — correspond to this accumulated structure. They are stable because the system that produces them is stable. They are geometric because geometry is what structured, constrained, predictive systems leave behind.

5.2 Why Collectivity Matters

Emotional systems do not operate in isolation. Through coupling, multiple emotional systems interact to form affective ecosystems. Within these ecosystems, affect circulates as a shared energetic substrate, while emotional field configurations influence one another through synchronization and feedback.

At the collective level:

- affective intensity may amplify or dampen across systems
- meaning may concentrate or disperse through shared concepts
- field configurations may polarize or align

These dynamics are emergent properties of coupling — arising from interaction, coordination, and constraint across systems rather than from summation or consensus alone.

The interpretability findings discussed earlier (Sofroniew et al., 2026) describe a related dynamic at a different scale. Within a single language model, emotion concept representations are reused across speakers, and the model's outputs propagate emotional configuration into its responses without persistently encoding any one character's state. This is a within-system parallel to what couples across systems at the collective level — the same operative emotional configuration, reactivated and rendered, shaping what comes next.

5.3 Why Rendering Is Plural

Because emotional structure is defined at the level of system configuration, it is

modality-agnostic. No single sensory or communicative modality is privileged as the primary or authentic carrier. Translation across modalities alters surface form while preserving underlying structure — what remains invariant is configuration, not expression.

Emotional coherence can persist across radically different forms of representation because the geometry that produces it is independent of the medium through which it appears. A computational rendering carries the same structural validity as a linguistic one. A color field is as precise a rendering as a gesture. The medium is an interface, not the emotion itself.

This has consequences for how systems that engage with emotion are designed, evaluated, and extended. Authenticity is not a property of the modality. It is a property of the configuration beneath it.

6. LIMITS AND OPENINGS

This framework is bounded by design. Where the architecture becomes transferable is where its boundaries are drawn. Precision at the system level requires leaving questions of biological implementation, empirical validation, and normative application open. Those openings are intentional.

What this framework provides is a set of structural conditions precise enough to be shared, reused, and extended. Its limits are its portability.

The openings are specific. Empirical validation of field-level dynamics across coupled systems remains to be done. Computational instantiation of the architecture — engines that operate on field, mass, trajectory, and gradient as primitives — is in active development at Studio Lab BH. Application to design, public health, and AI alignment, where collective affective dynamics are operative but undertheorized, is where the framework expects to earn its keep. This paper supplies the architecture. The work that follows will test it.

7. CONCLUSION

Emotion has structure. That structure is geometric, and the geometry is the same whether the system in question is a person, a population, or a model. Individual-level affective science describes how one mind constructs an emotional state. Computational interpretability now shows that artificial systems trained on human language recover the same primary axes — valence, arousal — that organize human affect. This framework names the architecture that connects them and extends upward into the collective scale, where emotional systems couple, fields interact, and meaning concentrates across populations.

Field, mass, trajectory, gradient. Affect as substrate. Emotion as configuration. Perception as movement. Action as descent along the gradient the field has shaped. These are the primitives. They are precise enough to compute, observe, and design with. The work ahead is instantiation, measurement, and application — turning architecture into instrument.

REFERENCES

- Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.
- Baudrillard, J. (1981). *Simulacra and Simulation*. Éditions Galilée.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204.
- Gabriel, I., & Keeling, G. (2025). A matter of principle? AI alignment as the fair treatment of claims.
- Hoemann, K., Nielson, C., Yuen, A., Gurera, J. W., Quigley, K. S., & Barrett, L. F. (2021). Expertise in emotion: A scoping review and unifying framework for individual differences in the mental representation of emotional experience. *Psychological Bulletin*.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
- Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4), 261–292.
<https://doi.org/10.1007/BF02686918>
- Mohammad, S. M. (2025). NRC VAD Lexicon v2: Norms for valence, arousal, and dominance for over 55k English terms. National Research Council Canada.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv:2203.02155.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Sofroniew, N., Kauvar, I., Saunders, W., Chen, R., Henighan, T., Hydrie, S., Citro, C., Pearce, A., Tarng, J., Gurnee, W., Batson, J., Zimmerman, S., Rivoire, K., Fish, K., Olah, C., & Lindsey, J. (2026). Emotion concepts and their function in a large language model. Transformer Circuits Thread.
- Vishnubhotla, K., Teodorescu, D., Feldman, M. J., Lindquist, K. A., & Mohammad, S. M. Emotion granularity from text: An aggregate-level indicator of mental health.
- Von Bertalanffy, L. (1968). *General System Theory: Foundations, Development, Applications*. George Braziller.