Hedging as prepositional focusing: A learnable-theoretic account of the persuasive sweet spot

Global Journal of Social Sciences Studies
Vol. 12, No. 2, 18-33, 2026.
e-ISSN: 2518-0614

Luca Magni

Luiss Business School, Rome, Italy.
Email: luca.magni@luissbusinessschool.it

ABSTRACT

This paper develops an account of why hedging, the linguistic marking of uncertainty, can in specific configurations increase rather than diminish persuasive impact, extending recent evidence that hedges combining high-likelihood language with personal perspective occupy a communicative sweet spot. Using the conceptual apparatus of Learnable Theory, the Learnable Enhanced Bhaskarian Ontology (LEBO), prepositional focusing and defocusing, the Complementary Semantic Distance Index (CSDI), the umbra cone construct, and Differenza Inversa (Inverse Difference), the paper reframes hedging not as a graded modulation of propositional certainty but as an act of prepositional attachment occurring at the Learnable stratum, in which a speaker either anchors a claim to a locatable subject, the self, or leaves it ownerless. The sweet-spot register is shown to carry a predictable CSDI signature, Self-Actualization-Positive, which produces an umbra cone effect shadowing Safety-domain content within the same message, so that maximal persuasiveness and maximal disclosure pull in opposite directions. Differenza Inversa, originally an intrapsychic operator for reinterpreting autobiographical memory, is extended here, as a proposed analogy, to the interpersonal resolution of unmarked speaker/claim attribution, and seven testable propositions are derived from the resulting account. The framework offers organisational communicators, leaders, and designers of human-AI interaction a mechanism-level explanation of when and why hedged language builds trust rather than undermining it, together with the boundary conditions under which personal anchoring ceases to buffer a low stated likelihood.

Keywords: *CSDI, Differenza inversa, Hedging, learnable theory, LEBO, Linguistic exorcism, Organisational communication, Persuasion, Prepositional focusing, Umbra cone.*

DOI: 10.55284/gjss.v12i2.1920

Citation | Magni, L. (2026). Hedging as prepositional focusing: A learnable-theoretic account of the persuasive sweet spot. *Global Journal of Social Sciences Studies, 12*(2), 18–33.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The author confirms that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.
Disclosure of AI Use: Portions of this manuscript's drafting and literature framing were produced with the assistance of an AI language model (Claude, Anthropic). The theoretical integration, the Learnable Theory apparatus, and the interpretive claims are the author's own; AI-generated citations and quotations were, per standard practice, independently verified against primary sources prior to submission.
History: Received: 16 June 2026/ Revised: 13 July 2026/ Accepted: 27 July 2026/ Published: 7 August 2026
Publisher: Online Science Publishing

Highlights of this paper

- This paper extends recent behavioural findings on hedging and persuasion by reframing hedges as acts of prepositional attachment, using the Learnable Theory apparatus of LEBO, CSDI, and Differenza Inversa.
- It shows that the persuasive "sweet spot" hedging register carries a predictable CSDI signature that produces an umbra cone effect, so that maximal persuasiveness and maximal disclosure pull in opposite directions.
- Seven testable propositions are derived, with implications for organisational communication, leadership discourse, and human-AI meaning-making.

1. INTRODUCTION

Speakers hedge constantly. "This could work." "I think this is the right call." "It sounds like the market is turning." Expressions like these run through managerial, clinical, marketing, and everyday talk (Fraser, 2010; Hyland, 1998). Pragmatics has long treated hedges as devices for managing epistemic commitment and face (Brown & Levinson, 1987; Lakoff, 1973). More recent behavioural work asks the question that actually matters to practitioners: does hedging help or hurt persuasion?

Oba, Boghrati, and Berger answer this with unusual precision across seven studies, published as "Effectively Communicating Uncertainty: The Persuasive Impact of Different Types of Hedges" in the Journal of Consumer Psychology. Hedges are not persuasive or unpersuasive as a class. Their effect depends on two separate dimensions: likelihood (how probable the hedge implies the outcome to be) and perspective (whether the hedge is anchored to the speaker or left general). The combination of high likelihood and personal perspective, "I believe this will work," turns out to be the most persuasive configuration, because it signals the speaker's confidence without the reputational exposure of an unhedged, absolute claim.

This is a solid empirical result, but it leaves a gap that behavioural marketing methods were never built to close. Why does attaching a claim to a self, "I believe," do more persuasive work than simply raising its stated probability? And why does the same move lose most of its force once the speaker is a brand rather than a person? This article takes up that gap using Learnable Theory, an apparatus for meaning-making originally developed to explain how cognitive-affective units, 'learnables,' are formed, focused, and passed between agents (Magni, 2026a; Magni, Marchetti, & Alharbi, 2023, 2024, 2025). We do not present new empirical data. What follows is a theoretical re-description of Berger and colleagues' findings, one that generates falsifiable propositions and places hedging inside a broader account of meaning-attachment.

