

HOW TO DESIGN AN EMOTION · WORKING PAPER 1 OF 8

The Theory of Constrained Identity: Identity Conditions Upstream, Typicality Downstream

Cézjah (Cecil John)

ORCID 0009-0005-1866-9700

New Intellectual LLC, Miami

Correspondence: cezjah@cezjah.com

24 October 2026 · Version 2 (26 October 2026)

DOI: Crossref DOI pending

Also available as https://doi.org/10.5281/zenodo.23676655 (Zenodo)

License: CC BY 4.0 (this paper only; the book is All rights reserved)

newintellectual.com/pages/how-to-design-an-emotion-wp-01

ABSTRACT

When emotion science gave up the search for emotional essences, it also lost an answer to a prior question: what makes an episode of anger an episode of anger rather than a useful guess under a familiar name? I argue that several influential accounts, among them the theory of constructed emotion, relocate identity conditions downstream into similarity, prediction, categorization, and agreement, and that this relocation leaves error, disagreement, and revision without a standard. I propose the Theory of Constrained Identity (TCI) as the repair. It orders four layers with a direction of dependence: a constraint layer that fixes identity conditions and admissibility, an instance layer that realizes them under determinate values, a category layer that groups admitted instances by similarity and typicality, and a goal layer that evaluates both instances and constraint structures without supplying identity. Five invariants keep the layers apart. The arrangement dissolves the classical–prototype dichotomy by assigning each theory its level. I apply the TCI to the folk concept of anger in contemporary American English and use it to separate misapplication, disagreement, cultural variation, and revision from drift. What is new is a strict four-layer order, a separate goal layer, and a test that makes the order inspectable in executable code.

The Theory of Constrained Identity: Identity Conditions Upstream, Typicality Downstream

Cézjah
New Intellectual LLC, Miami
Open Researcher and Contributor ID (ORCID): 0009-0005-1866-9700

Working paper 1 of 8 in the series on the frameworks of *How to Design an Emotion*, October 2026
Digital object identifier (DOI): 10.5281/zenodo.23631050
License: CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)

Abstract

When emotion science gave up the search for emotional essences, it also lost an answer to a prior question: what makes an episode of anger an episode of anger rather than a useful guess under a familiar name? I argue that several influential accounts, among them the theory of constructed emotion, relocate identity conditions downstream into similarity, prediction, categorization, and agreement, and that this relocation leaves error, disagreement, and revision without a standard. I propose the Theory of Constrained Identity (TCI) as the repair. It orders four layers with a direction of dependence: a constraint layer that fixes identity conditions and admissibility, an instance layer that realizes them under determinate values, a category layer that groups admitted instances by similarity and typicality, and a goal layer that evaluates both instances and constraint structures without supplying identity. Five invariants keep the layers apart. The arrangement dissolves the classical–prototype dichotomy by assigning each theory its level. I apply the TCI to the folk concept of anger in contemporary American English and use it to separate misapplication, disagreement, cultural variation, and revision from drift. What is new is a strict four-layer order, a separate goal layer, and a test that makes the order inspectable in executable code.

Keywords: Theory of Constrained Identity; identity conditions; admissibility; constraint without universality; classical theory of concepts; prototype theory; typicality; categorization; goal-derived categories; theory of constructed emotion; anger; conceptual revision; rule-following

Relation to the book. This paper restates, in new prose, the argument of Part II, Chapters 2 to 6, and Appendices A to E of *How to Design an Emotion: The Deconstruction and Computation of Emotional Meaning from First Principles* (Cézjah 2026, chaps. 2–6, apps. A–E). A selection of the book's diagrams in the Unified Modeling Language (UML), excerpts of its C# programs, and their console input and output appear in the appendix to this paper. Every source cited here is cited in the book, with the same locators.

1. Introduction: When the Essence Dies

Researchers have looked for a region of the brain, a pattern of the body, or a face that belongs to anger alone, and the search has not produced one. Lisa Feldman Barrett reviews evidence that does not fit the view that there are kinds of emotion with boundaries fixed in nature (Barrett 2006, 28), and her theory of constructed emotion builds on that evidence (Barrett 2017). I accept the evidence. The picture of emotion as a hidden substance with its own fingerprint is gone, and I do not defend it.

The end of essence leaves a question that evidence of variation cannot answer by itself. Either something decides which episodes count as anger, or the word anger means whatever its users do with it. If nothing decides, then a person who calls disappointment anger has made no mistake; two observers who disagree about an outburst have no common subject; and a community that changes its use of the word has drifted rather than revised anything. The question is not empirical in the ordinary sense. It concerns what a concept must be for error, disagreement, and revision to be possible at all.

My answer is the Theory of Constrained Identity (Cézjah 2026, chap. 6). It keeps everything the constructionist critique gets right about variation, context, and culture, and it restores what that critique discards: identity conditions that sit upstream of every process that sorts, predicts, or evaluates. Section 2 sets out the vocabulary of words, concepts, constructs, and categories. Section 3 diagnoses how identity conditions have been orphaned. Section 4 dissolves the dichotomy between classical and prototype theories. Section 5 states the theory, and Section 6 applies it to anger. Section 7 describes how the book makes the architecture executable. Sections 8 to 11 state what is shared and what is new, answer objections, mark the scope, and conclude.

2. Words, Concepts, Constructs, and Categories

The argument depends on four terms that ordinary speech runs together (Cézjah 2026, chap. 2). A word is a label, a token that points. A concept states what anything must satisfy to belong to its kind; it is a unit of meaning, and in the book's terms, an idea is a unit of meaning that holds one or more claims that can be true or false. A construct is one real case that fills in a concept's open values. A category gathers valid cases without becoming their meaning. A recipe, the cakes baked from it, and the shelf that holds them stand in the same relations: no cake rewrites the recipe, and the shelf decides nothing about what a cake is.

