

Emotional Differentiation

Categorization, Prediction, and the Formation of Conscious Feeling

Flyxion

August 27, 2026

Abstract

Bodily signals do not arrive already divided into anger, fear, hunger, attraction, exhaustion, or grief. Similar physiological changes participate in many different experiences, while instances assigned to the same emotional category can differ substantially in physiology, expression, and action. This essay argues that the brain resolves this variability by using prior experience, present context, interoceptive condition, sensory evidence, and anticipated consequence to construct a workable interpretation of the organism's state. This interpretation is predictive because it prepares the organism for what is likely to happen next; categorical because it treats variable instances as relevantly similar; and allostatic because its governing function is the prospective regulation of bodily need. Emotional differentiation is the acquisition of sufficiently precise distinctions to separate affective conditions that would otherwise remain undifferentiated. Conscious feeling, on this account, is not a transparent reading of the body but an organized regulatory achievement – one that can clarify experience, distort it, or require revision.

Introduction

The essay that follows treats emotion as a problem of differentiation rather than detection. It does not argue that emotions are illusions, that bodies are irrelevant to feeling, or that any interpretation of a bodily state is as good as any other. It argues instead that the nervous system does not begin with fully formed emotions waiting to be found. It begins with changing bodily conditions, ambiguous sensory signals, a history of prior categorizations, and predictions about what the present condition portends. Emotion is what results when these materials are organized into a category precise enough to guide action while keeping the organism's regulatory budget viable.

Two commitments run through the argument. The first is that categorization is not a decorative label applied after a feeling is already complete, nor an arbitrary social or linguistic imposition on raw sensation. It is a biologically constrained operation through which an organism differentiates its condition, interprets its significance, and narrows the field of available continuations. The second is an evidential discipline. Contemporary neuroscience supports strong claims about predictive regulation, distributed signal routing, interoception, and the developmental acquisition of categories. It does not, by itself, establish the stronger interpretive claims this essay will sometimes venture – about admissibility, about which distinctions ought to count as constitutive of feeling, about consciousness itself. Where the argument moves from what a study measured to what the essay proposes, that movement

will be marked rather than smoothed over. A theoretical framework earns the right to interpret evidence; it does not thereby inherit the evidence's authority.

1 The Problem of Emotional Kinds

1.1 The apparent obviousness of emotion

Nothing about emotional experience feels tentative from the inside. Anger arrives already labeled. Fear announces itself, sadness settles in, joy lifts without argument. This immediacy encourages an assumption that goes largely unexamined in ordinary life: that emotions are natural kinds, internally distinct states with characteristic causes, physiological signatures, facial expressions, and behavioral consequences, waiting to be correctly identified the way a species is identified from its markings. Ordinary language reinforces the assumption. Each emotion word is treated as naming a discrete mechanism, stable across people, cultures, and circumstances, differing only in intensity or disguise.

This assumption generates a specific empirical commitment. If emotions are biologically fixed kinds in something like the way species or elements are fixed kinds, then instances filed under the same emotional name should display strong, reliable physiological and behavioral similarity across occasions and across people. The commitment is testable, and it does not survive contact with the relevant evidence. A racing heart accompanies fear, but it also accompanies attraction, exertion, anger, fever, embarrassment, anticipation, and caffeine. Tears accompany grief, but also relief, physical pain, frustration, joy, and, on occasion, a calculated bid for sympathy. The question this essay takes as its starting point is therefore not primarily how the brain detects an emotion that is already there. It is how variable bodily and situational events come to be treated as instances of an emotion at all – and why the resulting groupings, however constructed, are not for that reason arbitrary.

1.2 Against emotional fingerprints

Call the view under scrutiny the fingerprint model: the expectation that each emotion possesses a distinctive neural, physiological, expressive, or behavioral signature, such that a sufficiently fine-grained measurement could in principle read the emotion straight off the body. The argument against this view does not require claiming that physiology is irrelevant to emotion, or that every emotional classification is equally defensible. It requires only establishing that the mapping between bodily pattern and emotional category is many-to-many rather than one-to-one: a given physiological configuration is compatible with several emotional interpretations, and a given emotional category can be physiologically realized in several different configurations, depending on context, history, and anticipated consequence.

The same structure appears in facial expression, which is often treated as the fingerprint model's strongest piece of evidence. A facial movement is an observable event; the assignment of a specific emotional meaning to that movement is an inference, not a readout. Context, social convention, prior interaction, posture, accompanying speech, and temporal sequence all alter what a given facial configuration is taken to indicate, and they can alter it enough to reverse the inferred emotion entirely while the muscular pattern stays constant. An observer who treats visible conduct as containing its interpretation transparently mistakes their own

categorization for direct perception – an error with consequences that extend well beyond the laboratory, and one this essay will return to in the discussion of imposed categories in Part VII.

1.3 Variation is not noise

Traditional models of emotion often treat variation within a category as measurement error scattered around a hidden prototype: real fear has an essential form, and any instance that departs from it is simply a noisy or atypical case of the same underlying thing. This essay reverses that interpretation. Variation is not noise obscuring the phenomenon. It is the phenomenon a theory of emotion has to explain.

Fear while hiding from an immediate physical threat is not physiologically or behaviorally identical to fear before an examination, fear of social rejection, or fear elicited by a film with no danger to the viewer at all. These instances belong to a common category not because each expresses one underlying physical template with cosmetic variation layered on top, but because a nervous system – and often a linguistic community alongside it – has learned to treat certain differences among them as irrelevant for some purpose, while treating other differences, invisible to a purely physiological description, as decisive. The explanatory burden accordingly shifts. The question is no longer where the essence of fear is hiding in the body. It is how an organism forms, learns, stabilizes, and on occasion revises the equivalence classes that let dissimilar events count as the same kind of thing to feel.

2 The Regulatory Ground

2.1 From homeostasis to allostasis

Two models of physiological regulation are worth distinguishing carefully, because the essay's account of emotion depends on which one is doing the explanatory work. Homeostasis, in its familiar sense, is the maintenance or restoration of physiological variables within a viable range: a thermostat that activates once temperature has already drifted outside its set bounds is a homeostatic device. Allostasis names something structurally different – prospective regulation, in which a system changes its activity in anticipation of an upcoming need rather than waiting for a deviation to register as an emergency.

A nervous system regulating allostatically must estimate, ahead of the fact, what the body is about to require. It prepares digestive and metabolic activity before a meal is fully processed, prepares cardiovascular and muscular readiness before exertion peaks, prepares defensive posture before a threat makes contact, and prepares social and attentional resources before a familiar interaction has fully unfolded. Prediction, on this view, is not an optional cognitive layer bolted onto a body that would otherwise regulate itself perfectly well through reaction alone. It is required by the plain temporal structure of a living system: regulation that begins only once damage or depletion has already occurred will, in enough cases to matter, begin too late to be useful.