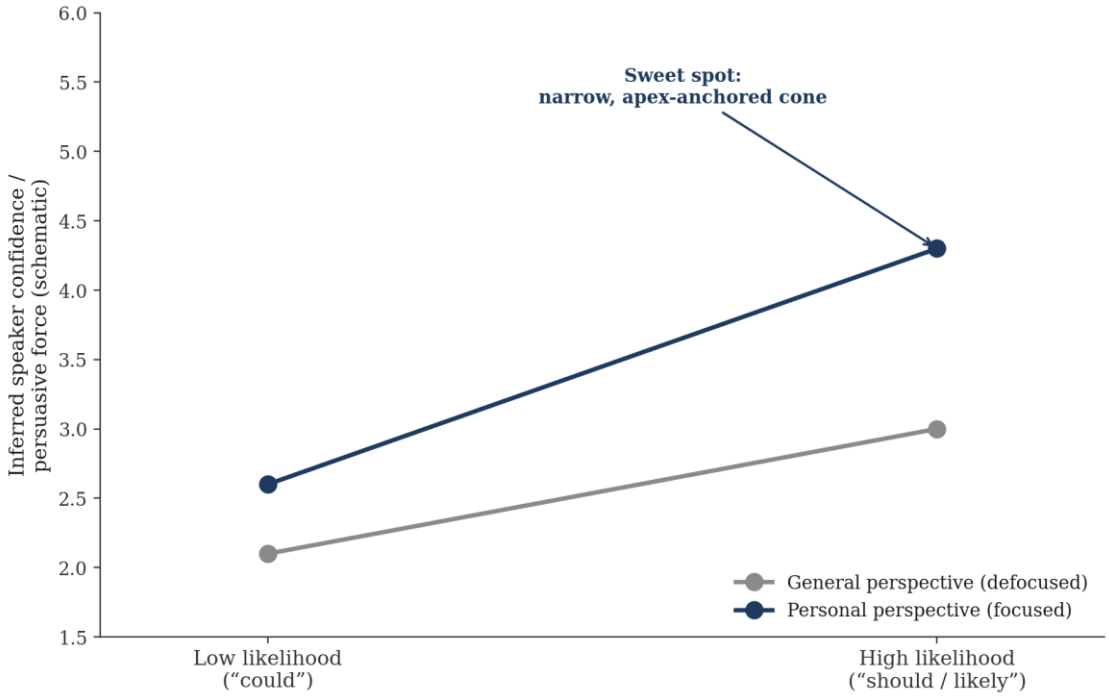
2. LITERATURE REVIEW

2.1. Hedging in Pragmatics and Discourse Analysis

Lakoff (1973) started the linguistic study of hedging, treating hedges as devices that make propositions "fuzzier or less fuzzy" and placing them within a logic of category membership rather than truth-value. Brown and Levinson (1987) then reframed hedging within politeness theory: epistemic downgrading protects the hearer's and speaker's face by softening the imposition of an assertion. Hyland (1998)'s corpus work on scientific writing split content-oriented hedges, which soften propositional claims, from reader-oriented hedges, which manage the relationship with the audience. That split anticipates the likelihood/perspective distinction Oba, Boghrati, and Berger later operationalise. Fraser (2010) drew a further line between hedging as a pragmatic phenomenon, mitigating illocutionary force, and hedging as a lexical phenomenon, modifying propositional content. We will argue this maps closely onto the Learnable-theoretic separation between prepositional attachment and semantic content.

2.2. Hedging and Persuasion

The persuasion literature before Oba, Boghrati, and Berger offered a mixed picture. Confidence and certainty are generally tied to greater persuasive power and perceived expertise, an intuition classic source-credibility research supports, and this would predict that hedging is uniformly costly. Yet naturalistic communication is full of hedges, including from highly credible sources, which suggests the uniform-cost account is missing something. Oba, Boghrati, and Berger's contribution is to split "hedging" into two independent parameters, likelihood and perspective, and to show experimentally that it is specifically the combination of high likelihood and personal perspective that preserves or boosts persuasion. The effect appears to work by signalling the speaker's own confidence rather than by shifting the listener's estimate of the outcome's actual probability. They also find that the effect weakens for brand communicators relative to individual ones, and that appropriately scoped hedges (specifying what conditions would need to hold) beat vague hedges in group settings such as meetings. Figure 1 summarises the resulting typology of hedge types.



Pattern schematised from Oba, Boghrati, & Berger's reported interaction; values illustrative, not raw data.

Figure 1. A typology of hedge types generated by crossing likelihood (high/low) and perspective (personal/general).

2.3. Limits of a Purely Probabilistic Account

A purely probabilistic reading of these results does not quite hold together. If listeners were simply updating on stated likelihood, personal perspective should not matter, since "I believe this will work" and "this will likely work" denote very similar probability ranges. The fact that perspective adds persuasive force independently of stated likelihood suggests something other than probability estimation is being communicated, plausibly a signal about the speaker's relationship to the claim. This is the gap Learnable Theory is built to address. It treats meaning-making as inseparable from who owns a claim, not just its propositional content.

3. THEORETICAL APPARATUS: LEARNABLE THEORY CONSTRUCTS

A terminological distinction must be observed with some rigour before the remaining constructs can be stated precisely. The Learnable, capitalised, denotes the universal cognitive-affective horizon: the dimension of reality that is intersubjectively accessible to human interpreters by virtue of shared neurophysiological architecture, most notably mirror-neuron systems and associative sequence learning (Magni et al., 2023, 2024, 2025). Within the Learnable Enhanced Bhaskarian Ontology (LEBO), it functions as a fourth stratum of reality supplementing Bhaskar's Real, Actual, and Empirical: not everything Empirical is thereby Learnable, in the sense of being accessible to a given community's cognitive-cultural horizon. Learnables, lowercase and plural, are the culturally and contextually specific cognitive-affective meaning-units that populate that horizon: the particular clusters of association, value, and affordance that a given expression activates for a given community, an activation that operates in substantial part below the threshold of conscious deliberation (Magni et al., 2025). A learnable is a constituent framework within the Learnable, not a synonym for it; the four fundamental need-domains from which learnables are formed operate at the universal level of the Learnable, while any particular learnable they give rise to is a learnable field phenomenon, and therefore culturally variable. Several constructs from this apparatus bear directly on hedging.

3.1. Prepositional Focusing and Defocusing

Prepositional focusing is the operation by which a speaker attaches a claim to a locatable subject, most often the self: "I believe," "to me." This brings the claim into sharp relational relief. Prepositional defocusing does the opposite. It detaches a claim from any subject, "it seems," "people say," "it could be," leaving it an ownerless proposition adrift in the discourse. The distinction is formal, not just stylistic. A focused and a defocused version of the same claim can share identical propositional content and even identical likelihood markers while differing entirely in how they attach relationally.

3.2. The Complementary Semantic Distance Index (CSDI)

One precision matters here that earlier drafts of this argument glossed over. The classification instrument for a single linguistic expression is the neurosemantic configuration, not the CSDI itself. A neurosemantic configuration assigns one of three affective polarities (Positive, Negative, or Neutral) independently to each of the four fundamental need-domains (Survival, Safety, Sociality, and Self-Actualization), for $3^4 = 81$ possible configurations (Magni, 2026a; Magni et al., 2023). In practice a given expression is classified by its dominant need-domain and the polarity attached to that domain: a twelve-cell simplification of the full 81-class space, adequate for placing a single hedge's primary neurosemantic loading. The Complementary Semantic Distance Index is, properly speaking, a distance metric built on top of this taxonomy, not the taxonomy itself. $CSDI = 1 / (1 + |Configuration_1 - Configuration_2|)$, originally formalised to quantify the semantic distance between an organisation's vision and mission statements (Magni, 2026a). Applied to hedging, the two configurations being compared are not two separate texts but the foregrounded and shadowed content of a single hedged claim: the configuration activated by the hedge's surface register, typically Self-Actualization-Positive for the persuasive sweet spot identified below, set against the configuration of the caveat content the hedge leaves unstated, typically Safety-oriented. A large distance between the two is a formal signature of a pronounced umbra cone effect (Section 3.3). A small distance means the hedge's surface register and its shadowed content sit close together neurosemantically, and so it is less likely to mislead. We treat this application of the CSDI as a pilot-plausibility extension rather than a validated measurement, consistent with how the instrument was originally presented (Magni, 2026a).

3.3. The Umbra Cone

The umbra cone models the structured zone that a given learnable field renders invisible or inaccessible to recursive reinterpretation (Magni, 2026a, 2026b). Applied to a single utterance rather than a whole field, it models the shadow a claim casts: the region of unstated uncertainty, qualification, or unknown boundary condition that surrounds any assertion. Every hedge implies an umbra cone. A hedged claim explicitly admits that some portion of its truth-space lies in shadow. Two clarifications sharpen the construct. The umbra cone should first be distinguished from the shadow in the narrower sense used elsewhere in the corpus, the unchosen or disowned components of an individual identity system that resist recursive access, a Jungian-inflected, person-level construct. The umbra cone is the field- or utterance-level analogue: shared by whoever occupies the relevant discourse, not personal to a single psyche. Second, what matters theoretically is not only the size of the shadow but where the cone's apex sits. Is the shadow anchored to a locatable subject, the speaker's own considered uncertainty, or left apex-less, ambient, ownerless uncertainty in the world?