The book draws its account of concept formation from Ayn Rand (1905–1982), as a philosopher rather than an empirical authority. Her account of concept formation by measurement omission states precisely the distinction the TCI needs: what varies across instances, and what must hold for each to count as the same kind. A three-foot oak table and a six-foot steel table are both tables, because the concept retains the relation that makes something a table and omits the particular

measurements. For Rand, “A definition is a statement that identifies the nature of the units subsumed under a concept” (Rand 1990, 40). On this view, a definition targets a concept, not a word, and its job is to fix what the concept includes and excludes, to hold steady across contexts, and to constrain reasoning rather than preference.

Once definitions hold, some valid propositions follow from meaning alone. In chess, bishops move diagonally, and no experiment can overturn that, because the rules decide it. Valid analytic propositions of this kind govern consistency within a conceptual system rather than existence. The book applies this demand to Basic Emotion Theory (Cézjah 2026, app. A). A theory that asserts that fear is a discrete, universal, evolved kind takes on an obligation to say what makes fear fear across cases: a face, a physiological pattern, a circuit, an action tendency, a feeling, an appraisal of threat, or a stated relation among them. Wide eyes, a racing heart, and freezing appear in fear, but the same signs do not single fear out. Until a theory says which condition governs, it has proposed inferences, not a definition.

3. The Orphaning of Identity Conditions

3.1 A Fair Hearing for Constructed Emotion

On Barrett’s account, the brain does not store emotion concepts as definitions. It assembles them in the moment, from experience, to predict incoming sensation and prepare action; a word such as anger ties together physically different episodes; and collective agreement sustains the grouping (Barrett 2017). She holds that emotion categories are conceptual because their functions “are imposed on physically disparate instances” through agreement (Barrett 2017, n. 15). The account explains variety, context, and culture better than the classical view of emotion, and I grant it all of that (Cézjah 2026, chap. 3).

What the account does not supply is an admissibility structure. A word can name a group, but it cannot admit members. Agreement reports what a community treats as anger; it cannot report when the community gets anger wrong. When concepts become patterns that survive because they predict well, success becomes the only judge, and success is a poor judge of genuineness: a counterfeit bill can buy lunch. Error shrinks to a failed forecast, disagreement to two different predictive histories, and persistence to repetition. A narrower version of the worry, about mistakes concerning one’s own emotions, appears in the philosophical literature (Andreotta 2024, §5). My concern is the general case.

3.2 Constraint Without Universality

The debate presents a false choice: either emotions carry essences that hold for every person, or they dissolve into constructions of the moment. I reject both and adopt what I call constraint without universality (Cézjah 2026, chap. 3). A concept states the conditions an instance must meet

within a system, for an agent, in a context. A British writer spells colour and an American spells color; the rule changes across the ocean, yet on each shore a misspelling remains a misspelling. Rules can vary without becoming arbitrary. I also separate identity conditions, the requirements a concept sets, from identity, the standing an episode earns when it meets them.

Construction and constraint do not compete. They work at different levels. The brain assembles each episode, and the constraint decides which episode it is. The same bricks can raise a chapel or a prison; the bricks do not decide which one stands.

3.3 Sorting Is Not Defining

The orphaning has a precise location: theorists mistake sorting for defining (Cézjah 2026, chap. 4). Michael I. Posner (b. 1936) and Steven W. Keele (1940–2005) trained subjects on distortions of dot patterns whose originals the subjects never saw; the subjects later classified the unseen originals well, and the authors report that "information about the schema is abstracted from the stored instances" (Posner and Keele 1968, 353). Barrett cites the paradigm in her account (Barrett 2017). I accept the result. The error begins when a fact about sorting becomes a theory of what a concept is. The subjects did not merely drift toward a centroid. They sorted each pattern into categories that the experimenter had built from different prototypes, and their sorting could be scored against those categories. Remove the definition, and the experiment cannot be scored. The book reconstructs the paradigm formally, with stimuli as vectors and classification by distance, to make explicit the constraint structure the experiment presupposes (Cézjah 2026, app. D).

Two further examples come from Barrett's public presentation of her account (Cézjah 2026, chap. 4). A rainbow is a smooth sweep of light without stripes, and the brain draws the lines. True; yet before anyone draws a boundary, a concept of color must say which dimensions count, such as wavelength, brightness, and saturation, and only then can samples be grouped. A grub prompts a call to the exterminator in Boston and becomes a meal in Iquitos, Peru. Culture changes the classification; it does not change the creature. Food and pest serve human goals, and neither defines the grub. Similarity explains clustering, not sameness.

4. The Faux Dichotomy

Concept theory often presents a forced choice: concepts are rule-governed definitions, or they are graded clusters of resemblance (Cézjah 2026, chap. 5). The classical view treats a concept as a list of features individually necessary and jointly sufficient for membership (Medin 1989, 1470; Armstrong, Gleitman, and Gleitman 1983, 267). Prototype theories answered with experiments in which membership came in degrees, and some members counted as more typical than others (Rosch and Mervis 1975, 573–75; Medin 1989, 1470–71).

Each camp sees something real, and each answers a different question. Classical theorists ask what makes something a member. Prototype theorists ask how people sort. Treat the two as rival answers to one question, and the result is a manufactured war. Each camp alone stumbles. A pure classical view struggles to say why a robin strikes people as a more typical bird than a penguin. A pure prototype view struggles to say why a whale remains a mammal for someone who, judging by appearance, would group it with fish (Medin 1989, 1470–77).