2.2 The body budget

The concept of a body budget gives this section its functional center. Neural computation, immune response, movement, digestion, thermoregulation, sustained attention, and learning all draw on a shared, finite pool of metabolic resources, and an organism must allocate that pool under uncertainty about what will be needed next and when. Emotions, on the account developed here, are one of the principal means by which this allocation gets organized. Fear can prepare withdrawal, heightened vigilance, freezing, concealment, or active defense, depending on which of these the situation and the organism's history make most plausible. Anger can prepare confrontation or the enforcement of a boundary. Fatigue can suppress exploratory behavior in favor of conservation. None of these should be read as a rigid, hard-wired command issued by a discrete emotional module. Each is better understood as a context-sensitive configuration of perception, action-readiness, physiological adjustment, and anticipated cost, assembled for the situation at hand rather than retrieved intact from storage.

This is also the point at which the argument must resist becoming purely semantic – a dispute about which words to use for feelings that would otherwise be untouched by the relabeling. Emotional differentiation matters because different interpretations of the same ambiguous bodily condition license different regulatory policies. Interpreting a racing heart as danger and interpreting it as excitement are not two descriptions of an identical underlying event; they are two different preparations for what should happen next, with different costs attached to getting the interpretation wrong.

2.3 Affect before emotion

It is useful at this point to distinguish affect from emotion, rather than treating the two as synonyms differing only in formality. Affect refers to more elementary dimensions of bodily significance – centrally, valence and arousal – without assuming that these dimensions arrive already parsed into a named emotional state. A person can feel generally bad, activated, heavy, restless, constricted, overheated, empty, or unsettled without yet knowing whether the underlying condition is grief, hunger, fear, incipient illness, anger, loneliness, or simple exhaustion. This is not an absence of bodily information reaching awareness. It is, on the account this essay develops, a state of comparatively low emotional differentiation: real affective signal, not yet organized into a sufficiently specific category to support targeted action.

Emotion, in this sense, arises when affective and sensory material is organized into a more specific, actionable interpretation – one precise enough to narrow what should happen next rather than merely registering that something matters. The relation between the two is not a clean conveyor belt in which raw sensation is first completed and only then handed off to cognition for labeling. Bodily sensing, prediction, categorization, and the preparation of action continually shape one another during the same interval, rather than occurring as separable, sequential stages.

3 Prediction and Interoception

3.1 The body is inferred

Interoception names the sensing and regulation of the internal body: signals associated with cardiovascular activity, respiration, temperature, digestion, metabolic demand, immune activity, pain, and related internal conditions. The conscious experience of one's own body is not a comprehensive internal measurement reaching awareness intact. Many bodily variables never become available to consciousness at all, and the signals that do arrive are ambiguous, delayed, filtered, and integrated across several processing stages before they take any experiential form. The brain must therefore infer bodily condition from evidence that is real but incomplete. What a person feels is neither a fictional construction detached from physiology nor a transparent copy of it. It is a controlled interpretation, constrained jointly by bodily signal, learned expectation, present context, and regulatory demand.

This inference does not occur against a neutral or context-free background. Cortical processing is massively recurrent rather than predominantly feedforward: within primary visual cortex, the substantial majority of synapses arriving are feedback or lateral connections originating elsewhere in cortex, rather than ascending signal from earlier sensory stages, with some anatomical estimates placing the proportion of feedback input near ninety percent (Markov et al., 2011; Barrett and Miller, 2026). Barrett and Miller review this anatomical evidence, along with converging results from electrophysiology and neuroimaging, and interpret the pattern as showing that categorization is not confined to a late, terminal stage of perception but participates in signal processing from its earliest points (Barrett and Miller, 2026). The anatomical facts and the interpretive claim are not the same kind of statement. The synaptic ratios are measured; the claim that this recurrence amounts to continuous, predictive categorization is Barrett and Miller's proposed reading of what the anatomy supports. Both are worth taking seriously, but only the first is properly called a finding.

For the purposes of this essay, the more modest and better-supported claim is the one that matters most: incoming interoceptive and exteroceptive signals do not arrive at an otherwise unprepared cortex. They are met, from the outset, by an active predictive context assembled from prior experience and present regulatory need. Whatever else that context does, it is not a passive backdrop against which raw sensation gets pictured.

3.2 Prediction is preparation

Prediction, as the term is used here, is defined operationally rather than mystically. A predictive system uses prior regularities to prepare perception and action before all relevant evidence has arrived. Predictions can influence which sensory differences receive attention, how ambiguous signals get interpreted, and which motor or autonomic responses are made ready. Incoming signals may then confirm those preparations, modestly adjust them, or, in sufficiently discrepant cases, destabilize them outright.

Direct evidence for expectation-sensitive routing of this kind comes from outside the domain of emotion altogether. In laminar recordings across five cortical areas in macaques performing a visual working-memory task, Bastos and colleagues found that predictable and unpredictable stimuli were processed through measurably different pathways: unpredictable

stimuli were associated with increased spiking and gamma-band activity propagating through superficial cortical layers, while predictable stimuli were associated with enhanced alpha and beta activity and with feedback connectivity concentrated in deeper layers (Bastos et al., 2020). These are reported measurements – differences in spike rate, oscillatory power, connectivity, and cortical layer, tied to whether a visual stimulus had been predictable. The authors interpret this pattern through a model they call predictive routing, on which lower-frequency feedback signals prepare the pathways associated with an expected input by inhibiting the gamma-mediated processing that would otherwise carry it forward, so that an input matching the prediction is comparatively suppressed while a mismatched input propagates with less resistance (Bastos et al., 2020). Predictive routing is the authors’ proposed explanation of the measured layer- and rhythm-specific effects; it is not itself something the recordings directly observed.

This essay takes the Bastos result as evidence that expectation can measurably alter which sensory differences propagate through a cortical hierarchy, and it takes predictive routing as one plausible account of how that alteration might be implemented. It does not treat the study as direct evidence about emotional categorization, interoception, allostasis, or the regulation of feeling. The experiment concerned visual predictability in a working-memory task; nothing in it was designed to speak to affect. Its relevance to the argument of this essay is illustrative rather than probative: a demonstration, in an unrelated domain, that prior expectation can reorganize which evidence advances rather than merely biasing how already-processed evidence gets read.

Read this way, the rejection of a simple “the brain filters out predictable information” picture becomes important. In the Bastos data, the modulation of alpha, beta, and gamma activity by predictability was stimulus specific: suppression was strongest at recording sites that preferred the particular stimulus being predicted, not applied as a uniform blanket across the visual field (Bastos et al., 2020). Whatever process is doing the preparing, it is targeted rather than global – consistent with the broader claim that predictive regulation, wherever it operates, adjusts specific pathways rather than imposing an undifferentiated gate on all incoming signal.

3.3 Prediction error without homunculi

A discrepancy between prediction and incoming signal need not be understood as a message inspected by an internal observer who then decides what to do about it. Prediction and sensory constraint are distributed processes, and different levels of neural organization constrain one another through recurrent activity rather than reporting upward to a central arbiter. A given discrepancy may be absorbed as noise at an early stage, may adjust a lower-level estimate without reaching anything that could be called awareness, may redirect attention toward the source of the mismatch, or may propagate far enough to reorganize the interpretation of the entire situation. Nothing in this list requires a homunculus receiving the error signal and grading its significance; the significance is realized in what the discrepancy is subsequently allowed to do.

