The asterisk de-idealized: Looking back at grammaticality, moving forward with conditioned stability

Brett Reynolds
Humber Polytechnic & University of Toronto
brett.reynolds@humber.ca

Abstract

The asterisk has done foundational work in theoretical linguistics, but it also hides a persistent conflation. The same diacritic is used to mark strings that defeat structural analysis, structurally viable utterances whose values don't stabilize under the constraints of an interpreted situation, interpretable forms that still aren't in the community's repertoire, and cases that are grammatically well-formed but feel unacceptable for reasons of processing or ideology. This article argues that the resulting competence–performance–usage triangulation hasn't resolved the impasse because grammaticality has been asked to answer several distinct questions at once.

Moving forward, I propose a minimal state theory that reconceptualizes grammaticality as conditioned stability of form–value relations within communicative situations. Grammatical status depends on (i) mapping viability, (ii) interpretive coherence, and (iii) repertoire status. Distinguishing grammatical status from the feeling of ungrammaticality makes principled sense of classic dissociations between acceptability ratings and repertoire membership. The proposal yields operational diagnostics for separating coherence failures from repertoire exclusion, motivates an opportunity-normalized notion of negative evidence, and states concrete conditions under which the framework would be disconfirmed.

Keywords: grammaticality; acceptability; form–value relations; norms; preemption; processing; communicative situation

1 Introduction

Every competent speaker of English knows that **Can the have running* is impossible, but the source of this certainty is still not settled. What, exactly, is being asserted when an utterance is labelled UNGRAMMATICAL? Consider the following cluster:

(1) a. **Can the have running?*
b. *Colorless green ideas sleep furiously.* (chomsky1957)
c. **I've finished it yesterday.*
d. *?I saw Joan, a friend of whose was visiting.*

e. *The bread the baker the apprentice helped made is delicious.*

f. A: *How old are you?* B: **I have 25 years.*

g. **Which did you buy car?*

These items share the folk verdict that "something's wrong", but they don't share a single type of failure. Example (1a) defeats structural analysis, while (1b) is structurally impeccable but conceptually bizarre, even if a construal can be recovered. (1c) is interpretively transparent but clashes in temporal value: tense and modifier pull in different directions. For many speakers, (1d) is not confidently rejected so much as judged marginal or uncertain in the repertoire. (1e) is often rejected in spontaneous use but becomes acceptable once a parse is stabilized, suggesting a processing-driven illusion. (1f) is viable and interpretable but isn't in the repertoire of the relevant English norm-centres, despite being ordinary in French and Spanish. (1g) is short and interpretable but treated as categorically excluded.

In short, the asterisk has been doing at least four jobs: marking structural crash (1a), marking interpretive incoherence (1c), marking repertoire exclusion (1f), (1g), and marking the feeling of anomaly driven by processing or ideology (1e). That heterogeneity is the phenomenon.

The history of grammaticality theory can be read as a sequence of attempts to compress such heterogeneity into a single explanatory core. Formal approaches treated grammaticality as categorical well-formedness; processing accounts treated gradience as performance; usage-based theories treated acceptability as the shadow of frequency and entrenchment; sociolinguistics treated grammaticality as norm-relative; experimental syntax refined measurement but didn't settle what's being measured. The result is a familiar triangulation in which the same data is alternately explained away as "competence", "performance", or "usage", often with little agreement on what would count as decisive evidence (schutze2016; sprouse2013).

This article is a contribution to the *Journal of Linguistics* section "Looking Back, Moving Forward". Looking back, I argue that the impasse persists because grammaticality has been asked to do the work of multiple distinct questions at once. Moving forward, I propose a minimal state theory that separates those questions, thereby restoring empirical vulnerability. The core proposal is that grammaticality is CONDITIONED STABILITY of form–value relations[1] within a communicative situation: grammatical status depends on (i) mapping viability (whether an expression-shape admits a well-typed structural analysis), (ii) interpretive coherence (whether the values encoded stabilize under the constraints live in the situation), and (iii) repertoire status (whether the form–value relation – especially at the operator stratum – is treated as a legitimate option in the relevant norm-centre). The same decomposition also clarifies how the FEELING OF UNGRAMMATICALITY arises as a metacognitive signal whose sources include, but aren't exhausted by, grammatical status; this distinction explains why some constructions feel ungrammatical while being licit, and why some illicit constructions escape detection (Fanselow2021).

The structure is as follows. Section 2 diagnoses the impasse by reviewing what the asterisk has been made to mean. Section 3 isolates the multiple questions that have been collapsed into one label. Section 4 introduces the state theory: conditioning states, the three constitutive quantities, and the stability score. Section 5 works through diagnostic profiles that the model predicts. Section 6

[1] I use VALUE for what a form conventionally contributes – primarily meaning, but extending to phonological and distributional regularities. Value is relational and contrastive: defined by opposition within a system, not by speaker intention (saussure1916).

addresses evidence and measurement, including a worked opportunity proxy. Section 7 frames key questions for future research. Section 8 states what would count against the framework. Readers who want the core model quickly can focus on §4.4.2–4.3, Table 3, and §8.

2 Looking back: what the asterisk has been made to mean

The modern theoretical role of grammaticality was shaped by the mid-century identification of grammar with a formal system generating a set of well-formed expressions. In this tradition, grammaticality is a categorical membership fact: a string is grammatical iff it's generated by the grammar (chomsky1957). This view captures the hard edge of cases like (1a), where the system crashes before any stable analysis is available. It also provides a clean division of labour: semantics and pragmatics interpret outputs; performance systems realize them.