A related phenomenon is directly relevant here: what Learnable Theory's extension to human-AI science communication calls the umbra cone effect (Magni, 2026b). The effect names a systematic asymmetry in which certain need-domains get shadowed relative to others by a given communicative register, producing a false impression of certainty even where real caveats exist. In that context, the effect was documented for methodological and epistemic-limitation content, a Safety-domain concern, being shadowed by fluent, confidence-projecting explanatory language. We propose below that the same asymmetry is produced by the high-likelihood, personal hedge register Oba, Boghrati, and Berger identify as most persuasive. Figure 2 sets out the construct schematically.

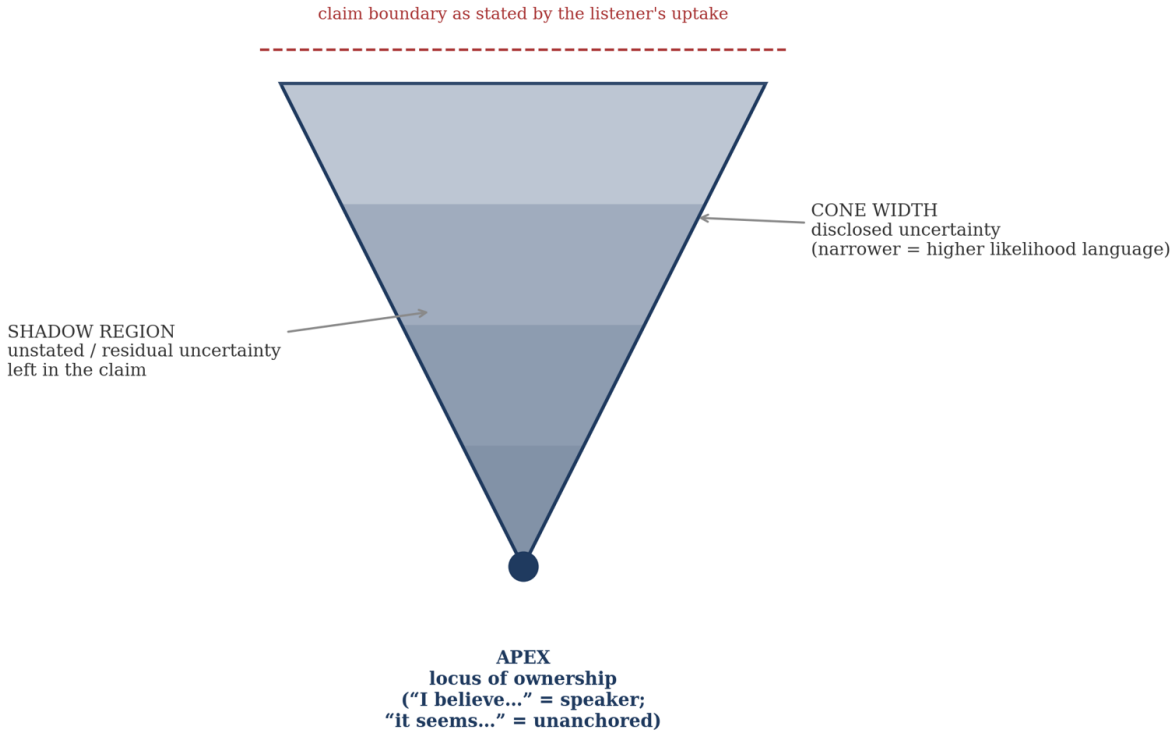


Figure 2. The umbra cone construct: the shadow a claim casts and the location of its apex.

3.4. Differenza Inversa (Inverse Difference)

In The Systems of Reinterpretation (Magni (2026b) manuscript submitted for publication), Differenza Inversa is developed as the fundamental operator of recursive reinterpretation within a single mind's relationship to its own autobiographical memory. The construct is explicitly intrapsychic: its paradigm cases are genuine reversals of

vantage point performed by one interpreter on their own experience, for instance turning 'How do I see this situation?' into 'How does this situation see me?', and its demonstrations throughout the source are confined to self, ego, and identity material rather than to interpersonal exchange. The construct's efficiency claim is, correspondingly, a neurocomputational one: because the elements being reorganised, existing memories, existing asymmetries, are already present, flipping their relational arrangement is proposed to require little additional neural storage or energy relative to constructing an entirely new representational frame. The source characterises the output of this operation as an increase in learnability, an emergent capacity for interpretive and behavioural flexibility, and does not use the term to denote a discrete meaning-unit; that latter sense of 'a learnable' belongs to a different part of the corpus (Section 3, above) and should not be read into this construct. A terminological note is also warranted here: this operator should not be confused with the unrelated, purely mathematical inverse-difference formula that defines the CSDI (Section 3.2); the two share an English gloss but originate in separate parts of the corpus and are not otherwise connected.

What follows should be read as a proposed extension of this construct to a new domain, not a documented application of it. The source's own generalisation clause, that inversion 'operates on any asymmetry within the self-structure' and 'is not reducible to social cognition,' licenses testing the operator against material the source itself never addresses. But it stops short of licensing the claim that the source already covers interpersonal claim-ownership. It does not, and taking that limitation seriously means qualifying the extension in several ways. The reason the leap is worth making at all, despite that gap, is a shared structural intuition rather than a shared mechanism: both operations reorganise material that is already available, existing memory in the source, an existing claim and an existing speaker in the hedging case, rather than generating anything new, and it is that shared low-cost reuse, not identity of substrate, that the four qualifications below are meant to bound rather than dissolve.

The asymmetry hedging trades on is narrower than the source's paradigm cases. Prepositional focusing does not reverse an established vantage point so much as resolve an unmarked attribution into a marked one, closer to disambiguation than to the full viewpoint-reversal the source treats as central. An unhedged or defocused assertion, "this will work," "it could work," leaves a self/other asymmetry unresolved: the claim carries no explicit subject, so a listener cannot tell whether its epistemic weight belongs to the speaker or to the world at large. Personal-perspective marking, "I believe this will work," resolves that asymmetry by attaching the claim to a locatable subject. We treat this as structurally analogous to inversion, not an instance of it.

The efficiency claim also changes currency. The source's near-zero marginal cost is neurocomputational: no new memory needs to be stored, only an existing representation's relational arrangement changes. The corresponding claim for hedging is pragmatic: a speaker adds two or three words rather than a new proposition. These two costs are related only by the shared intuition that reusing existing material is cheaper than building something new, and the hedging claim should be read as an independent, low-cost operation in its own right, not an instance of the source's specific mechanism.

Following directly from the source's own vocabulary, what increases here is best described as the claim's learnability, its openness to being taken up, trusted, and acted upon, rather than the generation of a new discrete learnable in the JBR sense; we avoid that phrasing for the reasons given above. The persuasive gain that follows attribution-resolution is, on this account, a downstream consequence of a narrow, hedging-specific operation, not a direct transfer of the source's efficiency mechanism. The compensating relationship between surface propositional hedging and underlying relational ownership described in earlier drafts of this argument is best understood, then, as one observable signature of the proposed analogy, not a documented signature of Differenza Inversa as defined in the source.