This is the point at which the wider theoretical framework this essay draws on becomes tempting to invoke directly, and where the temptation should be resisted rather than indulged. The neuroscience reviewed above supports selective propagation and differential

influence: some signals advance further than others, depending on how well they match an active prediction, and that differential advancement is layer-specific, rhythm-specific, and stimulus-specific rather than uniform. Describing this pattern as a kind of *admission* – treating unmatched signal as gaining access to further processing that matched signal is denied – is a natural redescription at the level of a broader theoretical vocabulary. But it is a redescription, not a finding independently established by the experiments themselves. Nothing in Bastos and colleagues’ recordings distinguishes a signal that has been “registered but not admitted” from a signal that has simply been suppressed; the vocabulary of admission adds no measurement the gating and inhibition language does not already supply. The honest position is that predictive routing *resembles*, at a certain level of abstraction, a mechanism for selectively determining which differences may advance – and that resemblance is worth naming, because it lets a single line of research bear on questions the original experimenters were not asking. It is not evidence that the nervous system implements admissibility in any stronger or more specific sense than that.

4 Categorization

4.1 Categories as controlled equivalences

A category allows different events to be treated as equivalent for a particular purpose. Categorization, on the view developed here, does not require discovering some identical property shared by every instance filed under a common name. It requires preserving the differences that matter for the purpose at hand while disregarding those that do not matter for it. This is close to the working definition offered by Minda and colleagues in their review of the categorization literature: categories are cognitive groupings whose members are treated as functionally equivalent, in a way that permits knowledge acquired from some instances to guide classification of, and action toward, others (Minda et al., 2024). An emotional category, by extension, can group together physiologically different events because they support sufficiently similar predictions, social interpretations, or action policies – and, just as importantly, nearly identical bodily states can be placed into different categories once a change in context alters their implications. Categories, on this account, are neither wholly internal to the individual organism nor simply imposed from outside by social convention. They emerge from the interaction among bodily regularity, environmental structure, learning history, language, and practical need, none of which is sufficient on its own to fix where a category’s boundary falls.

This functional-equivalence picture leaves open an important empirical question that the categorization literature has not settled: whether the acquisition of categories of different structural kinds relies on one underlying learning system or several. Minda and colleagues’ review surveys rule-based, similarity-based, exemplar-based, single-system, and multiple-system accounts of category learning, and their central conclusion is that the dispute between single-system and multiple-system architectures remains empirically unresolved (Minda et al., 2024). This matters for the present essay because it forecloses a tempting oversimplification: nothing in Barrett and Miller’s account of predictive, context-driven categorization by itself settles how the categories in question are learned in the first place. A broad claim about what categorization *is* – a continuous, predictively organized process spanning the

cortical hierarchy – is compatible with considerable diversity in how any particular category, emotional or otherwise, comes to exist. The essay’s later discussion of emotional development (Part V) should be read with this openness in view rather than as an extension of one committed learning architecture.

4.2 Categorization is prospective

The strongest link between categorization and the regulatory picture developed in Part II is that categorization is valuable precisely because it reduces uncertainty about what should happen next. To categorize a sensation as hunger, panic, grief, infection, attraction, or exhaustion is to alter the field of actions that currently register as relevant. The category brings some continuations forward and relegates others to the background; it changes which further evidence gets sought, which memories become salient, and which bodily or social interventions come to seem appropriate. A category, on this view, is not simply a report about the present condition. It is a prediction-laden regulatory policy, adopted because of what it permits the organism to anticipate and do.

This relation can be represented schematically, without claiming more precision for the notation than the underlying claim actually has:

$$(B_t, E_t, H_t) \longrightarrow C_t \longrightarrow \hat{B}_{t+1}, A_t,$$

where B_t is the organism’s estimated bodily condition at time t , E_t is the current environment, H_t is learned history, C_t is the category constructed from these, $\hat{B}_{t+1}$ is the anticipated future bodily condition, and A_t is the action or regulatory policy the category makes available. The diagram states a structural claim – that categorization mediates between present condition and anticipated future in a way that shapes available action – rather than a quantitative model to be fit against data.

4.3 What “categorical” does not require

A natural but mistaken inference threatens to follow from everything said so far: if categorization is pervasive in neural processing, as Barrett and Miller propose, then the underlying neural populations responsible for a given category ought to be discoverable as discrete, cleanly bounded clusters – a population of “fear neurons” distinguishable from a population of “anger neurons” the way one brain region is distinguishable from another. Recent large-scale evidence weighs directly against this inference, and it is worth stating the point with some care, because it complicates rather than merely supplements the account this essay is building.

Posani and colleagues analyzed more than fourteen thousand recorded units across forty-three cortical regions and found that categorical organization depended heavily on the scale of analysis (Posani et al., 2026). At the scale of the whole cortex, categorical structure was present: task-relevant conditions could be reliably separated by simple readouts applied to the population as a whole. Within most individual cortical regions, however, cleanly categorical neural populations – discrete clusters of neurons whose activity divided sharply along category boundaries – were rare outside primary sensory areas. Local representations were instead diverse and high-dimensional, and it was precisely this diversity, rather than any

local clustering into category-specific subpopulations, that made the population’s activity separable by a downstream readout (Posani et al., 2026).

This finding constrains what the essay can claim without contradicting itself. Categorical experience, and categorical behavior at the level of the whole organism, does not require categorically partitioned neurons at the local level. An emotion category may be perfectly legible from a distributed, high-dimensional population of neural activity without being embodied in any internally homogeneous emotional module – without there being a dedicated fear circuit or anger center whose activation is both necessary and sufficient for the category to be instantiated. Read this way, Posani and colleagues’ result stands in a productive rather than contradictory relation to Barrett and Miller’s proposal. Barrett and Miller argue that categorization is pervasive across neural processing; Posani and colleagues show that this pervasiveness does not imply local categorical clustering. The two claims concern different meanings, and different scales, of the word “categorical,” and the essay’s later discussion of differentiation (Part V) will depend on keeping those meanings separate: differentiation need not mean that the brain sorts continuous activity into sharply bounded internal boxes. It may mean instead that a diverse, high-dimensional neural state becomes sufficiently separable for a context-sensitive readout to support different predictions and different actions – a considerably more modest, and more defensible, claim than the boxed-modules picture it replaces.

5 Emotional Differentiation

5.1 What differentiation means

Emotional differentiation is the capacity to make useful distinctions among affective states, rather than collapsing many forms of distress or activation into broad labels such as “good,” “bad,” “stressed,” or “upset.” The relevant measure of a distinction is not verbal ornamentation – acquiring more emotion words is not, by itself, differentiation in the sense intended here. A distinction is useful when it tracks a real difference in cause, in bodily condition, in likely development, or in the action a situation warrants. Differentiation therefore has both an epistemic and a regulatory role at once: distinguishing shame from guilt, hunger from anxiety, grief from a more diffuse depression, or attraction from threat can change what response is called for, not merely how the response is described afterward.