The cost of this idealization is that it forces the field to treat gradience as epiphenomenal. The competence–performance distinction (chomsky1965) allowed formal theory to preserve categorical grammar by relocating variability to processing and attention, but the move is methodologically hazardous: once invoked, it can immunize the grammar from counterevidence by labelling inconvenient data as performance noise (schutze2016). Much subsequent work can be read as a search for principled ways to reintroduce gradience without abandoning the insight that some failures are genuinely categorical.

Meaning, coherence, and the limits of well-formedness present a second theme. Chomsky's (1b) was designed to show that structural well-formedness doesn't reduce to semantic plausibility. That point remains foundational: a theory that equates grammaticality with "interpretability" will misclassify many robust structural constraints. But (1b) also revealed a complementary fact: humans routinely accept structurally well-formed utterances whose values are conceptually odd, while rejecting other utterances whose intended interpretation is transparent. This tension motivated a long tradition of work linking acceptability to interpretive pressures, including semantic motivation for constraints (lakoff1971; mccawley1968) and constructional meaning (goldberg1995constructions).

These traditions didn't establish that meaning replaces grammar; rather, they showed that the stability of interpretation is itself a locus of constraint. An utterance may be structurally viable but fail because the values encoded by its parts can't be reconciled under the constraints that are live in a situation. The present perfect plus *yesterday* in (1c) is a canonical case (huddleston2002): the intended meaning is obvious, but the morphosyntactic temporal value conflicts with the adjunct anchoring. Conversely, in English at least, many lexical clashes are tolerated as long as they don't implicate morphosyntactic value.

Processing and the reallocation of gradience constitute a third response. If grammar is categorical but judgements are gradient, one obvious move is to treat gradience as a function of processing. The processing literature has supplied a large inventory of robust effects – dependency locality, interference, garden-path reanalysis – that depress ratings and slow reading times for structures that are otherwise analyzable (gibson2000; GrodnerGibson2005). Classic centre-embedding examples like (1e) are often treated as the poster children: they are grammatical in the sense of analyzable and interpretable, but they trigger strong negative responses because incremental parsing is strained.

Processing explanations, though, don't exhaust the landscape. Certain constructions remain sharply rejected even when short and interpretable, and even when repeated exposure doesn't improve ratings. The literature on satiation and adaptation was partly motivated by precisely this need

(snyder2000grammaticality; Snyder2022; reynolds2025hpcbook): some degraded structures improve with exposure, others don't, and the difference can't be reduced to length or memory load alone. While processing accounts for why such structures feel ungrammatical, it doesn't, by itself, constitute a theory of grammatical status.

Usage, norms, and the social life of grammaticality provide a fourth strand. Usage-based approaches shifted attention to the role of frequency and entrenchment: speakers learn the distributions of forms, and those distributions shape what feels acceptable (bybee2006; bybee2010; reynolds2026lbe). A key advance in this tradition is the recognition of PREEMPTION: a form can be rejected because a competitor is consistently selected in the same niche, even if the discarded form remains structurally possible (Goldberg2011). The contrast between *I'm 25 years old* and **I have 25 years* in English illustrates the point: the latter is transparent and structurally viable, but is systematically excluded from the repertoire of the relevant norm-centres.

Sociolinguistic accounts, meanwhile, emphasize that grammaticality resides not in the abstract properties of a language, but in a community's normed repertoire (labov1972). Indexical values attached to forms can shift what a situation admits to the repertoire, and speakers routinely disagree about what counts as "the" grammar because they construe different norm-centres as relevant (Silverstein1976; Eckert2012). Far from an embarrassment, this constitutes part of the phenomenon. The problem's that, in much theoretical practice, norm-relativity's treated as a complication external to grammar rather than as a constitutive feature of what grammatical status amounts to.

3 The impasse diagnosed: three questions collapsed into one label

Rather than noise, the heterogeneity in (1) is structural. Four things have been collapsed, but only three are constitutive of grammatical status itself; the fourth – the feeling of ungrammaticality – is a distinct phenomenon that needs to be separated. Grammaticality theory has repeatedly attempted to treat grammatical status as a unified phenomenon when it is, in fact, the intersection of three distinct questions.

First, structural viability. Some inputs fail because no structural analysis is available that yields a well-typed morphosyntactic representation. In such cases, the failure is categorical and doesn't depend on meaning, social norm-centres, or processing effort; the analysis crashes. Example (1a) is emblematic: the category sequence prevents the construction of a viable constituent structure.

Treating this failure mode as real is non-negotiable: without it, the notion of grammar loses its basic explanatory purchase. The mistake lies in elevating this single prerequisite into the definition of grammaticality itself.

Second, interpretive coherence. Many strings are structurally viable but unstable in value. Sometimes the instability is semantic (temporal alignment, argument structure satisfaction); sometimes pragmatic or information-structural (topic/focus fit); sometimes indexical (social meaning clashes with footing). The common thread lies in the stability of a dominant construal under the constraints that are live in the relevant situation, rather than in a folk notion of "meaningfulness".

Example (1c) illustrates: the intended interpretation is obvious, but the morphosyntactic value encoded by the present perfect conflicts with the temporal anchoring provided by *yesterday*. The result is interpretive instability grounded in conventional form–value relations, rather than structural nonsense.

Third, repertoire status. A third class of cases are structurally viable and interpretively coherent,

but rejected because they aren't in the community's repertoire. Here the role of usage and norms is constitutive: the community hasn't conventionalized the relevant form–value relation as a legitimate option under the norm-centres that define the communicative situation.

Example (1f) is again emblematic. The form isn't nonsensical, and it's interpretable. Its rejection is a fact about English community conventions, not about universal cognitive limits. The same form is licit in other languages, demonstrating that the relevant factor is repertoire status, not viability or coherence.