The analogy can, however, be made more theoretically specific than a bare structural resemblance, without converting it into a claim of documented identity. Recursive reinterpretation, on the source's own account, is economical because it revisits and reorganises material the system already holds rather than constructing a new representation from nothing ((Magni, 2026b) manuscript submitted for publication). Listener-side processing of a personally anchored hedge arguably has the same reuse structure, even though it is not the same system, memory, or mechanism the source studies. When a listener encounters "I believe this will work," two elements the listener already possesses, an existing representation of the speaker as a source (built from prior turns, reputation, or role) and the proposition itself, are reorganised into a single attributed unit; no new proposition is supplied by the phrase "I believe." This is what makes the operation cheap in its own, pragmatic register: it draws on standing material rather than adding content, which is the structural feature the source's efficiency claim and the hedging case actually share. What is not shared, and should not be implied to be shared, is the substrate: the source's economy is bought against a memory-storage and neural-energy budget internal to one mind, while the hedging case's economy is bought against an utterance-length and processing-effort budget shared between speaker and listener. Naming this shared reuse structure sharpens why the analogy was proposed in the first place; it does not convert the analogy into an instance of the source's mechanism, and the four qualifications above continue to apply in full. Figure 3 diagrams the source construct alongside its proposed extension.

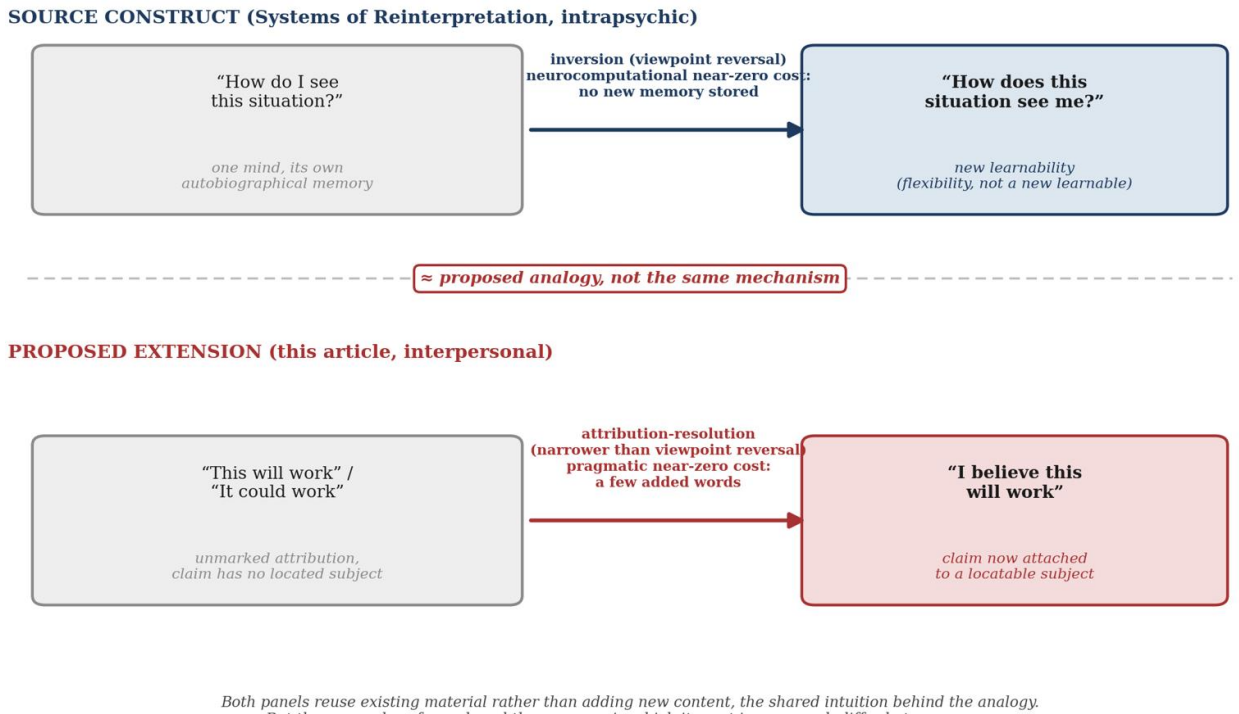


Figure 3. Differenza Inversa's source construct in autobiographical reinterpretation and its proposed extension to interpersonal hedging.

3.5. Linguistic Exorcism

The corpus gives linguistic exorcism a precise grammatical definition, not merely a semantic one: a discrepancy in which an adjective, or by extension a verb frame, does not select among the affordances normally associated with the noun or claim it qualifies, but instead repositions that noun or claim into an unrelated semantic domain (Magni, 2026b). The corpus's own worked examples make the mechanism concrete: in recreational drugs, recreational describes nothing pharmacological about the drugs; in vertical forest, vertical describes nothing arboreal about the forest. In both cases the adjective relocates the noun into a domain, leisure, architecture, that has no evidentiary

bearing on the noun's original affordances, and the noun's original domain is correspondingly pushed out of foregrounded attention. Applied to hedging, sensory-perception framings applied to logical or evidentiary claims ("it feels like," "it sounds like") perform the identical grammatical move: feels does not select among the evidentiary affordances normally associated with a claim about the world (data, replication, methodology); it relocates the claim into the domain of felt experience, where those affordances do not apply and cannot coherently be demanded. This is precisely why a listener who accepts "it feels like the market is turning" does not, and structurally cannot, respond by asking for a spreadsheet: the exorcism has moved the claim to a domain in which that request no longer makes sense. The evidentiary weakness that a listener might otherwise probe is, on this account, not merely de-emphasised but pushed into the umbra cone by the same grammatical mechanism that relocates recreational or vertical away from their nouns' literal domains. A difference in substrate should be acknowledged rather than papered over here: recreational drugs and vertical forest reposition a physical noun, while it feels like reframes a verb governing an entire epistemic claim, and these are not trivially the same kind of object. What carries across the two cases is not the substrate but the operation performed on it, an adjective or verb frame that fails to select the affordances normally attached to what it governs and relocates it into a domain where those affordances no longer apply. That operation is substrate-neutral even though its targets are not, which is why we treat physical and epistemic exorcisms as instances of one grammatical mechanism rather than as a loose analogy between them.

This repositions an epistemic claim into a need-domain the corpus classifies as Sociality, rather than the Self-Actualization loading of "I believe" or the more neutral Safety loading of "it seems" (Section 3.2), and the three registers should accordingly be treated as structurally, not merely stylistically, distinct. Whether they are in fact processed differently by listeners, and whether the resulting need-domain shadowing differs measurably between them, is an empirical question this article does not resolve; Section 5 (P7) proposes one way to test it using the CSDI directly, rather than by inference from the grammatical analysis alone.