5.2 Development of emotional categories

Differentiation has a developmental history rather than arriving as a fixed endowment. Infants begin with broad affective contrast, bodily rhythm, and a strong regulatory dependence on caregivers, and only gradually acquire the capacity to distinguish recurrent patterns that were initially undifferentiated. Caregivers participate directly in this process: by interpreting a child’s signals, naming situations, and responding differentially to bodily and behavioral change, they do more than teach vocabulary. They help organize the expectations that come to govern what a given internal event is taken to mean and what actions are understood to follow from it. The resulting repertoire of categories grows in precision over development, unevenly, shaped by culture, and open to revision throughout life.

The claim that learning history shapes eventual category structure is not specific to human emotional development, and the strongest evidence for it currently comes from elsewhere. Collina and colleagues trained mice to categorize sounds along a continuum and found that individual learning trajectories differed substantially across animals, with stimulus-independent tendencies such as choice bias and perseveration measurably associated with where the resulting category boundary fell, particularly for the most ambiguous stimuli (Collina et al., 2025). This is direct evidence that category boundaries can be path-dependent products of the route by which they were learned, rather than a fixed function of the stimulus distribution alone – established, in this case, for auditory categorization in a nonhuman species with no bearing on emotion. Whether human emotional categories develop through a comparable process is a further, presently unestablished claim. Establishing it would require longitudinal evidence connecting documented human learning histories to measured variation in later emotional category boundaries; the mouse result licenses treating the hypothesis as worth investigating, not treating it as already confirmed.

A related but stronger inference is available from work conducted directly in humans. Roark and Chandrasekaran examined how the same individuals learn categories built from different distributional structures – rule-based categories that can be captured by a single verbalizable criterion, and information-integration categories that require combining information across dimensions in a way resistant to explicit verbal strategy – and found a mixture of stable individual tendencies and flexible, structure-dependent behaviors, shared across category types in some respects and distinct in others (Roark and Chandrasekaran, 2023). This is direct evidence that human category learning does not proceed through one uniform behavioral route even within an individual learner. It supports a pluralistic account of human categorization generally. Applying that pluralism specifically to emotional differentiation – the further claim that some emotional distinctions are acquired through verbalizable rules while others emerge gradually from exposure to multidimensional patterns the learner cannot fully articulate – remains a theoretically motivated extension of the finding rather than a result the study itself reports.

5.3 Granularity, uncertainty, and action

Greater emotional granularity can improve regulation because it narrows the range of plausible causes and responses. “I feel bad” leaves a great deal open; “I am embarrassed because I violated an expectation in front of people whose judgment matters to me” supplies a considerably more constrained model of what happened and what might help. But finer differentiation is not automatically better without limit. Categories can become excessively narrow, compulsively elaborated, or detached from any action they usefully inform; an endlessly proliferating taxonomy of feeling can create the appearance of insight while obscuring regulatory causes that several apparently distinct labels actually share. The appropriate target is not maximal differentiation but sufficient differentiation – enough precision to guide effective action without manufacturing unnecessary conceptual fragmentation.

This trade-off has at least two points of contact with categorization research conducted outside the domain of emotion. Using EEG during object categorization, Greene and Rohan found that information corresponding to the intermediate, “basic” level of category abstraction became usable very early in processing, at approximately fifty milliseconds, with

more specific, task-dependent distinctions emerging measurably later (Greene and Rohan, 2025). Their interpretation is that basic-level categorization is comparatively automatic and obligatory, a default level of description the visual system reaches before more particular or more abstract levels become available. This is direct evidence about object categories, not emotion categories, and it should not be read as establishing that emotional concepts possess a demonstrated neural “basic level” of their own. The more defensible extension is comparative: emotional differentiation may face an analogous optimization problem, in which categories must be specific enough to guide action without becoming so particular that they cease to generalize across occasions, and whether emotion concepts show a basic-level advantage comparable to what Greene and Rohan found for objects is an open empirical question rather than a settled parallel.

A second point of contact concerns the cost side of the trade-off rather than its early-processing side. Garner and Dux, reviewing evidence on knowledge generalization and multitasking interference, argue that representations shared across tasks facilitate rapid generalization but also produce interference when those tasks must be performed concurrently, and that extended practice can segregate task-specific representations in a way that reduces interference while correspondingly reducing flexibility and transfer to new situations (Garner and Dux, 2023). This is a theoretical synthesis of a body of reviewed evidence rather than a single new experiment, and its subject is task representation in general rather than emotion specifically. Applied to the present argument as a hypothesis rather than an established parallel, it suggests that emotional differentiation may involve a comparable trade-off: broad categories generalize widely across situations but supply coarse guidance, while narrowly segregated categories reduce ambiguity within a given situation at the cost of becoming less transferable to others. The goal identified in the preceding paragraph – sufficient rather than maximal differentiation – can be restated in these terms as a search for a workable balance among precision, generalization, flexibility, and mutual interference, rather than a race toward ever finer distinction for its own sake.

5.4 Misclassification and recursive reinforcement

An emotional category, once adopted, can influence the evidence subsequently encountered rather than remaining a passive description of evidence already gathered. Categorizing an ambiguous arousal as danger can increase vigilance, heighten bodily activation, sharpen threat detection, and prompt avoidance; those consequences can then be read as confirmation that danger was genuinely present, closing a loop that began with an interpretation rather than with an independently verified fact:

ambiguous sensation → threat category → defensive preparation → stronger sensation → apparent confirmation

This is a stronger claim than ordinary confirmation bias, in which a prior belief merely shapes which evidence gets noticed or how it gets weighted. Here, the prior category is implicated in producing some of the very evidence that will later appear to confirm it – the heightened arousal is not simply better attended to, it is partly caused by the categorization that will subsequently be read as tracking it. None of this implies that the resulting experience is unreal or somehow less genuinely felt. The physiological and phenomenological consequences of the loop are real even when the interpretation that initiated it was incomplete or mistaken.

What matters is that an interpretation has become causally involved in producing the state it appears merely to describe, rather than remaining a neutral report standing outside the process it characterizes.

It is worth being explicit about why this loop resists a common folk-psychological remedy. A widespread intuition treats emotional intensity as something like an accumulated quantity – pressure, charge, a reservoir – that builds until it is discharged, so that giving a feeling behavioral or verbal expression should relieve it in something like the way opening a valve relieves pressure in a sealed vessel. Nothing in the account developed here supports that picture, and the loop just described suggests why it fails. If a category is a self-sustaining pattern of the kind sketched above – one that reshapes the evidence available to it rather than merely awaiting discharge – then rehearsing the interpretation that sustains it, replaying the grievance, dwelling on the danger, restating the anger to oneself or to others, is more plausibly continued engagement with the same predictive category than release from it. Nothing needs to be borrowed from outside the argument already built to reach this conclusion: a category that organizes attention and preparedly shapes what counts as confirming evidence will tend to keep finding what it is already positioned to find, and expressing that category’s content is not obviously different, from the point of view of the loop, from any other event that re-engages it. Whether expression happens to interrupt the loop or extend it becomes, on this view, an empirical question about the particular case rather than something guaranteed by the act of expression itself.