A fourth label deserves separation: the feeling of ungrammaticality. The three components above are constitutive for grammatical status. But speakers' judgements also reflect a FEELING OF UNGRAMMATICALITY: a metacognitive negative signal triggered by instability or high repair cost. This feeling is an important object of study, but it isn't identical to grammaticality. It yields false positives, where licit constructions feel bad, and false negatives, where illicit constructions pass undetected (Fanselow2021). Equating ratings with grammatical status invites conceptual confusion.

Four concepts need to be kept apart. APPROPRIATENESS is the genus: the fit between a form and the context in which it's used. GRAMMATICALITY is one species of appropriateness – the coupling between grammatical form and the values it conventionally expresses. It has a fact of the matter: either the form–value relation is in the repertoire for the relevant conditioning state or it isn't – even when that fact isn't directly accessible to measurement.[2] ACCEPTABILITY is the measurement channel: how speakers rate utterances, informed by grammatical status but also by processing factors, repair costs, and ideological filtering. CORRECTNESS is the prescriptive overlay: what gatekeepers enforce, what gets codified and moralized – often an ideologized version of one variety's appropriateness norms imposed as if universal. A predictable objection to the framework below is that C_t (repertoire status) merely relabels prescriptive correctness. The answer is no: correctness concerns what should be enforced; repertoire status is a constitutive fact about what a norm-centred population actually treats as a legitimate resource. Enforcement can distort evidence, but the repertoire state isn't the prescription (pullum2019-normativity).

The remainder of the paper proposes a minimal state theory that makes these distinctions explicit, thereby clarifying what it is for an utterance type to be grammatical *in a communicative situation*.

4 MOVING FORWARD: GRAMMATICALITY AS CONDITIONED STABILITY OF FORM–VALUE RELATIONS

The state theory commits to just three things: analyzability, stability of construal, and repertoire status – each relativized to a construed situation. The notation that follows is bookkeeping, not a new philosophical burden.

4.1 CONDITIONING STATES AND COMMUNICATIVE SITUATIONS

Consider a speaker who says *I seen it*. In a classroom presentation, this is likely to be heard as ungrammatical; at lunch with friends, it may pass without comment. The string hasn't changed; what's changed is which norm-centre is in play and what's at stake. This is what conditioning captures.

[2] The distinction between ontological fact and epistemic access is developed in reynolds2025hpcbook<empty citation>, which argues that category boundaries are structurally determinate but located at thresholds we cannot finitely specify; gradient judgments arise from discrete categories filtered through processing noise, not from gradient membership.

Let c be a CONDITIONING STATE: a construed communicative situation together with whatever norm-centre is treated as relevant (wiese2023). The point isn't to reify c as a fixed external context; interlocutors can misalign about which c is in force, and c can be renegotiated. The modelling commitment is simply that grammatical status is always assessed relative to some such conditioning.

Rather than an optional sociolinguistic add-on, this move is the minimal way to state the empirical fact that grammars are socially situated repertoires: the same speaker can treat different resources as in-repertoire in different situations, and different speakers can rationally disagree about repertoire membership when they construe different norm-centres. This is the core of the realist commitment: grammaticality isn't an abstract property of the string, but a measurable state of the relation between form, value, and agents in a constructed situation.

The conditioning state c can be decomposed into at least three anchors (reynolds2026varieties):

- SITUATION (S): the here-and-now interactional frame – activity type, medium, footing, institutional context.
- ASCRIPTION (A): what the speaker is treated as – the social categories assigned by self and others, which condition expectations about baseline repertoire.
- IDENTIFICATION (I): whose norms are being oriented to – the reference population the speaker treats as the standard for what counts as legitimate.

Together with situational stakes, these yield a conditioning vector $c \approx \langle S, A, I, \text{stakes} \rangle$. The decomposition matters because apparent disagreement about grammaticality often reflects misalignment in A or I rather than genuine conflict about the state of the form–value relation. The same token can be processed as dialectal (in-repertoire under one ascription) or as an error (out-of-repertoire under a different norm-centre), depending on which conditioning anchors the listener infers. This is why the classroom/lunch contrast for *I seen it* isn't just about formality (S); it's also about whose norms are in play (I) and what categorization the listener assigns to the speaker (A).

4.2 THREE CONSTITUTIVE QUANTITIES

For an utterance type u in conditioning state c at time t, define three state quantities.

The first quantity is mapping viability. Let $\mathsf{map}(u, c) \in \{0, 1\}$ be a binary indicator of whether there exists at least one viable morphosyntactic analysis for u in c for which there's a well-typed representation (where 1 is viable and 0 isn't). map is intended to capture genuine analyzability failure and only that. It's the categorical prerequisite highlighted by the well-formedness tradition – the "entry ticket" to the system – but it doesn't on its own guarantee either interpretive coherence or repertoire membership. Many ungrammatical strings remain easily parsed and "interpreted" in a folk sense.

The second quantity is interpretive coherence. Let $K(u, c) \in [0, 1]$[3] represent the stability of interpretation: the degree to which the utterance yields a dominant, non-contradictory construal under the constraints live in c (ranging from 0, complete instability, to 1, perfect coherence). Formally, K can be modelled as concentration of a distribution over candidate construals; for present purposes, the important point is that K is distinct from map. Structural viability doesn't guarantee coherence.

[3] The curly braces $\{0, 1\}$ denote a two-member set (exactly 0 or 1); the square brackets $[0, 1]$ denote the continuous interval from 0 to 1 inclusive.