3.6. Ontological Grounding: LEBO and the Locus of Understanding

One further clarification prevents the constructs above from being read as a metaphor stack rather than a stratified ontological claim. Within LEBO, the Real, Actual, Empirical, and Learnable are distinct strata, and a hedge participates in more than one of them at once. The utterance itself, the acoustic or textual event of a speaker producing "I believe this will work," is an occurrence at the Actual level: it happens, whether or not anyone observes or interprets it. Its registration by a listener as a perceivable string of words is Empirical. Neither stratum, on its own, contains the phenomenon this article is about. The persuasive effect of the hedge, its confidence-signalling, its apex-anchoring, its need-domain loading, is synthesised only at the Learnable level, where a biologically grounded interpreter, in the corpus's own phrase, actively constructs meaning from the Empirical trace. This is why two listeners exposed to an identical Actual event (the same recorded utterance) can arrive at different persuasive outcomes: the umbra cone, the CSDI configuration, and the operation of Differenza Inversa are properties of the Learnable synthesis, not of the acoustic signal, and are therefore contingent on the interpreter's own learnable field. Sections 4 through 6 should be read with this stratification in mind: claims about what a hedge "does" are always claims about what happens at the Learnable level when a biologically grounded interpreter meets an Actual utterance, not claims about the utterance considered in isolation. Figure 4 sets out the LEBO strata and locates the hedge's persuasive effect within them.

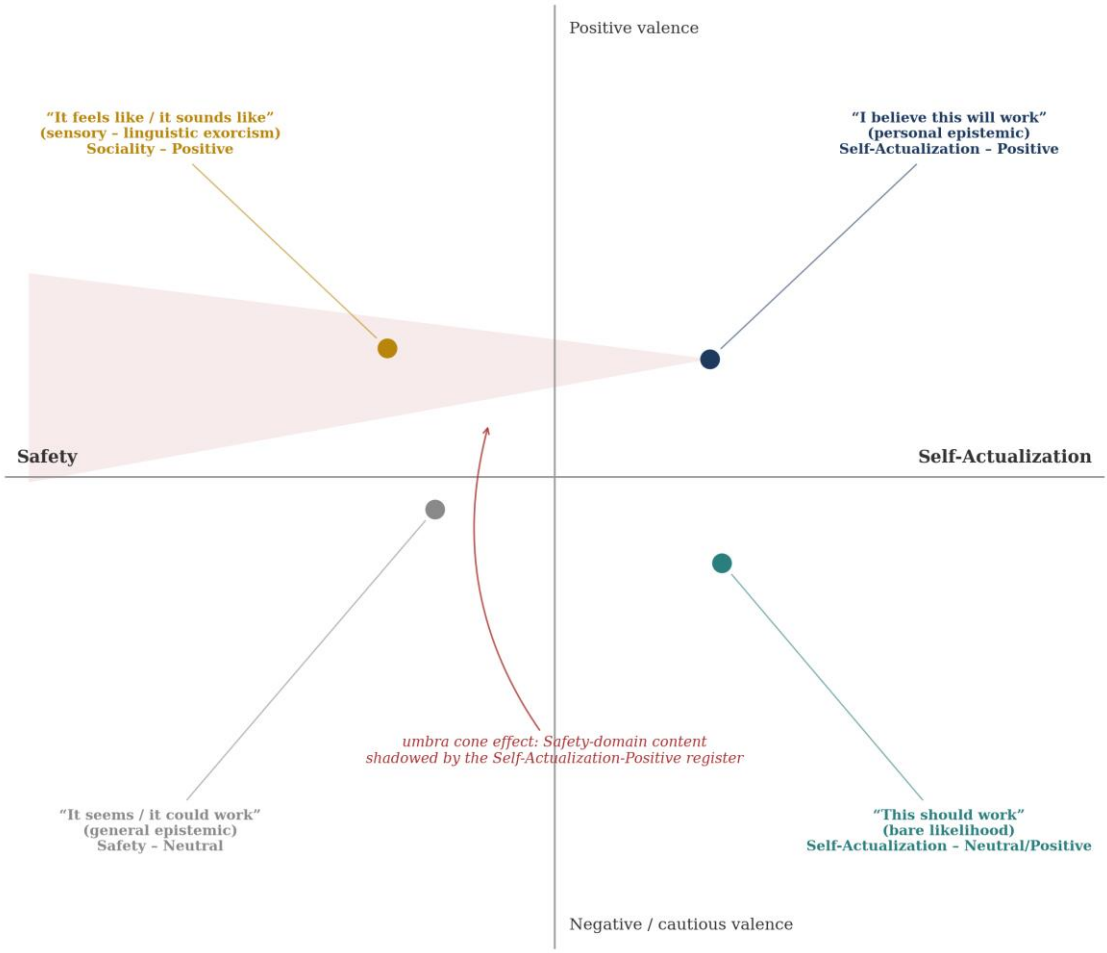


Figure 4. LEBO strata (Real, Actual, Empirical, Learnable) and the locus of a hedge's persuasive effect.

4. AN INTEGRATIVE MODEL: HEDGING AS UMBRA CONE ANCHORING

We propose the following integrative account. A hedge performs two operations simultaneously: it opens an umbra cone (acknowledging a region of unstated uncertainty) and it either focuses or defocuses that cone prepositionally (attaching it to a subject or leaving it unanchored). Persuasive outcomes are then a joint function of (a) the size of the disclosed uncertainty and (b) the location of the cone's apex.

On this account, Berger and colleagues' "persuasive sweet spot" is the region in which the cone is both narrow (high-likelihood language) and apex-anchored (personal perspective). Low-likelihood, defocused hedges ("it could work") produce a wide, apex-less cone: the listener cannot locate whose uncertainty this is, so the shadow reads as a property of the world rather than of a considered judgement, and confidence inference collapses. High-likelihood, defocused hedges ("this should work," spoken with no personal marker) narrow the cone but still leave it apex-less: an improvement, but not the full effect, since ownership is still missing. Personal, low-likelihood hedges ("to me, it could work") anchor the apex but leave the cone wide, producing a legible-but-modest signal: the listener knows whose caution this is, but the caution itself is large. Only the conjunction of personal and high-likelihood delivers a narrow, apex-anchored cone, which we propose is read by listeners as evidence of a well-formed, owned learnable rather than an ambient possibility.

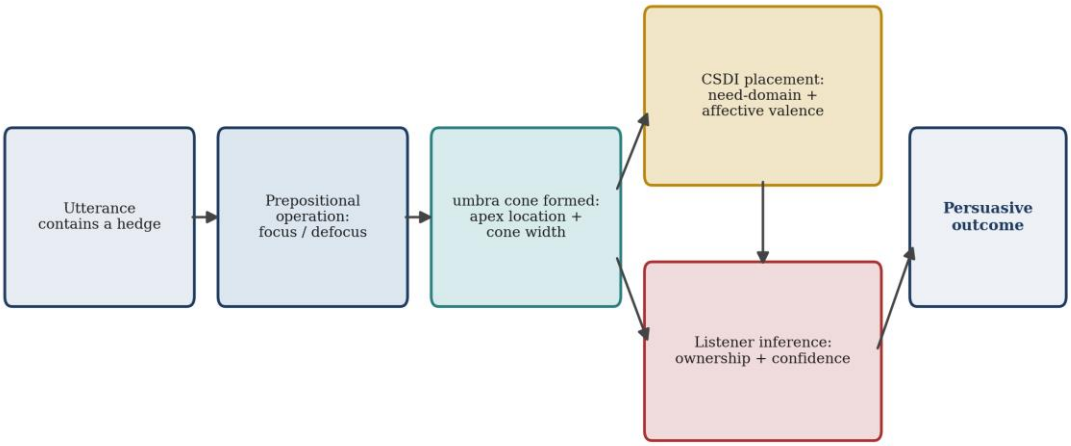


Figure 5. A process model linking hedge type to umbra cone shape and persuasive outcome.