5.5 Regulation is not the same operation as recognition

If re-engagement with a category can reinforce rather than resolve it, a further distinction becomes necessary between two operations that ordinary language tends to run together. *Expressing* a state – giving it behavioral or verbal form, describing it, acting it out – is not the same operation as *recategorizing* it: revising the interpretation that organizes the state in the first place. The first can leave the underlying category untouched, or, as the preceding section argued, can strengthen it through the same mechanism that produced it. The second requires something the first does not guarantee: new evidence, a shifted point of comparison, or a changed context salient enough to compete successfully with the interpretation already in place, sufficient to dislodge or reorganize it rather than simply restate it.

This distinction is offered here as a hypothesis about mechanism rather than as an established finding, and it is worth being precise about its evidential status. It follows from combining the account of prediction error given in Part III – where a discrepancy was shown to require more than mere occurrence in order to produce revision, since it must survive local processing, propagate, and remain available for later reconstruction – with the account of self-sustaining categorization just given in this section. No single cited study establishes that expression and recategorization are dissociable in the way proposed; the claim is essay-level synthesis, built from material already on the table, and it should be read that way. Its usefulness will be judged by what it does in Part VII, where the distinction is put to work in discussing regulation through recategorization, and by whether it generates a discriminable prediction rather than merely restating, in new vocabulary, the observation that some interventions help and others do not.

6 Consciousness and the Self

6.1 Feeling as regulated significance

Conscious feeling can now be characterized more precisely than the introduction allowed: it is bodily change experienced under an interpretation of its significance. This formulation is meant to resist a temptation that has recurred throughout the essay in different guises – the temptation to treat body and meaning as two separable stages, a raw physiological event first, an interpretive gloss added afterward. A conscious emotion, on the account developed here, is not a bodily perturbation accompanied by a verbal label attached once the perturbation is already complete. It is an organized state in which bodily condition, situational understanding, memory, action-readiness, and anticipated consequence become mutually constraining during the same interval, in the way Part III argued that sensing, prediction, and categorization already are.

Feeling, on this view, is what makes ongoing regulation experientially available to the organism undergoing it. This is a modest claim, and it is worth stating what it does not include. It does not claim that feeling reveals every mechanism producing the state – the account given across Parts II through V involves a great deal of processing, much of it distributed and none of it introspectively accessible, that never becomes part of what is felt. What conscious feeling does supply is a presentation of the organism’s condition as mattering in some particular way: as calling for withdrawal, for confrontation, for rest, for repair. The mattering is not decorative. It is the aspect of the state that is available to guide the organism’s own further action and communication, in a way the unfelt substrate, however causally important, cannot directly do on its own.

6.2 The self as a differentiated regulatory model

The self, on the account this essay has been building, develops in substantial part through the repeated differentiation of bodily and emotional experience described in Part V. Over time, a person learns which changes tend to be attributed to the body, to the immediate environment, to other people, to remembered events, to anticipated futures, or to standing dispositions that persist across many particular occasions. This accumulated attribution produces something usefully described as a model: a working sense of what tends to happen to oneself, what one tends to need, what one tends to do under given conditions, and what kind of situation a present circumstance is. The self, understood this way, is neither a fixed essence sitting behind experience nor a momentary fiction reconstructed from nothing on each occasion. It is a historically stabilized model, built out of exactly the kind of differentiation Part V described, that supports regulation across stretches of time longer than any single episode.

Emotional categories play a double role with respect to this model, and the ambivalence is worth stating plainly rather than resolving prematurely in either direction. They help maintain continuity: a person’s sense of what they tend to feel and why is part of what makes a life recognizable to itself across time. But the same mechanism can also imprison the continuity it maintains. Repeatedly interpreting a range of varied bodily and situational states through a single identity claim – “I am an anxious person,” “I am an angry person,” “I am broken” – can convert what began as a contingent, situation-specific categorization

into something treated as an assumed invariant, applied in advance of the evidence rather than derived from it on each occasion. The recursive-reinforcement structure described in Part V applies here with particular force: an identity claim of this kind is itself a category, and categories, as that section argued, can shape the evidence that will later appear to confirm them.

6.3 Differentiation without reification

This raises a central paradox for the account as a whole, one the essay has been circling rather than confronting directly until now. Consciousness, on the view developed here, requires distinctions – an undifferentiated affective state, as Part II observed, is precisely one that has not yet been organized precisely enough to guide targeted action. Yet distinctions, once formed, are exactly the kind of thing that can become falsely substantial: treated not as a currently useful way of organizing variable experience for a particular purpose, but as a discovery of some fixed kind that experience simply turns out to contain.

The resolution this essay proposes is not to abandon distinction in favor of some more fluid, category-free description of experience – nothing in the argument so far suggests such a description is available, and Part I’s critique of the fingerprint model was never a case against categorization as such, only against categorization understood as the discovery of essences. The resolution is instead that a useful emotional category must remain revisable in a way its user can recognize. It must organize experience without claiming to exhaust it. The category *grief*, to take a case the essay has not yet used, may capture something real and important about a condition without implying that every feature of that condition follows from a single, invariant grief mechanism waiting to be more precisely characterized. Differentiation succeeds, on this account, exactly when it increases an organism’s discriminative and regulatory capacity while preserving the recognition that the categories doing this work are models of variable underlying processes – useful, situation-relative, and answerable to revision – rather than fixed partitions the world or the body was already carrying.

7 Clinical and Social Consequences

7.1 Emotional knowledge and epistemic humility

People generally have privileged access to their own feelings and considerably less reliable access to the causes of those feelings, and the account developed across the preceding sections explains why both halves of that asymmetry should be expected. A person is well positioned to report the interpretation their own nervous system has currently settled on – the category presently organizing their attention and action-readiness – because that interpretation is, on this view, constitutive of what they are consciously experiencing. They are considerably less well positioned to report why that particular interpretation was reached, since the process described in Parts II through IV draws on interoceptive signal, learned regularity, present context, and anticipated allostatic need in ways that are not introspectively transparent even to the person undergoing them. First-person testimony about feeling should not be dismissed on the grounds that emotional experience turns out to be constructed rather than simply detected; construction, on the account given here, is the ordinary means by which

experience becomes intelligible in the first place, not a mark of unreliability.

At the same time, sincerity does not guarantee causal accuracy. A person may correctly report fear while misidentifying its source, correctly report exhaustion while interpreting it as a moral failing rather than a bodily state calling for rest, or correctly report anger while attributing it to the wrong provocation among several plausible candidates. The appropriate response to this gap is neither an external skepticism that discounts first-person report as unreliable window-dressing on processes better read off the body directly, nor an internal infallibilism that treats every report of felt cause as automatically correct because it is felt. Emotional knowledge, on the position this essay defends, is genuine – it tracks something real about the organism’s regulatory state – and corrigible, in the specific sense that its causal attributions can be mistaken and can be revised by evidence the person did not initially have access to.

7.2 Regulation through recategorization

The distinction drawn in Part V between expressing a state and recategorizing it now has direct clinical purchase. A racing heart, encountered without further context, can be interpreted as evidence of danger, of physical exertion, of excitement, of incipient illness, or of recent stimulant intake, and each interpretation licenses a different further action: vigilance and avoidance in the first case, rest in the second, engagement in the third, medical attention in the fourth, patience in the fifth. Recategorization, in the sense meant here, is what happens when new evidence, a shifted point of comparison, or a changed context makes one of these interpretations sufficiently more plausible than the one currently organizing the person’s attention and action – not when the person simply gives the current interpretation fuller expression.