The resolution is an order of layers. Classical structure belongs to the concept as the constraint that fixes identity conditions. Prototype effects belong downstream, in how admitted instances distribute and how categories form over time. In the terms of Georg Wilhelm Friedrich Hegel (1770–1831), each theory is both preserved and overcome (Norris 2022, 73): the truth in each survives while its error does not. Classical theory is right that a concept has identity conditions and wrong to make them rigid rules for sorting. Prototype theory is right about typicality and graded membership and wrong to make them the concept.

Machine learning offers a concrete case (Cézjah 2026, chap. 5). A deep learning system learns from data and assigns each category a score, which resembles prototype theory in silicon. Yet before training begins, engineers fix the architecture, the objective function, and the space to search (LeCun, Bengio, and Hinton 2015, 436). The architecture constrains learning, but it does not supply the identity conditions of what the system learns. Those come from the task: the labels, the loss, and the purpose the engineers fixed. The outputs look prototype-like, and the task definition is what lets anyone call an output right or wrong.

5. The Theory of Constrained Identity

5.1 Four Layers

The TCI separates four functions that theory often conflates (Cézjah 2026, chap. 6).

The **constraint layer** fixes identity conditions and admissibility. A concept, on this account, is an intensional constraint structure: the set of structural requirements that governs permissible combinations of properties and relations. It is not a list of observed cases, a region of similarity, or a statistical summary. Identity conditions fix what makes a concept the same concept across time, agents, and contexts; they specify the permitted relations among values, not every value.

The **instance layer** realizes those conditions under determinate values. Differences in magnitude, configuration, embodiment, or context are different realizations of one concept, not different concepts. Bodily state and environmental input shape execution without supplying identity. This is why a concept is not identical to the set of its instances: instances can multiply, scatter, or vanish while the concept stands, because the actual never exhausts the possible.

The **category layer** groups admitted instances by similarity, typicality, context effects, and graded membership. It neither fixes identity nor determines admissibility. No instance is similar in the abstract; comparison requires a choice of respects, and "For similarity to be a useful construct, one must be able to specify the ways or respects in which two things are similar" (Medin, Goldstone, and Gentner 1993, 254).

The **goal layer** evaluates. A goal is a regulatory criterion applied to admitted instances: a standard of adequacy, efficiency, relevance, or compliance relative to a purpose. Goals can regroup the same instances in different ways without changing the concept, and an instance can satisfy a concept and still fail a governing goal. Goals also evaluate whether a constraint structure is a good one to have. They never decide what qualifies. The book tests this division against Barrett's goal-based examples, including the fish someone would choose as a pet rather than for dinner, and Lawrence W. Barsalou's (b. 1951) ad hoc categories formed in the service of goals (Barsalou 1983, 211), and it separates five layers those examples compress: identity conditions, instantiation, goal selection, runtime construction, and classification under several concepts (Cézjah 2026, app. C).

5.2 Five Invariants

Five invariants keep the layers distinct (Cézjah 2026, chap. 6). Identity conditions remain stable unless the constraint structure changes. Every admitted instance satisfies those conditions before categorization begins. Categorization organizes admitted instances without redefining the concept. Goals regulate categorized instances and evaluate constraint structures, but they do not determine admissibility. Learning and optimization refine execution without altering the identity conditions they presuppose. From these follows a restriction: no downstream success condition alone constitutes a concept. Statistical convergence, predictive accuracy, cultural uptake, and practical utility each evaluate execution under constraint, and each presupposes the identity conditions it evaluates.

5.3 Context Selects

The layers also explain how one word can serve several concepts. If right indexes a moral claim in one context and a legal entitlement in another, the shift is no evidence that the concept is fuzzy or assembled on the spot. The same linguistic pointer can reach different stable architectures, each with its own identity conditions (Cézjah 2026, chap. 6). Context selects; it does not define. Prediction works inside the rules; it does not write them. Learning sharpens skill; it does not rewrite identity.

6. A Worked Case: Anger

Abstraction can hide weakness, so the book applies the TCI to a single concept in a single system: the folk-psychological concept of anger as adult speakers of contemporary American English use it in ordinary interpersonal life (Cézjah 2026, chap. 6). It is not anger in every culture, and it is not a scientific kind. Fixing the system first is what constraint without universality requires.

Constraint. Within this system, an episode is admissible as anger only if it meets four conditions. The agent construes some event as a wrong, slight, or obstruction attributable to an agent. The agent holds that agent accountable: not merely "this is bad" but "you did this, and you should not have." The episode is realized in core affect whose valence is negative, at high arousal or low. And the episode includes a readiness to confront, correct, retaliate, or restore, even if the agent suppresses it. Each condition tracks a feature that appraisal research and studies of the folk concept associate with anger: an offense, an accountable agent, unpleasant affect, and a readiness to respond (Smith and Lazarus 1993, 238; Frijda, Kuipers, and ter Schure 1989, 212; Russell 2003, 145; Shaver et al. 1987, 1061). In appraisal terms, "the core relational theme for anger is 'other-blame'" (Smith and Lazarus 1993, 238). The book offers the four conditions as a revisable hypothesis about the English concept, not as a proof.

Instantiation. Road rage realizes the conditions with high arousal and an urge to shout. Cold resentment at a colleague realizes them with low arousal and a plan to withhold cooperation. A written complaint to a landlord who ignores a broken heater realizes them through a demand for repair. The episodes differ in physiology, behavior, and feeling, and each satisfies one constraint structure. Variation lives in the values; identity lives in the constraints.

Categorization. Only admitted instances are grouped. A shouted confrontation over an obvious insult is a highly typical instance in this system; quiet resentment is less typical. The rating is real and measurable, and research on anger reports exactly such graded judgments, concluding that anger's "membership is graded, its borders are fuzzy" (Russell and Fehr 1994, 186). On the TCI, that gradient is a fact about the distribution of admitted instances. Typicality grades what the constraint has admitted; it admits nothing of its own.