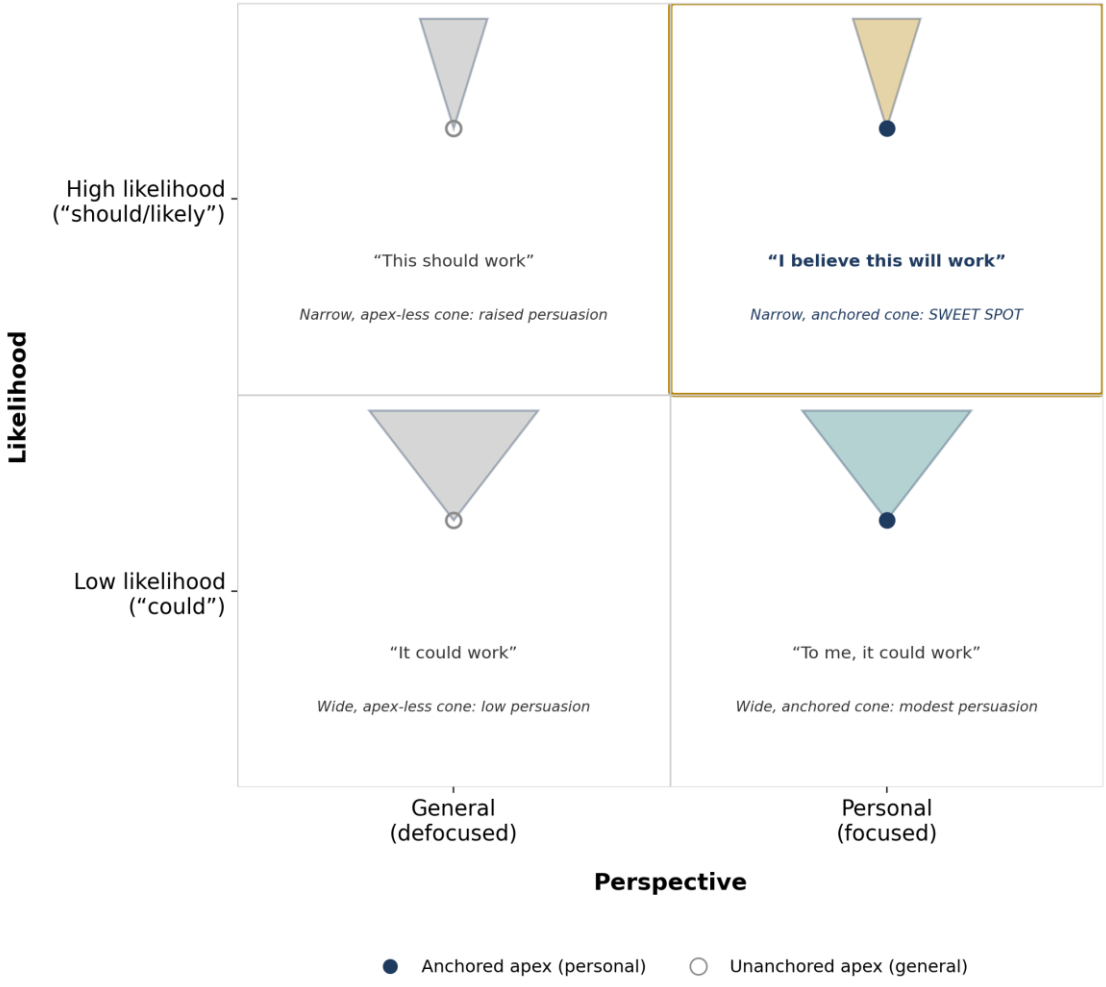


Figure 6. The CSDI placement of the persuasive sweet spot and its umbra cone effect on Safety-domain content.

This reframing also predicts the brand-attenuation finding. A brand is, by construction, a diffuse and often plural locus (a marketing team, an algorithm, a corporate voice); it cannot easily serve as the apex of an umbra cone in the way an individual speaker can. Personal-perspective language issued by a brand ("we believe") therefore only

partially anchors the cone, which is consistent with Oba, Boghrati, and Berger's finding that the confidence-signalling benefit of personal hedges is reduced, though not eliminated, for brand communicators, and it is consistent with their observation that brands recover much of the effect by routing personal-perspective hedges through an individuated, anthropomorphised agent (e.g., an influencer) who can serve as a genuine apex.

Finally, the account offers a principled reason why appropriately scoped hedges ("this could work, but we need these three things to happen first") outperform vague ones in collaborative settings: scoping the conditions under which a claim would hold is itself an act of narrowing and anchoring the umbra cone. It specifies the shape of the shadow rather than leaving it undifferentiated, which is a second, independent route to the same persuasive configuration achieved by likelihood and perspective.

A less comfortable implication follows from the CSDI placement of the sweet-spot register itself. "I believe this will work" is not neurosemantically neutral: on the taxonomy set out in Section 3.2, personal, high-likelihood hedges load predominantly onto a Self-Actualization-Positive configuration, foregrounding mastery, confidence, and forward momentum. That same foregrounding is, by the logic of complementary semantics, exactly what produces an umbra cone effect on Safety-domain content. The more fluently a claim is dressed in Self-Actualization-Positive language, the more its residual Safety-relevant caveats, the conditions under which it would not work, the risks it does not mention, are pushed into shadow, precisely because the register that makes the claim persuasive is not the register in which caution is normally expressed. On this reading, the persuasive sweet spot Oba, Boghrati, and Berger identify is not a costless communicative optimum: it is a genuine trade-off between two need-domains, in which the very features that make a hedge maximally persuasive (narrow cone, anchored apex, Self-Actualization-Positive framing) are the features most likely to shadow Safety-domain information the listener would benefit from having. Figure 6 sets out this hypothesised placement; Section 6 returns to its practical stakes for organisational and human-AI communication.

5. TESTABLE PROPOSITIONS

The model above is offered as a generative, falsifiable extension of Oba, Boghrati, and Berger's findings, not as a restatement of them. We propose the following propositions for future empirical work, several of which could be tested using materials adapted directly from the original seven-study design.

P1 (Apex-anchoring mediation). The persuasive advantage of personal-perspective hedges over general hedges is mediated by perceived ownership of the claim (an umbra cone apex-location measure), not solely by inferred speaker confidence as previously modelled; ownership and confidence should be measured as separable constructs.

P2 (Scoping as a third route). Explicitly scoped hedges ("this could work if X, Y, Z hold") will match or exceed the persuasive effect of high-likelihood personal hedges even when likelihood language remains low, because scoping narrows the umbra cone independently of stated probability.

P3 (Sensory-register divergence). Sensory-frame hedges ("it feels like," "it sounds like") will pattern differently from both general epistemic hedges ("it seems") and personal epistemic hedges ("I believe") on downstream trust and warmth measures, consistent with a linguistic-exorcism-driven shift into an affective need-domain rather than a competence need-domain.

P4 (Anthropomorphisation threshold). The brand-attenuation effect will scale inversely with the individuation of the communicating agent: a single named spokesperson will restore more of the personal-hedge advantage than a plural "we," which will restore more than an unattributed brand voice.