This should not be overstated into a universal remedy, and the essay’s earlier caution against theoretical over-reach applies here as much as anywhere. Some conditions arise from material threats, disease processes, genuine deprivation, or structural circumstances that no change of interpretation resolves; a person correctly categorizing chronic pain as pain, or a dangerous relationship as dangerous, is not in need of recategorization, and treating persistent distress as invariably a categorization problem risks becoming a way of denying causes that are real and require their own remedy. What the argument built across Parts III and V does license is a narrower and more useful claim: where recategorization is genuinely available as a route to relief, it works through the mechanism already described – by introducing evidence, comparison, or context sufficient to compete with a standing interpretation – and not through the discharge of an accumulated quantity that expression is supposed to drain. Effective emotional regulation, on this view, may in different cases call for bodily care, environmental change, social repair, medical treatment, altered expectations, or genuine conceptual revision, and the theory should retain all of these possibilities rather than reducing regulation to any single one of them.

7.3 The politics of imposed categories

Categorization of the kind this essay has described also occurs at a level above the individual organism. Institutions classify people as threatening, unstable, irrational, resilient, uncooperative, or appropriately distressed, and these classifications are consequential in

ways that extend well beyond description: they affect whose testimony is believed, which conduct is permitted or sanctioned, and which interventions become available to a person so classified. The error identified in Part I's discussion of the fingerprint model recurs here in an institutional register. An observer – a clinician, an employer, a judge, a bystander – can mistake their own act of categorization for direct perception of an internal state the categorized person actually has, treating a facial expression, a tone of voice, or a pattern of conduct as though it transparently contained the emotional or characterological fact being read into it, rather than as evidence an inference has been drawn from, an inference shaped by the observer's own context, convention, and prior expectation in exactly the way Part I described.

The political question this raises is accordingly broader than the question of who is permitted to feel a given emotion in a given circumstance, though that question matters too. It is a question of authority: who has the standing to classify another person's affective or characterological state, on what evidential basis, and with what consequences attached to the classification once made. Nothing in the argument of this essay supplies an answer to that question on its own – it is a question about institutions and power as much as about cognition – but the essay's account of categorization as constructed, context-sensitive, and revisable rather than as the discovery of a fixed internal fact does at least remove one common justification for treating an imposed classification as beyond challenge: the assumption that the classifier was simply reading off something the classified person's body or conduct already, transparently, contained.

8 Relation to the Wider Framework

The preceding seven parts have deliberately stayed close to what predictive categorization, understood on Barrett and Miller's terms and constrained by the further evidence reviewed in Parts IV and V, can support. This part introduces a narrower, explicitly framework-level vocabulary for redescribing that material. Nothing in what follows should be read as a further empirical claim, or as evidence that the preceding sections' sources anticipated or independently confirm the vocabulary being applied to them. The question this part exists to raise is whether the vocabulary earns its keep by adding genuine discriminatory power to the account already built, or whether it merely renames distinctions the empirical sections stated more plainly on their own. Part IX takes up that question directly. This part's task is only to state the vocabulary carefully enough for the question to be answerable.

8.1 Record, recognize, admit, commit

Four stages, distinguished here for analytic purposes, can be used to organize the material assembled in Parts II through V without adding to it. A bodily signal may be *recorded* without being consciously felt: interoceptive information of the kind discussed in Part III can be present in the nervous system's processing without reaching anything that qualifies as awareness. A recurrent pattern may be *recognized* without receiving an emotional name: this is the condition Part II called affect prior to emotion, in which a person registers that something matters without yet possessing a category precise enough to say what. An interpretation may be *admitted* as relevant to the organism's active model of its situation

without being accepted as certain – corresponding to the prediction-error material in Part III, where a discrepancy can propagate far enough to influence further processing without thereby settling the question it raises. A category may become *committed* through action, speech, memory, or lasting regulatory change, in the sense given to categorization as a prospective, action-shaping policy in Part IV.

These four terms belong to this essay, not to Barrett and Miller, Bastos and colleagues, or any other source cited above. No cited study distinguishes recording from recognition, or admission from commitment, in these terms or in terms equivalent to them. Introducing the distinction is useful to the extent that it lets the essay refer economically to transitions the empirical sections already described – the movement from unfelt interoceptive signal, to undifferentiated affect, to a provisionally active interpretation, to a settled category with behavioral consequence – without requiring those transitions to be restated in full each time they recur. It is not useful, and would be actively misleading, if it were taken to imply that the neuroscience reviewed above independently discovered a four-stage architecture matching these terms.

8.2 Emotional categories and reachable futures

A second piece of framework vocabulary concerns what an active category does to the field of action available to an organism, extending the claim made in Part IV that categorization is prospective rather than merely descriptive. Fear changes the salience of escape, concealment, defensive preparation, and reassurance-seeking; grief changes the temporal weight given to memory, to attachment, to withdrawal, and to the slow work of reconstruction. In each case, the category does not simply describe a present condition – it alters which of the organism’s possible next actions currently register as available, as costly, or as relevant at all.

This can be written compactly, with the same caveat given earlier in the essay about notation of this kind:

$$C_t \mapsto \mathcal{R}_{t+1}(C_t),$$

where $\mathcal{R}_{t+1}(C_t)$ denotes the set of continuations rendered salient, apparently costed, or actionable once category C_t is active. The expression states a structural claim already implicit in Part IV’s account of categorization as a regulatory policy: that the effect of a category is not confined to how a present signal is described, but extends to which future actions currently present themselves as live options. It is worth being exact about what the expression does not claim. The category does not create the members of $\mathcal{R}_{t+1}(C_t)$ – escape, confrontation, withdrawal, and the rest remain physically available whether or not the organism is currently disposed to consider them. What the category changes is their accessibility, their apparent cost, and their regulatory relevance to the organism’s present situation. This is a redescription of material already established in Part IV, offered because the redescription may prove useful elsewhere in this framework’s broader corpus, not because the neuroscience reviewed in this essay discovered anything answering to $\mathcal{R}_{t+1}$ as an independent object.

8.3 Admissibility and its limits

The vocabulary of admission was introduced provisionally at the end of Part III, where it was already flagged as an analogy rather than a finding. It is worth returning to that analogy now, with the full apparatus of the essay in view, in order to state precisely where it holds and where it should be expected to break.

What the neuroscience reviewed in this essay actually shows is this: prior expectation, present sensory input, and task context jointly and measurably shape which neural signals propagate further through a cortical hierarchy, with that shaping differentiated by cortical layer, oscillatory frequency, and the specific stimulus involved (Bastos et al., 2020); and categorization, on Barrett and Miller’s synthesis of independent anatomical, electrophysiological, and imaging evidence, is plausibly distributed across this kind of signal processing rather than confined to a late, terminal stage (Barrett and Miller, 2026). Describing this pattern, at the level of the framework developed across this essay’s broader body of work, as a form of selective *admission* – some signals gaining further access to processing that others are denied – is a defensible redescription at that level of abstraction. It is not a claim this essay is entitled to attribute to the underlying studies, and it should be held only as long as the following four points of disanalogy remain visible.