The third quantity is repertoire status. Let $C_t(u, c) \in [0, 1]$ be the population-level posterior probability that the form–value relation u is in the community's repertoire for c: the probability that an individual drawn from the relevant norm-centred population treats u as a legitimate resource rather than as an error, performance slip, or alien form (where 1 represents universal acceptance and 0 indicates total exclusion). This quantity is related to what usage-based work calls entrenchment, but it's explicitly conditioned on c and includes normative dimensions that pure entrenchment doesn't capture. While C_t is informed by frequency, it represents inferred repertoire membership rather than a simple tally of token occurrences.

C_t is where norms live. It's also where many apparently categorical exclusions can be located without positing hard representational bans: a form can be structurally viable and interpretable while being near-universally excluded from the repertoire in a given situation.

The $S/A/I$ decomposition of c introduced above clarifies what fixes "the community" for a given evaluation. Identification (I) is the natural anchor for whose repertoire counts in C_t: the reference population is whoever the speaker is orienting to. Situation (S) and stakes are the natural anchors for the decision regime $\tau(c)$: high-stakes institutional contexts raise the threshold. Ascription (A) explains a major source of apparent inconsistency: the same string can be treated as in-repertoire when attributed to one ascribed group and as an error when attributed to another, even if the underlying variety grammar is the same. That isn't a change in map or K; it's a change in how listeners map tokens to norm-centres.

Symbol	What it tracks	Diagnostic evidence
map	Structural analyzability (0/1)	Parse failure, no stable category assignment
K	Interpretive coherence (0–1)	Paraphrase dispersion, construal instability
C_t	Repertoire status (0–1)	Production rates, "would you say this?", corpus frequency normalized by opportunity

Table 1: The three constitutive quantities at a glance.

4.3 A stability score and a membership predicate

Define a graded stability score:

$$\widetilde{G}_t(u, c) = \mathsf{map}(u, c) \cdot K(u, c) \cdot C_t(u, c) \in [0, 1]. \tag{4.1}$$

This multiplicative scoring means that if any single component is zero – if the mapping fails, if interpretation is impossible, or if the form isn't in the community's repertoire – the entire relation is ungrammatical. Stability underwrites gradience: lowering any component reduces the overall score.

In plain terms: the product rule says that being in-repertoire can't make up for being incoherent, and being perfectly coherent can't make up for being out-of-repertoire. Deficits compound rather than average out.

This decomposition reflects a broader organization of linguistic infrastructure. Expression-shape constraints (phonotactics, morphotactics) regulate whether an utterance is recognizable as a token of

the system; their violation yields "not a word". Operator-like constraints – closed-paradigm contrasts that configure public update, allocate participant roles, and authorize uptake – are targeted by K (for value coherence) and C_t (for community repertoire); their violation yields "you can't say that". Payload resources (open-class lexicon, indexical stance) remain negotiable and extensible; their misuse invites clarification or social judgment, not structural rejection. The stability score $\widetilde{G}_t$ integrates across these levels: a form that crashes at any level is unstable, but the *type* of instability differs diagnostically.

Communities also often treat grammaticality as a categorical membership fact: either a resource is in the repertoire or not. Model this by thresholding:

$$G_t(u, c) = \mathbb{I}\Big[\widetilde{G}_t(u, c) \geq \tau(c)\Big], \tag{4.2}$$

where $\tau(c)$ is a situation-specific decision criterion. The point is that $\tau(c)$ is a property of how strict the situation is about what counts as "in" the repertoire. High-stakes institutional contexts can set a high threshold; low-stakes in-group contexts can set a lower one.

To illustrate: in a classroom presentation, using a stigmatized dialectal form risks being marked down; the threshold for "grammatical enough" rises. At lunch with friends, the same form may index solidarity; the threshold drops. Same form, same $\widetilde{G}$, different verdicts.

A concern is that $\tau(c)$ might immunize the theory if it can vary freely. Two constraints matter. First, $\tau(c)$ isn't construction-specific: it's fixed for a conditioning state and therefore shifts the boundary for *all* utterance types evaluated in that state. Adjusting τ to rescue a single problematic case entails collateral predictions for a broad set of anchor items. Second, $\tau(c)$ can be motivated by a standard decision-theoretic rationale in which classification losses differ by situation. Let $L_{\text{FA}}(c)$ be the loss of treating an item as in-repertoire when it is not, and $L_{\text{FR}}(c)$ the loss of treating an item as not-in-repertoire when it is. A natural constraint is

$$\tau(c) = \frac{L_{\text{FA}}(c)}{L_{\text{FA}}(c) + L_{\text{FR}}(c)}.$$

This ties τ to independently characterizable properties of c (stakes, institutional norms, gatekeeping pressure). Empirically, $\tau(c)$ can be estimated by calibrating participants on an anchor set spanning clear in-repertoire and clear not-in-repertoire items for the target c, rather than being tuned post hoc to accommodate the construction under dispute.

This formalizes an intuition that's often stated informally but rarely built into the state theory: what counts as "grammatical" for practical purposes depends on the decision regime of the situation, not just the resource itself.

4.4 Why this combination rule isn't decorative

(If you accept the decomposition, the point of this subsection is just to show that the multiplication yields discriminable interaction predictions. Skim if pressed.)

The three-way decomposition is the theoretical commitment; the choice of a specific combination operator is a modelling decision. But the operator should be constrained by desiderata that make it empirically non-trivial.

First, the core is non-compensatory: mapping failure, catastrophic incoherence, or categorical repertoire exclusion should each be sufficient to drive grammatical status to zero in the relevant c.

This excludes simple weighted sums as a model of grammatical status, since they permit a high value on one dimension to compensate for near-zero on another. Second, the graded score should reflect compounding instability: two moderate deficits should typically be worse than either deficit alone. Third, the operator should be monotone in each argument, and it should allow a transparent generalization to relative weighting if later work justifies it.