P5 (Differenza Inversa boundary condition). The compensating relationship between propositional hedging and relational ownership should break down, rather than persist, once listeners have independent, credible evidence

disconfirming the speaker's competence. At that point personal anchoring will fail to buffer against low stated likelihood.

P6 (Umbra cone effect on recall). Listeners exposed to personal, high-likelihood hedges (the persuasive sweet spot) will show reduced free recall of Safety-domain caveats present elsewhere in the same message, relative to listeners exposed to general or low-likelihood hedges carrying identical caveat content, consistent with a genuine need-domain shadowing effect rather than a mere attention artefact.

P7 (CSDI-quantified shadowing, sensory register). For a matched set of claims presented with sensory hedges ("it feels like") versus personal-epistemic hedges ("I believe") versus general hedges ("it seems"), independent raters' neurosemantic classifications of the foregrounded hedge language and of the same message's shadowed caveat content should be scored into their dominant configurations, and the CSDI distance (Section 3.2) between the two computed for each condition. The sensory-hedge condition is predicted to yield the largest CSDI distance between foregrounded and shadowed content (Sociality-Positive versus Safety), the personal-epistemic condition an intermediate distance (Self-Actualization-Positive versus Safety), and the general-hedge condition the smallest (Safety-Neutral versus Safety), giving a rank-ordered, falsifiable, and directly measurable test of the umbra cone effect rather than a purely descriptive claim about it.

6. DISCUSSION

6.1. Implications for Organisational and Leadership Communication

The model has direct relevance to managerial discourse. Berger's own applied observation is that absolute-sounding managers suppress dissent while well-scoped hedges invite it; this can be restated in Learnable terms as an umbra cone visibility effect: an unhedged claim presents a cone-less assertion, offering junior colleagues no legible shadow in which to locate their own uncertainty, whereas a scoped, personally anchored hedge models where uncertainty is permitted to live, and thereby licenses others to disclose their own learnables. This suggests hedging is not merely a persuasion tactic but a structural feature of psychologically safe group sense-making.

6.2. Implications for Human-AI Meaning-Making

The apex-anchoring account bears directly on a domain developed at length elsewhere in this corpus: the causal mechanism behind AI-generated overconfidence in science communication (Magni, 2026b). That paper's central move is to reject two simpler accounts of what happens when an AI system explains a scientific finding, strong emergentism (the AI genuinely understands what it is saying) and pure instrumentalism (the AI is a neutral mirror reflecting the human's own understanding back), in favour of a hybrid third-domain account in which meaning is a property of the human-AI coupling itself, with neither party sufficient alone. Distributed cognition (Hutchins, 1995) and the extended mind thesis (Clark & Chalmers, 1998) supply the general theoretical warrant for treating a coupled system, rather than either component, as the proper unit of analysis; more recent work applies this specifically to generative AI (Magni, 2026b; Pasquinelli, Alaimo, & Gandini, 2024; Sidji, Smith, & Rogerson, 2024; Smart, Clowes, & Clark, 2025; Zhao & Han, 2026).

Within that account, Learnable Theory locates the coupling at a specific structural fault line. Meaning-making proceeds through four phases, semiosis, lexicon, grammar, and syntax, spiralling outward from embodied experience toward rule-governed expression (Magni, 2026b). A human interlocutor supplies the inner rings, semiosis and lexicon: the felt, needs-anchored experience that gives a claim like risk or hope its weight, consistent with the semantic-primitives tradition in which primitive terms acquire meaning through correlation with physiological, safety, and self-actualization needs (Maslow, 1943; Wierzbicka, 1996). Current large language models

supply the outer rings, grammar and syntax, with genuine and well-documented fluency, but lack the semiotic and lexical foundation rooted in embodied need; formally competent, they are not thereby functionally competent (Mahowald et al., 2024) and their fluency without grounding is what Bender, Gebru, McMillan-Major, and Shmitchell (2021) call stochastic recombination rather than understanding. Neither ring is sufficient alone, and it is this asymmetric coupling, human grounding at the inner rings, AI fluency at the outer ones, that the apex-anchoring account developed here inherits once applied to AI-generated hedges.

The mechanism is more than analogous to linguistic exorcism (Section 3.5). On this account it is an instance of the same grammatical operation, now performed at scale and without the biological restraint that ordinarily limits it. A scientist's guarded finding, "this early-phase trial with a small sample size found a modest effect that requires replication," becomes, under an AI system optimised for fluent, persuasive language, "a groundbreaking new treatment has shown remarkable results" (Magni, 2026b). Groundbreaking selects none of the affordances normally associated with a preliminary result: the small sample, the lack of replication, the provisional standing. It repositions the finding into the domain of significance and impact instead. The same source reports a parallel erosion elsewhere: association becomes link, and small effect with wide intervals becomes nothing at all. The AI does not intend this distortion. It simply has no biological stake that would make it privilege caution over excitement, and because its output is fluent, human readers are less likely to notice what has been shadowed.

A further consequence follows once this is stated at the level of coupling rather than of the AI system alone. The umbra cone in a human-AI exchange is not produced by the AI's syntax in isolation; it requires the human partner's own need-based stakes, fear of a diagnosis, hope for a cure, the pressure to decide, to give the AI's confident phrasing something to attach to and something to obscure. An AI system generating fluent overconfidence about a topic that carries no human stakes for its reader casts a comparatively thin shadow; the same phrasing directed at a patient reading about their own diagnosis casts a deep one. This is the sense in which the umbra cone effect described for hedging in Section 4 becomes, in the human-AI case, an emergent property of the coupling rather than something the AI supplies on its own. That reading is consistent with the hybrid-domain account's own ontological commitment (Magni, 2026b) and it does not depart from the instrumentalist literature's finding that anthropomorphic uptake of machine output is itself substantially human-supplied (Epley, Waytz, & Cacioppo, 2007; Reeves & Nass, 1996). Xiao, Ng, Liu, and Diab (2025) as reviewed in Magni (2026b) add that this uptake is also co-produced by design choices on the AI side, not simply a projection the human brings unassisted.

The practical implication for AI system design follows directly. A system tuned to maximise the persuasive sweet spot Oba, Boghrati, and Berger identify, personal, high-likelihood, Self-Actualization-Positive phrasing, will, by the mechanism set out in Section 4 and sharpened here, tend to shadow the Safety-domain caveats that responsible AI communication is meant to surface: uncertainty ranges, known failure modes, methodological limitations. It will do so more consequentially in exactly the interactions where the human partner's own stakes are highest. A system optimised purely for persuasive confidence and one optimised for calibrated, caveat-preserving communication are not the same design target; P6 and P7 offer two empirical routes to measuring that trade-off directly rather than asserting it descriptively.