First, neural suppression is not automatically refusal. Refusal, in the sense this framework’s broader vocabulary intends, implies something evaluative: a determination that a given signal fails to meet a standard, made on some basis that could in principle be examined. The alpha- and beta-band suppression of expected-input processing documented by Bastos and colleagues is a measurable resistance to propagation, with no evidence that it constitutes, or depends on, any evaluative act with content of that kind. Calling it refusal imports a normative structure the measurement does not itself contain.

Second, propagation is not commitment. A signal that survives inhibition and continues into further cortical processing has not thereby become behaviorally committed in the sense given to that term earlier in this part. Further competition, further gating, and further revision downstream all remain possible; propagation past one stage is not the same event as the durable regulatory or behavioral consequence that commitment, as this essay uses the word, requires.

Third, prediction error is not verification. A prediction error, as Part III described it, is a mismatch signal – a discrepancy between an active prediction and incoming evidence. Nothing about the existence of that discrepancy certifies which side of the mismatch was correct. The incoming signal could be noisy, spurious, or itself the product of an upstream error; the prediction could have been wrong; both are logically available readings of the same mismatch, and the mismatch signal alone does not adjudicate between them. Prediction error is properly understood as a trigger for possible revision, not as proof that revision was warranted – a distinction Part III’s discussion of what a discrepancy must additionally survive in order to produce learning already depended on, and one this section makes explicit.

Fourth, and most importantly, cortical feedback does not establish an ontological boundary between what exists and what is admitted. This is the disanalogy most likely to be lost if the framework vocabulary is applied carelessly. The anatomical and electrophysiological evidence reviewed in this essay shows differential propagation: some signal patterns advance further, faster, or with less resistance than others, depending on predictive context. Nothing

about differential propagation licenses the further, stronger claim that a suppressed signal has thereby been demoted to a different order of existence than a signal that propagates freely. A suppressed signal is a signal that has been measurably suppressed; it is not evidence of a metaphysical distinction between the admitted and the merely present, and the framework's vocabulary should not be read as asserting one.

What Part VIII adds to the essay, if it adds anything, is therefore a compact and explicitly labeled vocabulary for referring to distinctions the empirical sections above already established on their own terms – not a discovery those sections independently confirm, and not a license to treat gating, suppression, and propagation as though they were already the technical notions of refusal, admission, and commitment that this framework's broader corpus develops elsewhere. Whether this vocabulary earns its place by adding real explanatory discrimination to the account, or merely restates what Parts II through V already said more plainly, is the question Part IX now takes up directly.

9 Objections and Limits

9.1 Does construction imply arbitrariness?

The answer is no. A bridge is constructed, but it is constructed under constraints imposed by materials, gravity, terrain, and intended use, not assembled from whatever the builder happens to prefer. Emotional categories, on the account this essay has defended, are likewise constructed under bodily, environmental, developmental, and social constraint. Not every categorization regulates effectively: some interpretations fail to predict what actually follows, obscure a disease process that requires attention on its own terms, intensify distress rather than resolving it, or produce action poorly matched to the situation. The possibility of a category being wrong, in ways that carry real cost, is evidence of constraint operating on the process, not evidence that the process is arbitrary.

9.2 Are there no universal emotions?

The rejection of fixed emotional fingerprints in Part I does not require denying every cross-cultural or biological regularity in emotional life. Organisms share anatomical systems, comparable metabolic needs, overlapping defensive capacities, attachment dependencies extending across long developmental periods, and recurrent environmental problems – predation, scarcity, social exclusion, injury – that differ in detail but not in kind across many populations. The relevant question is what level of description these shared regularities actually support. Universal regulatory problems of this kind do not by themselves imply universal emotion modules with invariant physiological signatures and invariant expressive form; they are equally compatible with shared underlying constraints giving rise to overlapping but variable emotional organizations across individuals and cultures, which is the picture this essay has argued for throughout.

9.3 Two burdens of proof

The essay has now made two distinct kinds of claim, and they require two distinct kinds of defense. The first is the predictive-categorization account itself, developed in Parts II

through V from Barrett and Miller’s synthesis together with the further evidence on category learning, granularity, and neural representation reviewed there. The second is the additional, explicitly framework-level vocabulary introduced in Part VIII – recording, recognition, admission, and commitment; reachable futures; the qualified use of admissibility. These two layers do not stand or fall together, and treating them as though they did would undo the separation this essay has been careful to maintain since Part III.

The predictive-categorization account owes a specific debt: it must say what would count as evidence against it, not merely what counts as evidence for it. A theory becomes empty if every neural event can be redescribed as prediction and every outcome, whatever it turns out to be, can be redescribed as successful updating after the fact. Part III already supplied the material this debt is paid with, without naming it as such: a discrepancy between prediction and evidence was said to require more than mere occurrence in order to produce revision, since it must additionally survive local processing, propagate, engage attention, and remain available for later reconstruction, with the likelihood of each step depending on precision, salience, learning history, and the cost of revision. The account is falsifiable to exactly the extent that these conditions can be specified independently, in advance of a given case, rather than read backward from whether revision happened to occur. A concrete failure condition follows directly: if a category persists despite evidence that is sustained, unambiguous, and high in salience by every measure the account itself specifies, and no story is available for why revision failed under those conditions, the account has been disconfirmed in that instance rather than rescued by relabeling the persistence as further evidence of the category’s predictive strength.

The wider framework introduced in Part VIII owes a different debt, and passing the first test does not discharge it. It must show that its distinctions produce different descriptions, different predictions, different interventions, or different error diagnoses than the source vocabulary of prediction, feedback, and categorization already supplies on its own. If the framework vocabulary can only ever restate a result already available in that source vocabulary, using different words for the same contrast, it has added no discriminatory power, whatever independent interest its terms might hold.

9.4 When the framework adds discrimination, and when it does not

The clearest case in which the four-stage decomposition of Part VIII earns its place is one in which it blocks an inference the source vocabulary alone does not obviously block. Recognizing a bodily pattern – registering an elevated heart rate as salient enough to draw attention – does not entail that any specific interpretation of that pattern has been admitted as action-guiding; a person can be plainly aware that something is happening in their body while no particular account of what it means has yet gained enough support to organize their behavior. And admitting an interpretation, in turn, does not entail committing to it through action, speech, or lasting self-description; a person can provisionally treat a sensation as probably anxiety, in a way that shapes their attention for a moment, without that provisional treatment producing any visible behavior or becoming part of how they subsequently describe themselves. Folk accounts of emotion, and some clinical ones, frequently run these transitions together, treating clear awareness of a bodily state as though it already settled what the state means, or treating a momentary interpretation as though it necessarily issues in corresponding

action. The decomposition earns its keep exactly where it prevents this collapse: it identifies, as separately observable and separately failable, a transition that a single undifferentiated notion of “prediction” or “categorization” does not by itself distinguish.