Goals. A therapist groups anger episodes by intensity and triggers, a lawyer by provocation and control, a parent by fitness for the moment. Each goal produces a different grouping, and none changes what counts as anger.

Misapplication. A man says he is angry at the weather for ruining his picnic. If he truly holds no agent accountable, the episode fails the second condition; he is disappointed, and his self-report misapplies the concept. If he half-blames a careless forecaster, the episode qualifies, though it is atypical. The constraint, not the word he chose, decides the case. Barrett's account explains why he reached for the word. The TCI explains whether the word was correct.

Disagreement. Two observers watch a woman speak sharply to her manager. One calls it anger; the other calls it assertiveness. On a purely predictive account, this is a difference between two predictive histories. On the TCI, it is a genuine disagreement with a determinate subject: did she construe the manager as accountable for a wrong? The observers can lack the evidence to settle it, but they disagree about something, and the constraint structure says what.

Cultural variation. In a community whose concept of anger requires a public violation of status, or merges what English separates, the admissibility conditions differ. Some episodes that English speakers call anger will not qualify there, and the reverse. Each system binds its own instances; no system binds all of them. Translation becomes an empirical question about overlapping constraint structures, not a search for one true anger.

Revision. Suppose English speakers come to accept anger at impersonal systems, such as the bureaucracy, with no accountable agent, and correct one another accordingly. They have revised the second condition, and the concept has changed. Revision differs from drift in three ways: it changes what counts as correct rather than how often a word appears; it enters correction, so the community corrects speakers who apply the old condition; and the community endorses it rather than merely undergoing it. Research on concept creep reports the expansion of harm-related concepts and separates changes in frequency from changes in meaning (Haslam et al. 2020, 254–59). A prediction error can motivate a revision, but an error that changes one agent's judgments without changing what the community corrects is drift.

7. Making the Order Inspectable

The book states the TCI in plain argument and then disciplines that argument with formal tools (Cézjah 2026, chaps. 2, 4–6, apps. B–E). Diagrams in UML and programs in C# separate type from instance, structure from realization, and identity from contextual variation; one program rejects any color sample that breaks the rules, and another admits a grub before any foraging or pest-control goal judges it. The book models structure, not neural mechanism, and it says so: diagrams and code check consistency, not truth.

The plainest case of the layering uses JavaScript, a language whose objects inherit from prototypes (Cézjah 2026, app. E). The word misleads: in JavaScript, a prototype is what an object inherits from, not an average of similar objects. A programmer writes the prototype first, and objects inherit from it and vary. Expanding an object's prototype chain in a browser's developer tools displays three levels: an instance with its particular values, a concept-level prototype that encodes the admissibility rule, and a runtime infrastructure that terminates the chain. The appendix then distinguishes three kinds of change: misinstantiation at the instance level, parametric revision of the concept-level object, and structural revision of its rule. On this reading, prototype theory collapses the concept level into the instance level, and classical essentialism treats the revisable concept level as if it had the fixed necessity of the infrastructure. I claim an architecture, not a new

metaphysics. The constraint layer fixes identity conditions for reuse, correction, and reference within a system; it does not claim that those constraints are metaphysically necessary features of reality.

8. What Is Shared and What Is New

Others joined classical and prototype theory before me, and I owe them the idea (Cézjah 2026, chap. 5). Edward E. Smith (1940–2012) and his colleagues modeled categorization with defining and characteristic features (Smith, Shoben, and Rips 1974). Daniel N. Osherson (1949–2022) and Smith (Osherson and Smith 1981), and Lila R. Gleitman (1929–2021) and Henry Gleitman (1925–2015) with their coauthor (Armstrong, Gleitman, and Gleitman 1983, 292–95), proposed a core that fixes membership and an identification procedure that drives quick judgments; the latter describe this dual theory as having "considerable currency among cognitive psychologists today" (Armstrong, Gleitman, and Gleitman 1983, 295). Others widened the field with exemplar, goal-based, and theory-based accounts, and hybrid models of category learning combine rules with exemplars (Medin and Schaffer 1978; Barsalou 1983, 211; Murphy and Medin 1985, 289; Erickson and Kruschke 1998, 107).

I differ in three ways. These are dual models with two parts; mine has four layers, and goals have a layer of their own. They describe how people represent concepts; I specify an architecture with a direction of dependence, in which nothing downstream defines what is upstream. And I state the order as type and instance in executable code, so that a collapse appears as an error. The book puts the difference in one pair (Cézjah 2026, chap. 5):

What is shared: concepts have a core, and typicality is real. What is new: a strict four-layer order, a separate goal layer, and a test that makes the order inspectable.

Chapter 6 states the claim of novelty in its strongest form (Cézjah 2026, chap. 6). Earlier accounts set the two parts side by side, as components of one representation or as systems competing to answer the same question. I know of no prior account that orders them as layers with a direction of dependence; gives goals a layer of their own that evaluates both instances and constraint structures without supplying identity; treats revision as a change in the constraint layer rather than as drift; and requires the layers to remain distinct when the architecture runs.

Two bodies of work come closer still, and I mark what I do not claim. Appraisal theories treat appraisal as both a component of an emotional episode and its elicitor (Moors et al. 2013, 120). Richard S. Lazarus (1922–2002) analyzes emotions by the core relational themes their appraisals produce (Lazarus 1991, 819); Robert C. Solomon (1942–2007) argues that emotions are evaluative judgments (Solomon 1988); and others treat emotions as evaluative attitudes whose correctness turns on properties such as danger for fear (Deonna and Teroni 2015, 303). My anger conditions owe much to this tradition, and I claim no new content for them. And the rule-following literature,

from Saul A. Kripke's (1940–2022) reading of Wittgenstein onward, asks what fixes the content of a rule from a finite record of past use (Kripke 1982, 21; Boghossian 1989, 508). I do not solve that problem.