Several standard operations meet the non-compensatory constraint. The minimum operator,

$$\widetilde{G}_t^{\min}(u,c) = \min\{\mathsf{map}(u,c), K(u,c), C_t(u,c)\},$$

treats the weakest link as decisive. This makes a clear prediction: once one component is identified as the bottleneck, further degradation elsewhere shouldn't matter for the objective score. At the other extreme, a weighted sum predicts systematic compensation:

$$\widetilde{G}_t^{\Sigma}(u,c) = w_{\mathsf{map}}\mathsf{map} + w_K K + w_C C_t,$$

which is often plausible as a model of subjective ratings but is a poor fit for a state theory of grammatical status precisely because it allows a community to "make up for" incoherence by repertoire status alone.

The product rule adopted in (4.1),

$$\widetilde{G}_t^{\times}(u,c) = \mathsf{map}(u,c) \cdot K(u,c) \cdot C_t(u,c),$$

is the simplest operator that's non-compensatory and compounding. It also has a useful interpretive property: in log-space, the components contribute additively ($\log \widetilde{G} = \log \mathsf{map} + \log K + \log C_t$), which aligns naturally with an evidence-accumulation picture in which distinct sources of instability contribute independent penalties. If future work motivates differential weighting, the product generalizes straightforwardly to a weighted geometric form, $\mathsf{map} \cdot K^{\alpha} \cdot C_t^{\beta}$, with $\alpha, \beta > 0$.

The choice among min and $\times$ is empirically discriminable. Consider a factorial manipulation that independently lowers coherence and repertoire status while holding mapping constant: for the same morphosyntactic frame, introduce a mild value-clash (lowering K) and, independently, present the construction under a norm-centre that treats it as non-native or marginal (lowering C_t).[4] The minimum rule predicts that once either K or C_t is the bottleneck, the second manipulation shouldn't further depress the objective score; the product rule predicts a systematic interaction (compounding), since the combined manipulation reduces stability more than either alone. This is a substantive prediction about the structure of the state space, not a restatement of the verbal story.

An information-theoretic perspective clarifies why this matters. The multiplicative structure has a natural interpretation in terms of how linguistic contrasts contribute to interpretation. Some form–value relations occupy small, closed paradigms but cause large downstream consequences: clause type constrains which responses are relevant; polarity flips entailment relations; case and agreement constrain role assignment. These relations function as control settings – protocol headers rather than payload content – carrying few bits in themselves but causing large entropy reduction in the space of licit interpretations (shannon1948; coverthomas2006). A wrong value doesn't merely produce a

[4] The repertoire-status manipulation can be implemented by norm-centre framing (ingroup/dialect/resource vs error) and by register/stakes manipulations, which are predicted to affect $\tau(c)$ and, in some cases, C_t itself.

surprising concept combination; it disrupts the mapping from form to publicly recognizable update. This is why a short, interpretable utterance like (1g) can trigger categorical rejection: the violation targets infrastructure, not content. The product rule captures this asymmetry: degradation in control-like dimensions (map, K for operator-relevant constraints, C_t for high-opportunity paradigms) compounds rapidly, while payload-level infelicities remain negotiable.

Empirical upshot: $K \times C_t$ manipulations should compound under the product rule; if they don't, the operator is wrong even if the decomposition stands.

4.5 Separating coherence from repertoire status: operational criteria

A recurring worry is that coherence failure $K \approx 0$ and repertoire exclusion $C_t \approx 0$ may collapse into one another, since both yield low stability. The separation requires distinct measurement signatures.

Low K (coherence failure)	Low C_t (repertoire exclusion)
Construal unstable; speakers disagree on meaning	Construal stable; speakers agree on meaning
Paraphrase dispersion high	Paraphrase agreement high
Repair-heavy, effortful interpretation	Readily interpretable
"What does that even mean?"	"I know what you mean, but we don't say that"
Diagnostics: paraphrase tasks, construal variability, RT to inference questions	*Diagnostics:* production probability, "would you say this?", corpus frequency / opportunity

Table 2: Separating coherence from repertoire status: predicted contrasts.

These diagnostics cut across the tempting verbal contrast between "values can't be reconciled" and "the community doesn't accept the reconciliation". In practice, the decisive question is whether the source of degradation is interpretive dispersion or repertoire exclusion.

This is why (1c) is a useful but non-trivial diagnostic. Many speakers can recover the intended meaning of *I've finished it yesterday* with little difficulty, which pushes it toward a low-C_t profile (repertoire exclusion of a specific tense–adverb pairing) rather than a pure low-K profile. On the other hand, if an experimental design reveals systematic competition between two construals (a present-perfect reading vs a coerced simple-past reading), then the same item will show low K by exhibiting dispersion in paraphrase and inference tasks even when participants are instructed to treat the form as a legitimate dialectal resource. The framework is falsifiable here: it predicts that the K-diagnosis and the C-diagnosis diverge in their measurement signatures.

4.6 Grammaticality versus the feeling of ungrammaticality

The state theory above defines grammatical status via $\widetilde{G}_t$ and $\tau(c)$. Speakers' ratings often track a different quantity: a subjective ungrammaticality signal driven by low stability, processing costs, and ideological overlays.

A useful way to characterize this signal is as inverse conditioning. If speakers condition production on S, A, and I, then listeners can infer those conditioning anchors from observed forms – Bayes' theorem running in reverse. The feeling of ungrammaticality is naturally tied to surprisal relative to the listener's inferred conditioning model: hearing a form that's low-probability under the

c the listener thinks is in force triggers the signal. Processing costs and ideological overlays layer on top, but the core input to the detector is $-\log P(u \mid c)$. This framing has two methodological payoffs. First, it gives a principled bridge from ratings to a measurement channel: ratings are observations of a detector whose input includes surprisal plus processing costs, not direct observations of C_t. Second, it makes the later discussion of language models less risky: LMs approximate something like $P(u \mid \text{context})$ for some training-conditioned mixture of c's, which is naturally closer to "detector input" than to "truth about G_t".