Continuation geometry earns its place under a comparable condition. Two people – or the same person on different occasions – might both be described, using ordinary emotional vocabulary and using Barrett and Miller’s predictive-categorization account alike, as instances of the same category: both are, by their own report and by any behavioral measure available, afraid. Yet one person’s fear may render escape and concealment newly salient and readily affordable while confrontation appears costly or unavailable, shaped by a history in which flight has reliably worked; the other’s fear, shaped by a different history, a different present resource, or a different social position, may render confrontation the more salient and affordable option while escape appears unavailable or carries its own cost. Calling both cases “fear,” and explaining both through the same predictive-categorization mechanism, is not incorrect, but it does not by itself register this difference. Asking, in addition, what set of continuations the active category has rendered reachable, salient, and costed for this organism in this context supplies a distinction the bare categorization vocabulary passes over – not because categorization is wrong, but because it was never built to track the difference in question.

The failure condition is equally worth stating without softening. If, in practice, “admission” is defined only by whichever signal is later found to have propagated, and “reachable future” is defined only by whichever action the organism in fact took, then both terms have been fixed circularly by the very outcome they were introduced to explain, and the vocabulary has become retrospective relabeling rather than a source of independent discrimination. A framework term earns its use by licensing a distinction or a prediction available *before* the outcome is known – by specifying, for instance, in advance of observing a person’s behavior, that recognition without admission should be independently detectable, or that two people sharing a category label should show measurably different reachable-future sets given measurably different histories. Where the vocabulary cannot do this, and can only be applied after the fact to whatever happened, it should be recognized as decorative and set aside, however elegant it may look on the page. Passing the falsifiability test given to the predictive-categorization account in the preceding subsection does not automatically discharge this further burden; the two must be assessed independently, and this essay’s own use of the wider framework’s vocabulary in Part VIII should be held to the discrimination test stated here, case by case, rather than assumed to have passed it simply because the underlying neuroscience is sound.

9.5 Does allostasis explain consciousness?

No. Allostasis, on the account given in Part II, explains why anticipatory bodily regulation is necessary for an organism operating under uncertainty and limited resources. It does not, by itself, explain why some regulatory states are consciously felt while a great deal of the processing described across this essay – including much of what governs allostasis itself – is never felt at all. The essay offers a considerably more constrained claim than a solution to that further problem: emotional consciousness, whatever ultimately explains its presence, is substantially shaped by the differentiation of interoceptive and contextual states under

predictive regulation described in Parts II through V. This identifies conditions bearing on what gets felt and functions it may serve once felt, without pretending to have resolved why anything is felt at all.

10 Conclusion: Feeling as an Achievement of Distinction

The essay opened with an appearance that ordinary experience does nothing to dispel: emotions seem simply to be found inside us, discrete objects awaiting correct identification. The account developed since has argued that the organism does not first discover a complete emotion and then attach a word to it. It continuously regulates a body under uncertainty, and in order to do so it predicts upcoming conditions, interprets ambiguous sensory and interoceptive evidence, and groups variable situations into categories precise enough to support action. Some of these categorizations become conscious feelings; the essay's account of why has been offered as a substantive but constrained theoretical claim rather than a complete theory of consciousness. Emotional differentiation is the refinement of the underlying capacity – the acquisition of distinctions that let an organism recognize that superficially similar discomforts may have different causes, call for different actions, and open onto different futures.

Two formulations can close the essay, and they should not be read as making claims of the same kind or resting on the same evidence.

Emotion is bodily regulation made intelligible through predictive categorization.

This is the essay's substantive theoretical thesis, built across Parts II through V from Barrett and Miller's synthesis of predictive, context-sensitive categorization, disciplined by the further evidence on category learning, neural representation, and the costs of differentiation reviewed alongside it. It is offered as a claim that could, in the manner specified in Part IX, be wrong.

To differentiate a feeling is to distinguish the futures to which it gives access.

This second formulation is the continuation-geometric redescription introduced in Part VIII. Its value is conditional in a way the first formulation's is not: it earns its place only where it exposes a difference in accessible, affordable, or costed action that the first formulation, on its own, would pass over – and it should be abandoned in any application where it fails to do so, without that abandonment threatening the theoretical thesis it was built to redescribe.

The essay ends, accordingly, with a caution rather than a summary. Categories are necessary because nothing can regulate what remains entirely undifferentiated; an organism that could not distinguish threat from opportunity, or exhaustion from illness, would not thereby experience the world more purely, only less usefully. But every category also excludes possibilities it does not name, suppresses differences it treats as irrelevant, and prepares some continuations while relegating others to the background. Emotional intelligence, on the view this essay has defended, consists not in accumulating an ever-finer stock of categories, but in knowing when a given category clarifies experience, when it distorts it, and when – by evidence, by comparison, or by a changed and salient context – it must be revised.

References

- Barrett, Lisa Feldman and Miller, Earl K. Categorization is ‘baked’ into the brain. *Nature Reviews Neuroscience*, 27:435–456, 2026. doi: 10.1038/s41583-026-01036-2.
- Bastos, André M., Lundqvist, Mikael, Waite, Ayan S., Kopell, Nancy, and Miller, Earl K. Layer and rhythm specificity for predictive routing. *Proceedings of the National Academy of Sciences*, 117(49):31459–31469, 2020. doi: 10.1073/pnas.2014868117.
- Markov, Nikola T., Ercsey-Ravasz, Maria, Van Essen, David C., Knoblauch, Kenneth, Toroczkai, Zoltan, and Kennedy, Henry. Weight consistency specifies regularities of macaque cortical networks. *Cerebral Cortex*, 21:1254–1272, 2011.
- Minda, John Paul, Roark, Casey L., Kalra, Priya, and Cruz, Alexis. Single and multiple systems in categorization and category learning. *Nature Reviews Psychology*, 3:536–551, 2024. doi: 10.1038/s44159-024-00336-7.
- Posani, Lorenzo, Wang, Shanshan, Muscinelli, Samuel P., Paninski, Liam, and Fusi, Stefano. Rarely categorical, highly separable representations along the cortical hierarchy. *Nature*, 2026. doi: 10.1038/s41586-026-10668-4.
- Collina, Julia S., Erdil, Gonzalo, Xia, Maria, Angeloni, Christopher F., Wood, Karen C., Sheth, Jinal, Kording, Konrad P., Cohen, Yale E., and Geffen, Maria N. Individual-specific strategies inform category learning. *Scientific Reports*, 15:2984, 2025. doi: 10.1038/s41598-024-82219-8.
- Roark, Casey L. and Chandrasekaran, Bharath. Stable, flexible, common, and distinct behaviors support rule-based and information-integration category learning. *npj Science of Learning*, 8, 2023. Article 14, doi: 10.1038/s41539-023-00163-0.
- Greene, Michelle R. and Rohan, Aravind M. The brain prioritizes the basic level of object category abstraction. *Scientific Reports*, 15:31, 2025. doi: 10.1038/s41598-024-80546-4.
- Garner, Kelly G. and Dux, Paul E. Knowledge generalization and the costs of multitasking. *Nature Reviews Neuroscience*, 24:98–112, 2023. doi: 10.1038/s41583-022-00653-x.