9. Objections and Replies

Objection 1: If the community fixes the constraints, agreement decides correctness after all. This is the strongest objection to the TCI, because it turns the critique of Barrett back on the account (Cézjah 2026, chap. 6). Part of it is right: communities institute concepts. But instituting a standard differs from applying it, in the sense of the distinction between justifying a practice and justifying a particular action that falls under it (Rawls 1955, 3). Chess illustrates the point. Players instituted the rules; once the rules stand, a move is legal or illegal whatever the onlookers believe, and a whole club can call a position checkmate and be wrong. As the same analysis puts it, "the rules of practices are logically prior to particular cases" (Rawls 1955, 25). Anger works the same way. My community can have instituted the requirement of an accountable agent; it does not follow that every episode the community calls anger is anger. A community can be wrong in bulk, and its own constraint convicts it. Barrett's account does not lack norms. What it lacks is an admissibility structure for agreement to fix, so that the word and the agreement do both jobs at once. The TCI lets agreement institute a constraint and then lets the constraint decide each case.

Two further checks constrain the constraints. A concept serves a function: anger registers a wrong by an accountable agent and readies a response, and a structure that fired at random would fail that function by the community's own goals. Function does not select one structure, but it rules many out. Ruth Garrett Millikan (b. 1933) offers a related appeal to function, but for her "Proper functions are determined by the histories of the items possessing them" (Millikan 1989, 284), whereas I measure function against the community's own goals. And the facts push back: whether a wrong occurred, and whether an agent construed it as one, are not up to the observers.

Objection 2: Revisable constraints are indeterminate constraints. They are not. At any given time, a concept has a constraint structure, and applications made then answer to it. A later revision does not make earlier errors correct; it creates a new structure. The rule-following literature puts the underlying point in general form: "Having a meaning is essentially a matter of possessing a correctness condition" (Boghossian 1989, 515). I claim only that content and application belong on separate layers.

Objection 3: The anger conditions are just appraisal theory. The resemblance is deliberate, but the claim differs (Cézjah 2026, chap. 6). Appraisal theories describe a process: how an agent evaluates an encounter and how that evaluation produces an emotion (Moors et al. 2013, 119). The constraint layer states what an episode must satisfy to count as anger within a system, however it was produced. An appraisal theorist explains how anger arises; the constraint layer says when an

episode is anger. The book draws the line between appraisal and interpretation in a separate appendix, which working paper 5 restates.

Objection 4: Deep learning vindicates prototype theory. It vindicates prototype effects, as patterns in instance distributions under fixed constraints. It does not vindicate prototype theory as a theory of concepts, because the task definition, fixed upstream, is what makes any output right or wrong (Section 4).

10. Scope and Limits

The TCI is an architectural clarification, not a metaphysics of concepts and not a psychological model of how brains store them. Its claims concern what a concept must be for error, disagreement, and revision to be intelligible, and how the functions that operate on concepts must be ordered. The worked case covers one concept in one system, and its four conditions remain a hypothesis about the English folk concept, answerable to community correction and to the functional test above. The TCI does not explain how a concept comes to stand for its referent or how interpretation binds identity to reference; the book takes up that question in its Constructivist Semiotics, which working paper 2 restates (Cézjah 2026, chap. 7). Nor does it explain emotion as such. Emotion enters the book's architecture later, as a goal-governed interpretive act that the TCI makes possible but does not by itself describe (Cézjah 2026, chap. 9).

11. Conclusion

Rejecting essences was right. Rejecting identity conditions along with them was not. I argued that several influential accounts orphan identity conditions by handing their work to similarity, prediction, categorization, and agreement, and that the result leaves error, disagreement, and revision without a standard. The Theory of Constrained Identity restores the standard by ordering four layers, constraint, instance, category, and goal, with dependence running one way, and by holding five invariants that keep the layers apart when the architecture runs. Applied to anger, it separates misapplication from difference, disagreement from divergent histories, and revision from drift.

The practical consequence is modest and personal. When a person calls herself angry, she can ask whom she holds accountable, and for what wrong. If the answer is no one, she faces something else: grief, worry, or disappointment. A concept she could never misapply would be a concept she could never learn from. The full argument, with its diagrams and programs, appears in Part II of *How to Design an Emotion* (Cézjah 2026, chaps. 2–6).

Appendix: Code, UML Diagrams, and Input/Output

The figures below are the book's own diagrams in the Unified Modeling Language (UML), reproduced unchanged. The listings are excerpts from the book's C# companion projects, and the input/output blocks were captured by running those projects with the .NET 8 software development kit; I keep only the lines that bear on the argument. The full programs and the complete transcripts sit beside this paper in `code-samples/`. As in the book, the code checks the consistency of the architecture, not the truth of the theory. The diagrams and programs come from Chapter 2 and Appendices C and E (Cézjah 2026, chap. 2, apps. C, E).

A.1 Concept and Construct

The smallest case separates a concept, which declares identity conditions, from a construct, which realizes them with determinate values (Section 2).

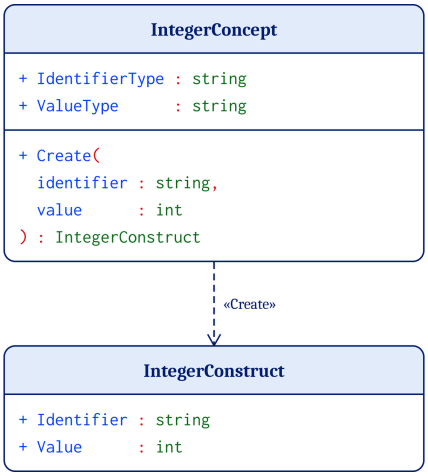


Figure 1. UML class diagram: IntegerConcept declares the identity conditions and creates an IntegerConstruct (book Figure 2-1).