This distinction predicts systematic dissociations:

- Licit but degraded: $\mathsf{map} = 1$, K high, C_t high, but processing costs depress ratings (classic centre embedding).
- Illicit but unnoticed: $\mathsf{map} = 1$ and the intended meaning is salient, so the ungrammaticality signal is weak even when a relevant coherence constraint is violated (agreement attraction and other slips in complex structures; wagers2009agreement).

Equating acceptability ratings with grammatical status conflates a state claim with a measurement channel (reynolds2025hpcbook). The methodological consequence is that claims about G_t should be supported by converging indicators, with ratings treated as evidence primarily about the ungrammaticality signal and only indirectly about repertoire status.

5 Diagnostic profiles: what different failures look like

Beyond being definitional, the value of a state theory lies in the diagnostic profiles it predicts. The decomposition in (4.1) yields a compact typology of recurrent instability modes. The typology reflects that an utterance can be structurally well-mapped and easily "interpreted" in a folk sense while remaining ungrammatical due to coherence failure or repertoire exclusion.

Profile	Canonical signature
$\mathsf{map} = 0$	Structural crash; categorical rejection; no amount of context stabilizes meaning (1a).
$\mathsf{map} = 1$, $K \approx 0$	Value incompatibility; intended meaning might be guessable, but conventional form–value constraints in c block stabilization (1c).
$\mathsf{map} = 1$, K high, $C_t \approx 0$	Repertoire exclusion; interpretable but treated as not in the repertoire; often cross-linguistically variable (1f, 1g).
$\mathsf{map} = 1$, K high, C_t low/uncertain	Rarity/indeterminacy; weak consensus; high variance across speakers (1d).
$\mathsf{map} = 1$, K high, C_t high, but high processing cost	Illusory ungrammaticality; improves with guidance; ratings track repair cost more than status (1e).

Table 3: Recurrent diagnostic profiles as regions of the state space.

To see the table at work, revisit the opening cluster. *Can the have running* (1a) is row 1: map $= 0$. *I've finished it yesterday* (1c) is row 2: the string parses, but temporal values clash (K low). *I have 25 years* (1f) is row 3: fully interpretable, but English doesn't have it in repertoire ($C_t \approx 0$). *A friend of whose* (1d) is row 4: the opportunity set is small, so speakers are uncertain rather than categorical. *The bread the baker...* (1e) is row 5: licit but processing-heavy, producing illusory ungrammaticality. The table isn't ornamental; it partitions the puzzle set.

Two contrasts are key for the future research agenda: stable repertoire exclusion versus rarity, and objective status versus felt ungrammaticality.

5.1 Stable repertoire exclusion versus rarity

A raw corpus absence is compatible with two very different states. A construction can be rare because the opportunity set is tiny, leaving speakers with little evidence either way; or it can be rare because, despite a large opportunity set, it's systematically preempted by competitors, driving repertoire status toward zero. The independent relative genitive in (1d) plausibly belongs to the first class for many speakers: the configuration that would make it useful is itself rare, so the absence of tokens doesn't straightforwardly imply categorical exclusion.

Left-branch extraction in (1g) behaves differently. The communicative niche is common, competitors are available (*Which car did you buy?*), and speakers show robust categorical rejection. This profile is analyzed as near-zero repertoire status in the relevant norm-centres, consistent with a preemption-based trajectory (Goldberg2011; reynolds2026lbe). In this view, categoricality needn't be located in map: the intended analysis can be available and interpretation can be coherent once stipulated, while the community treats the relation as excluded from the repertoire.

5.2 Illusory ungrammaticality and misattribution

Processing-driven illusions illustrate why the feeling of ungrammaticality can't be equated with grammatical status. Centre embedding (1e)'s analyzable and interpretable, but incremental parsing strains working memory and dependency integration, triggering strong negative affect. Similarly, garden-path items can feel nonsensical until reanalyzed:

(1) *The old man the boats.* (ritchie1984)

A first-pass parse yields nonsense; reanalysis yields a coherent, licit structure. In such cases, ratings track repair difficulty, not repertoire status. Conversely, illicit structures can pass unnoticed when meaning is compelling, yielding false negatives (pullum2009).

The repair system provides converging evidence. When repair does occur, mismatches targeting operator-like dimensions – tense errors, agreement failures, clause-type confusions – are predicted to elicit open-class repair initiation (*what?*, *who did it?*) and explicit rejection, because they disrupt the publicly accountable control settings on which uptake depends. Mismatches targeting payload or indexical dimensions are predicted to elicit stance negotiation and accommodation (*did you mean...?*, *why are you talking like that?*), because the utterance's update potential remains intact even when its content or social positioning is problematic. This asymmetry is independent of the feeling of ungrammaticality: a processing-heavy but licit structure may feel terrible without triggering the repair profile associated with genuine operator failure.

The state theory predicts such dissociations whenever the ungrammaticality signal pools multiple sources of difficulty.

6 Evidence and measurement: what it would take to test the state theory

A "moving forward" programme has to specify what would count as evidence. The constitutive variables suggest a principled division of labour among data types.

For mapping viability, evidence comes from analyzability: whether speakers can assign a stable category structure, whether repairs consistently fail, and whether comprehension collapses even under supportive contexts. Structural crash cases are rare but diagnostically clean.

For coherence, evidence comes from interpretive stability under controlled manipulations of the relevant constraints (temporal alignment, argument structure, information structure, indexical consistency). Here experimental pragmatics and semantics supply tools for isolating which constraints are doing the work, while corpus work can reveal conventional distributional restrictions that track those constraints.