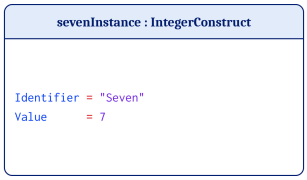


Figure 2. UML object diagram: the construct sevenInstance realizes those conditions (book Figure 2-2).

Listing 1. Concept and construct as separate types (`code/chapter-02/project-1/Program.cs`**, lines 4–25)**

```
class IntegerConcept
{
    public string IdentifierType => "string";
    public string ValueType => "int";

    public IntegerConstruct Create(string identifier, int value)
    {
        return new IntegerConstruct(identifier, value);
    }
}

class IntegerConstruct
{
    public IntegerConstruct(string identifier, int value)
    {
        Identifier = identifier;
        Value = value;
    }

    public string Identifier { get; }
    public int Value { get; }
}
```

Listing 2. Input/output: the concept stays fixed; only the assigned value realizes the construct (`code/chapter-02/project-1`**; captured with .NET 8)**

```
$ dotnet run -- Seven 7
Integer Concept
identifier type: string
value type: int

Construct Created
identifier: "Seven"
int value: 7

Result
The identifier names the construct; the concept declares the integer identity
conditions.
Identity conditions remain fixed. Only the assigned value realizes the construct.
```

A.2 The Constraint Layer: Admissibility Before Instantiation

The constraint layer admits or rejects a candidate before any goal or context appears (Section 5.1). The Fish constructor enforces two simplified admissibility conditions; a candidate that fails either never becomes an instance.

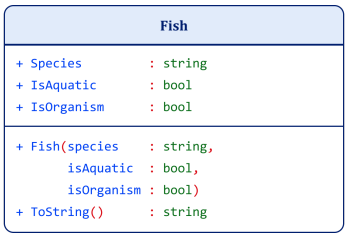


Figure 3. UML class diagram of the Fish concept, whose constructor enforces the admissibility conditions (book Figure AC-1).

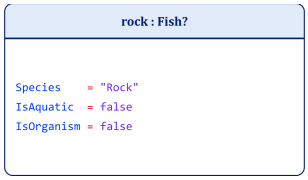


Figure 4. UML object diagram of the candidate rock, which the Fish constructor rejects (book Figure AC-4).

Listing 3. Identity conditions as admissibility constraints (`code/appendix-c/project-1/Program.cs`**, lines 6–31)**

```
class Fish
{
    public string Species { get; }
    public bool IsAquatic { get; }
    public bool IsOrganism { get; }

    // Simplified admissibility conditions, for illustration: a fish must be
    // aquatic, and it must be an organism. They do not vary with context.
    public Fish(string species, bool isAquatic, bool isOrganism)
    {
        var failed = new List<string>();
        if (!isAquatic)
            failed.Add("not aquatic");
        if (!isOrganism)
            failed.Add("not an organism");

        if (failed.Count > 0)
            throw new ArgumentException($"{species} is not a Fish ({string.Join(", ",
        failed)})");

        Species = species;
        IsAquatic = isAquatic;
        IsOrganism = isOrganism;
    }

    public override string ToString() => Species;
}
```

Listing 4. Input/output: an admitted candidate and a rejected one (`code/appendix-c/project-1`; captured with .NET 8)

```
$ dotnet run -- goldfish
Valid Fish created: Goldfish

$ dotnet run -- rock
Rejected: Rock is not a Fish (not aquatic, not an organism).
```

A.3 The Goal Layer: Selection Without Redefinition

The goal layer selects among admitted instances and assigns them roles, without touching identity conditions or instantiated values (Section 5.3). The same context yields different selections under different goals.

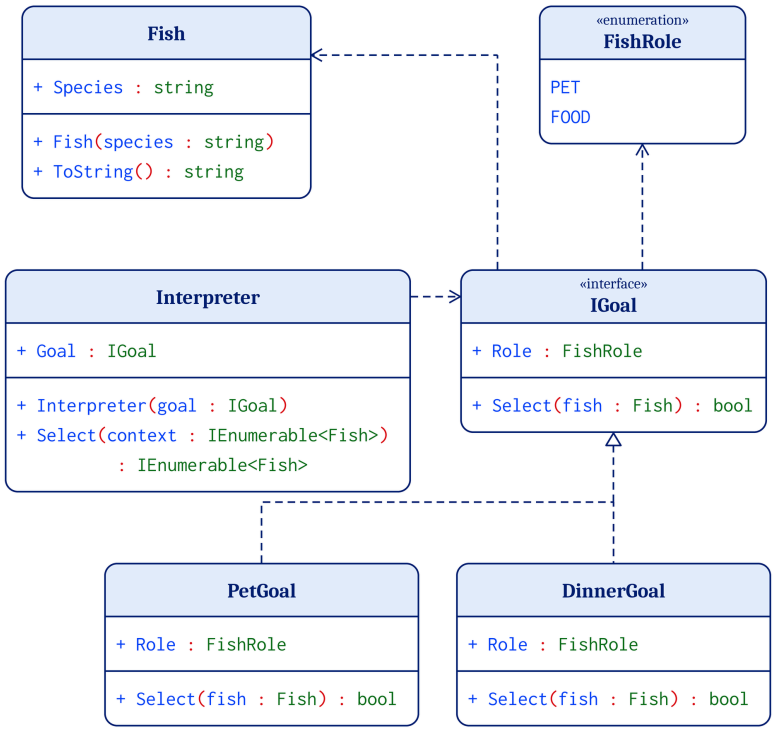


Figure 5. UML class diagram of goal-governed selection: PetGoal and DinnerGoal realize IGoal, and the Interpreter selects under the active goal (book Figure AC-8).

Listing 5. Goals select and assign roles; the fish stays the same instance (code/appendix-c/project-3/Program.cs, lines 25–59)

```
interface IGoal
{
    FishRole Role { get; }
    bool Select(Fish fish);
}

// Stipulation, for illustration, following the example this appendix opens with:
// when the purpose is buying a pet, a goldfish serves, and a salmon does not. The
      role,
// PET or FOOD, belongs to the goal, not to the fish: the fish stays the same instance
// under every goal.
class PetGoal : IGoal
{
    public FishRole Role => FishRole.PET;
    public bool Select(Fish fish) => fish.Species == "Goldfish";
}

class DinnerGoal : IGoal
{
    public FishRole Role => FishRole.FOOD;
    public bool Select(Fish fish) => fish.Species == "Salmon";
}

// The Interpreter selects within a Context; which fish are relevant, and in what role,
// is a function of its Goal.
class Interpreter
{
    public IGoal Goal { get; }

    public Interpreter(IGoal goal)
    {
        Goal = goal;
    }

    public IEnumerable<Fish> Select(IEnumerable<Fish> context) =>
        context.Where(Goal.Select);
}
```

Listing 6. Input/output: one context, two goals, two selections (code/appendix-c/project-3; captured with .NET 8)

```
$ dotnet run -- dinner
Active Goal: DinnerGoal
Context: Goldfish, Salmon

Selected: Salmon (FOOD)

$ dotnet run -- pet
Active Goal: PetGoal
Context: Goldfish, Salmon

Selected: Goldfish (PET)
```

A.4 The Prototype Chain in JavaScript

Appendix E makes the layering visible in a browser (Section 7). The concept-level object encodes the admissibility rule; an instance inherits from it and supplies its own values; the developer tools display the chain. Both panels are the book's figures.

Console Input </> JavaScript

```
// ═══════════════════════════════════════
// Layer 1 : Concept-Level Constraint
// ═══════════════════════════════════════

const Concept = {
  identity : "Triangle",     // Parametric universal
  sidesRequired : 3,         // Parametric universal
  valid() {                  // Structural universal (evaluative rule)
    return this.sides === this.sidesRequired;
  }
};
```

Figure 6. JavaScript input: the concept-level object with identity, sidesRequired, and the rule valid().

Chrome Console </> JavaScript

```
▼ {type: 'equilateral', sides: 3}
    sides: 3
    type: "equilateral"
  ▼ [[Prototype]]: Object
      identity: "Triangle"
      sidesRequired: 3
    ▶ valid: ƒ valid()
    ▶ [[Prototype]]: Object
```

Figure 7. Chrome console output: the equilateral instance, with the concept-level prototype above it.

References

Andreotta, Adam J. 2024. "Partial First-Person Authority: How We Know Our Own Emotions." *Review of Philosophy and Psychology* 15 (4): 1375–97. https://doi.org/10.1007/s13164-023-00698-6.

Armstrong, Sharon Lee, Lila R. Gleitman, and Henry Gleitman. 1983. "What Some Concepts Might Not Be." *Cognition* 13 (3): 263–308. https://doi.org/10.1016/0010-0277(83)90012-4.

Barrett, Lisa Feldman. 2006. "Are Emotions Natural Kinds?" *Perspectives on Psychological Science* 1 (1): 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x.

Barrett, Lisa Feldman. 2017. "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization." *Social Cognitive and Affective Neuroscience* 12 (1): 1–

23. https://doi.org/10.1093/scan/nsw154.

Barsalou, Lawrence W. 1983. “Ad Hoc Categories.” *Memory & Cognition* 11 (3): 211–27. https://doi.org/10.3758/BF03196968.

Boghossian, Paul A. 1989. “The Rule-Following Considerations.” *Mind* 98 (392): 507–49. https://doi.org/10.1093/mind/XCVIII.392.507.

Cézjah. 2026. *How to Design an Emotion: The Deconstruction and Computation of Emotional Meaning from First Principles*. Universal Moral Concepts III. Miami: New Intellectual LLC. International Standard Book Number (ISBN) 979-8-9898135-9-9 (paperback), 979-8-9898135-3-7 (Kindle). https://doi.org/10.5281/zenodo.23630995.

Deonna, Julien A., and Fabrice Teroni. 2015. “Emotions as Attitudes.” *Dialectica* 69 (3): 293–311. https://doi.org/10.1111/1746-8361.12116.

Erickson, Michael A., and John K. Kruschke. 1998. “Rules and Exemplars in Category Learning.” *Journal of Experimental Psychology: General* 127 (2): 107–40. https://doi.org/10.1037/0096-3445.127.2.107.

Frijda, Nico H., Peter Kuipers, and Elisabeth ter Schure. 1989. “Relations among Emotion, Appraisal, and Emotional Action Readiness.” *Journal of Personality and Social Psychology* 57 (2): 212–28. https://doi.org/10.1037/0022-3514.57.2.212.

Haslam, Nick, Brodie C. Dakin, Fabian Fabiano, Melanie J. McGrath, Joshua Rhee, Ekaterina Vylomova, Morgan Weaving, and Melissa A. Wheeler. 2020. “Harm Inflation: Making Sense of Concept Creep.” *European Review of Social Psychology* 31 (1): 254–86. https://doi.org/10.1080/10463283.2020.1796080.

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.