For repertoire status, C_t is latent and can't be inferred from ratings alone. It has to be estimated from converging indicators: production probability in elicitation, corpus frequency normalized by opportunity sets, repair behaviour, recognition latency, and social evaluation under explicit norm-centre manipulations. The state theory motivates an explicit measurement model for C_t in which acceptability ratings are treated primarily as observations of the ungrammaticality signal, not of repertoire status.

One central challenge involves operationalizing opportunity. Preemption-based accounts require not only token counts but niche counts: how often the communicative job arises in the relevant c. A key empirical task for the moving-forward agenda is to develop operational definitions of niches for different constructions and to measure non-occurrence relative to those opportunities.

6.1 A worked opportunity proxy: independent relative genitives

Opportunity-normalisation is the hinge between mere corpus rarity and evidence of systematic exclusion. The general problem is that niches aren't directly annotated in corpora: we rarely observe "the speaker needed to express X" as an explicit variable. A workable starting point is to use competitor forms as a lower-bound proxy for opportunities. If speakers reliably realise a niche using an established competitor, then each observed competitor token witnesses an opportunity in which the target variant could, in principle, have been selected.

For the independent relative genitive in (id), a plausible niche is: "predicate something of an associate of a discourse-salient possessor while packaging the possessor relation in a relative dependency". Directly counting such niches is difficult. But we can approximate an opportunity lower bound by counting competitor realizations in the same discourse environments, such as: (i) *I saw Joan; a friend of hers was visiting*; (ii) *I saw Joan; one of her friends was visiting*; (iii) *I saw Joan, whose friend was visiting*; (iv) *I saw Joan; Joan's friend was visiting*. Each token of (i)–(iv) is evidence that the niche occurred and was realized by a competitor.

Let N^* be the competitor count in a corpus slice intended to approximate the relevant c (genre, register, period). Then N^* provides a conservative lower bound on the opportunity set $N_t(n, c)$. The question "does zero attestation matter?" becomes "is zero attestation surprising given N^* and a plausible counterfactual choice probability?". If a target variant would have been chosen with probability ρ among in-repertoire competitors, then the expected number of tokens is $N^*\rho$. Even small values of ρ yield strong expectations when N^* is large; conversely, when N^* is small, absence is weak evidence and should predict uncertainty rather than categorical exclusion.

A toy illustration: suppose $N^* = 10{,}000$ competitor tokens and $\rho = 0.005$ (the target form would be chosen one time in 200 if in-repertoire). Expected tokens: 50. Observing zero is then striking – strong evidence the form isn't in-repertoire. But if $N^* = 50$ and $\rho = 0.005$, expected tokens are 0.25; observing zero tells us almost nothing.

This proxy operationalization is deliberately coarse, but it's already discriminating: it separates cases where "no tokens" is probative (large N^*) from cases where it isn't (small N^*). It also makes clear what a "moving forward" corpus programme has to provide: explicit definitions of competitor sets for niches and principled choices of corpus slices approximating c.

7 Key questions for future theoretical research

The state theory reframes several longstanding debates as tractable research questions.

How should conditioning states be operationalized? If grammaticality is conditioned, then specifying c isn't optional. Future work has to develop operational proxies for norm-centres and communicative situations: genre, medium, stance, audience design, institutional stakes, and community membership. An important prediction is that constructions whose status is driven by C_t will be more sensitive to c-manipulation than map-failures and many coherence-failures.

What are the right objects of repertoire membership? The theory treats u as an utterance type, but in practice the granularity of u matters. Is *gave the dog a bone* in-repertoire as a specific string, as an instance of the ditransitive construction, or as part of a broader caused-possession family? A moving-forward programme has to articulate principled criteria for individuating u in a way that makes the repertoire term empirically meaningful rather than vacuous.

What is the etiology of stable gaps? The present paper has remained mostly constitutive. The natural next step is an etiological module: a model of how $C_t(u, c)$ trajectories arise under positive evidence, error evidence, and opportunity-sensitive preemption. Instead of debating whether preemption exists, the crucial question is its effective strength across niches and how it interacts with processing difficulty and social evaluation. This is where classic "categorical" constraints become a test case: the moving-forward claim is that at least some of them can be redescribed as stable repertoire exclusion sustained by strong preemption in robust opportunity sets.

A related question is which form–value relations attract sharp repertoire boundaries in the first place. The present framework is neutral on this, but a natural hypothesis is that repertoire policing clusters around OPERATOR contrasts: closed-paradigm choices that configure public update, allocate participant roles, and constrain uptake – clause type, argument linking, tense–aspect where grammaticalized, evidential anchoring. If this is correct, then C_t trajectories are shaped not only by opportunity mass but by the functional load of the contrast: high-entropy-reduction dimensions attract categorical policing because a wrong value causes coordination failure even when the utterance is otherwise intelligible. This reframes the "categorical vs. gradient" debate as a question about which dimensions of the state space are operator-like, rather than about whether gradience is real (reynolds20260operators).

A complementary etiological resource is coordination equilibria in the sense of evolutionary game theory (oconnor2019games). On this view, communicative situations are payoff structures, and repertoire boundaries stabilize because they solve recurring coordination problems. Some partitions become sharp and policed because category salience enables coordination: once a contrast is salient, speakers and listeners expect each other to respect it, and deviation is costly. This explains why C_t can

remain near zero for forms that are structurally viable and interpretable – the coordination equilibrium excludes them. It also explains why certain boundaries resist erosion even under exposure: the equilibrium is self-sustaining because unilateral deviation is penalised. This game-theoretic module is compatible with the constitutive framework but adds an explanation of why some gaps are stable and others drift. The threshold $\tau(c)$ fits naturally into this picture: high-stakes situations are precisely those where coordination failure is costly, and institutions often encode the expected equilibrium as explicit gatekeeping.