Lazarus, Richard S. 1991. “Progress on a Cognitive-Motivational-Relational Theory of Emotion.” *American Psychologist* 46 (8): 819–34. https://doi.org/10.1037/0003-066X.46.8.819.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” *Nature* 521 (7553): 436–44. https://doi.org/10.1038/nature14539.

Medin, Douglas L. 1989. “Concepts and Conceptual Structure.” *American Psychologist* 44 (12): 1469–81. https://doi.org/10.1037/0003-066X.44.12.1469.

Medin, Douglas L., and Marguerite M. Schaffer. 1978. “Context Theory of Classification Learning.” *Psychological Review* 85 (3): 207–38. https://doi.org/10.1037/0033-295X.85.3.207.

Medin, Douglas L., Robert L. Goldstone, and Dedre Gentner. 1993. “Respects for Similarity.” *Psychological Review* 100 (2): 254–78. https://doi.org/10.1037/0033-295X.100.2.254.

Millikan, Ruth Garrett. 1989. “Biosemantics.” *Journal of Philosophy* 86 (6): 281–97. https://doi.org/10.2307/2027123.

Moors, Agnes, Phoebe C. Ellsworth, Klaus R. Scherer, and Nico H. Frijda. 2013. “Appraisal Theories of Emotion: State of the Art and Future Development.” *Emotion Review* 5 (2): 119–24. https://doi.org/10.1177/1754073912468165.

Murphy, Gregory L., and Douglas L. Medin. 1985. “The Role of Theories in Conceptual Coherence.” *Psychological Review* 92 (3): 289–316. https://doi.org/10.1037/0033-295X.92.3.289.

Norris, Andrew. 2022. “Hegel and Cavell on Meaning and Sublation.” *Conversations: The Journal of Cavellian Studies*, no. 9: 72–84. https://doi.org/10.18192/cjcs.vi9.6247.

Osherson, Daniel N., and Edward E. Smith. 1981. “On the Adequacy of Prototype Theory as a Theory of Concepts.” *Cognition* 9 (1): 35–58. https://doi.org/10.1016/0010-0277(81)90013-5.

Posner, Michael I., and Steven W. Keele. 1968. “On the Genesis of Abstract Ideas.” *Journal of Experimental Psychology* 77 (3): 353–63. https://doi.org/10.1037/h0025953.

Rand, Ayn. 1990. *Introduction to Objectivist Epistemology*. 2nd ed. Edited by Harry Binswanger and Leonard Peikoff. New York: Meridian.

Rawls, John. 1955. “Two Concepts of Rules.” *Philosophical Review* 64 (1): 3–32. https://doi.org/10.2307/2182230.

Rosch, Eleanor, and Carolyn B. Mervis. 1975. “Family Resemblances: Studies in the Internal Structure of Categories.” *Cognitive Psychology* 7 (4): 573–605. https://doi.org/10.1016/0010-0285(75)90024-9.

Russell, James A. 2003. “Core Affect and the Psychological Construction of Emotion.” *Psychological Review* 110 (1): 145–72. https://doi.org/10.1037/0033-295X.110.1.145.

Russell, James A., and Beverly Fehr. 1994. “Fuzzy Concepts in a Fuzzy Hierarchy: Varieties of Anger.” *Journal of Personality and Social Psychology* 67 (2): 186–205. https://doi.org/10.1037/0022-3514.67.2.186.

Shaver, Phillip, Judith Schwartz, Donald Kirson, and Cary O’Connor. 1987. “Emotion Knowledge: Further Exploration of a Prototype Approach.” *Journal of Personality and Social Psychology* 52 (6): 1061–86. https://doi.org/10.1037/0022-3514.52.6.1061.

Smith, Craig A., and Richard S. Lazarus. 1993. “Appraisal Components, Core Relational Themes, and the Emotions.” *Cognition and Emotion* 7 (3–4): 233–69. https://doi.org/10.1080/02699939308409189.

Smith, Edward E., Edward J. Shoben, and Lance J. Rips. 1974. “Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions.” *Psychological Review* 81 (3): 214–41.

https://doi.org/10.1037/h0036351.

Solomon, Robert C. 1988. “On Emotions as Judgments.” *American Philosophical Quarterly* 25 (2): 183–91.

How to cite

The Theory of Constrained Identity: Identity Conditions Upstream, Typicality Downstream

APA

Cézjah. (2026). *The Theory of Constrained Identity: Identity conditions upstream, typicality downstream* (Working paper 1). New Intellectual LLC. https://newintellectual.com/pages/how-to-design-an-emotion-wp-01 (Crossref DOI pending)

CONTACT THE AUTHOR

Cézjah (Cecil John)
cezjah@cezjah.com
ORCID 0009-0005-1866-9700
newintellectual.com

ALSO AVAILABLE AS

https://doi.org/10.5281/zenodo.23676655 (Zenodo, version 2; all versions: https://doi.org/10.5281/zenodo.23631049)

LINKS

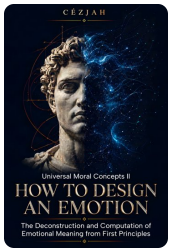
Landing page: newintellectual.com/pages/how-to-design-an-emotion-wp-01

Product page: newintellectual.com/products/working-paper-01-constrained-identity

License: CC BY 4.0 (this paper only; the book is All rights reserved)

KEYWORDS

Theory of Constrained Identity · identity conditions · admissibility · constraint without universality · classical theory of concepts · prototype theory · typicality · categorization · goal-derived categories · theory of constructed emotion · anger · conceptual revision · rule-following



This paper restates, in new prose, the argument of Part II of *How to Design an Emotion* (Cézjah 2026, chaps. 2–6, apps. A–E). The book is the full treatment.