How should typological generalizations be interpreted? If grammatical systems are normed repertoires shaped by stability dynamics, typological regularities are naturally viewed as recurring attractors in design space rather than as exceptionless laws. The task is to identify which combinations of form–value relations are robustly stable across lineages and which are contingent on local history and norm-centres. Large-scale typology becomes evidence about the global stability landscape rather than a direct route to categorical universals.

What role should language models play? Language models are now unavoidable instruments in linguistic practice. The state theory suggests a principled way to use them without mistaking their outputs for grammatical truth. If a model is treated as a proxy for the ungrammaticality signal, it may be useful for predicting processing difficulty and surprisal-like effects; if it's treated as evidence about repertoire status, it has to be grounded in opportunity-normalized distributions and norm-centre conditioning. The resulting agenda is methodological: what, exactly, are models approximating when they mimic human judgements, and which variable in (4.1) does that approximation correspond to?

8 What would count against this framework?

A framework that decomposes grammatical status into multiple components risks appearing too flexible unless each component is tied to independent evidence. The present proposal is disconfirmed, or at least seriously pressured, by any of the following patterns.

Each condition targets a specific earlier commitment: (1) targets the K/C_t separation (§4.5); (2) targets the opportunity methodology (§6.1); (3) targets conditioning and thresholds (§4.4.3); (4) targets the combination rule (§4.4); (5) targets the decomposition as a whole.

1. If there's no measurable dissociation between coherence and repertoire status – if constructions diagnosed as low-C_t (repertoire exclusion) systematically exhibit the same interpretive-dispersion profile as constructions diagnosed as low-K, and if tasks designed to separate these signatures fail across a range of phenomena – then the K/C distinction isn't empirically supported.

2. If opportunity-normalized absence doesn't discriminate stable gaps from rarity – if constructions widely treated as "categorical" don't show strong opportunity proxies via large competitor counts (N^*) while rare/uncertain constructions do – then the central methodological claim about opportunity-sensitive negative evidence is undermined.

3. If norm-centre and stakes manipulations don't affect the predicted targets, the conditioning architecture is mis-specified. The $S/A/I$ decomposition yields a specific experimental toolkit: manipulate S via genre/register framing, institutional roleplay, or audience-design cues; manipulate A via speaker ascription cues (biographical metadata, voice/ethnolectal markers, explicitly stated background); manipulate I via explicit norm-centre orientation cues ("speaking

as a member of X community", "aiming for formal norms", "in-group banter"); manipulate stakes via consequence framing that should shift $\tau(c)$ without necessarily shifting C_t. If such manipulations don't systematically shift threshold behaviour (as indexed by anchor sets) and don't preferentially affect constructions hypothesized to be repertoire-sensitive, the $S/A/I$ structure isn't just a conceptual repackaging; it's a design toolkit that tells you what counts as a clean manipulation of c rather than a vague "context effect".

4. If the combination rule makes the wrong interaction predictions – if factorial manipulations that independently target coherence and repertoire status show no compounding interaction where the product rule predicts one – then either a different non-compensatory operator is required (e.g. min), or the assumption that the components contribute independently to objective stability is incorrect.

5. If there are robust cases of categorical exclusion with high independent evidence of repertoire membership – if a construction is demonstrably used productively in the relevant c (high production probability and opportunity-normalized corpus rates) and yields stable construals, but is still treated as categorically ungrammatical in repertoire-membership tasks by the same population – then the proposal that grammatical status is constituted by repertoire status plus coherence plus mapping is incomplete.

These conditions are intentionally stated as empirical profiles rather than as verbal counterexamples, since the point is to align theoretical claims with distinct measurement channels.

9 Conclusion

Looking back, grammaticality has functioned as a foundational organizing notion in theoretical linguistics, but it has been burdened with incompatible tasks: marking structural crash, signalling coherence failure, recording community norms, and reporting subjective ungrammaticality. The resulting conceptual overloading has fuelled recurring disputes about whether data is "competence", "performance", or "usage".

Moving forward, grammaticality can be reconceptualized as a state property: conditioned stability of form–value relations within a communicative situation. A minimal decomposition into mapping viability map, interpretive coherence K, and repertoire status C_t yields a compact diagnostic typology and clarifies why acceptability ratings are an imperfect thermometer.

The same framework reframes categorical exclusions as potentially emergent stable repertoire exclusion sustained by opportunity-sensitive preemption, and it motivates a concrete research agenda: operationalizing conditioning states, defining the objects of repertoire membership, measuring opportunity sets, and building convergent estimators for repertoire status that don't collapse grammatical status into subjective affect.

If grammaticality is to remain a useful concept for theoretical linguistics, it has to become a target of explanation rather than a presupposed label. The state theory proposed here is intended as a step in that direction: it doesn't replace existing insights about structure, meaning, processing, or norms, but is a minimal architecture that makes their interaction explicit and testable. In this, the asterisk is de-idealized: it stops being a mark of abstract ill-formedness and becomes a realist diagnostic of stability failure in a situated communicative state.

Acknowledgements

Thanks to Peter Evans, Geoff Pullum, Muhammad Ali Khalidi, Ryan Nefdt, Irene Kosmas, and Mostafa Hasrati for comments and suggestions. Henri Kauhanen reviewed the formalization.

I used the large language models Claude 3.5 & 4.5; ChatGPT o1 pro & 5.2 pro; Gemini 3; and DeepSeek V3 in drafting and editing this paper.