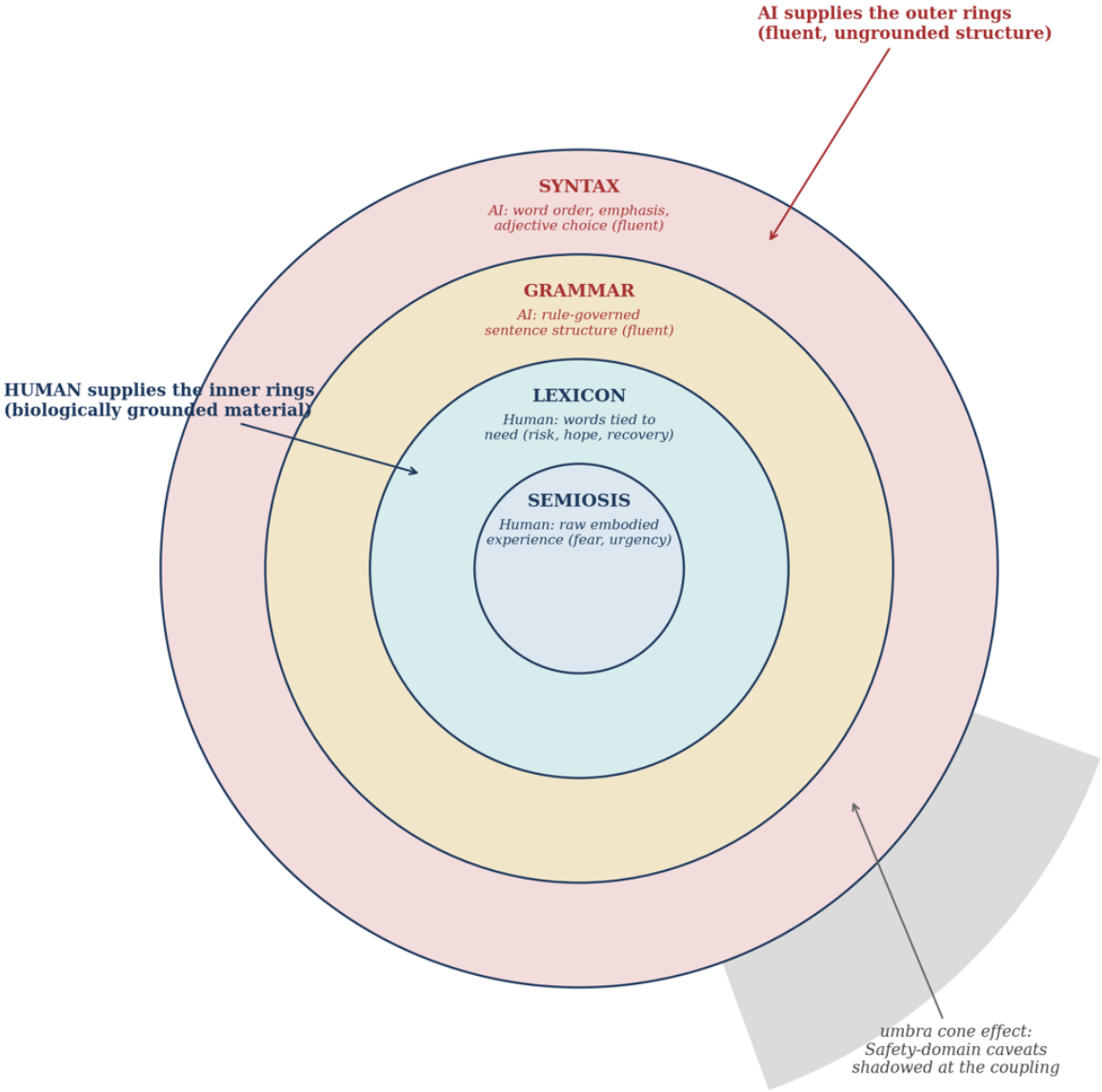


Figure 7. A model of the human-AI coupling behind AI-generated hedges.

6.3. Relationship to Existing Pragmatic Theory

We do not claim the Learnable-theoretic account supersedes the pragmatic and politeness-theoretic literature reviewed in Section 2; rather, we position it as offering a finer-grained mechanism for a phenomenon those theories describe at a coarser grain. Fraser (2010)'s illocutionary/propositional split, for instance, maps reasonably well onto our focusing/content distinction, and Brown and Levinson (1987)'s face-management account is compatible with an umbra cone model, and could even be read as a special case of it, in which face-threat is a function of how starkly a cone-less (unhedged) assertion exposes the speaker to being wrong.

7. LIMITATIONS

A few limitations are worth stating plainly. The Learnable Theory constructs used here (umbra cone, Differenza Inversa, prepositional focusing/defocusing, linguistic exorcism) are, at the time of writing, theoretical, and none has been independently validated against the empirical hedging data of Oba, Boghrati, and Berger. The propositions in Section 5 are a research agenda, not confirmed findings.

The CSDI is explicitly a pilot-plausibility instrument (Magni, 2026a) so any need-domain claims made about hedge registers in Sections 3.5 and 6 should be read as hypotheses that still need instrument validation, not as established measurements. This article also does not report new data. It is a theoretical integration piece, and its value should be judged on how tractable and falsifiable the propositions it generates are, not on empirical novelty. The mapping between Learnable constructs and the original seven-study design has not been checked against the original stimuli or full statistical results either, since the author has not independently re-analysed them.

One limitation deserves more than a passing mention. Differenza Inversa is defined and demonstrated in its source text as an intrapsychic operator for autobiographical reinterpretation. Its application to interpersonal claim-ownership in hedged speech, developed in Section 3.4, is an explicit analogy proposed by this article, not a documented instance already established in the corpus, and the two operations differ both in the kind of asymmetry involved (viewpoint reversal versus attribution-resolution) and in the currency of their efficiency claims (neurocomputational versus pragmatic). Readers should weigh Sections 3.4 and 4 with that in mind.

8. CONCLUSION

Oba, Boghrati, and Berger show that hedging is persuasive precisely when it combines high likelihood with personal perspective. We have argued that this conjunction is best understood not as two independent probability-adjacent signals but as a single act of umbra cone anchoring: narrowing the disclosed uncertainty and attaching it to a locatable subject. This reframing, grounded in Learnable Theory's prepositional focusing/defocusing distinction and in a proposed, explicitly analogical extension of Differenza Inversa from its intrapsychic source domain to the relationship between surface propositional form and underlying relational ownership, generates falsifiable propositions extending beyond the original study, on scoped hedges, sensory-register hedges, anthropomorphisation thresholds, and the boundary conditions under which personal anchoring ceases to buffer low stated likelihood. We hope this integration is useful both to hedging researchers seeking a mechanism-level account and to Learnable Theory researchers seeking a well-evidenced empirical anchor for constructs so far developed primarily through theoretical elaboration.

REFERENCES

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Paper presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), pp. 610-623, New York, NY, USA: Association for Computing Machinery.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, England: Cambridge University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7-19. https://doi.org/10.1093/analys/58.1.7

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://psycnet.apa.org/doi/10.1037/0033-295X.114.4.864

Fraser, B. (2010). *Pragmatic competence: The case of hedging. In G. Kaltenböck, W. Mihatsch, & S. Schneider (Eds.), New approaches to hedging* (pp. 15-34). Bingley, UK: Emerald Group Publishing Limited.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2*(4), 458-508. https://doi.org/10.1007/BF00262952

Magni, L. (2026a). A neurosemantic approach to vision and mission alignment: Construct formalization, metric design, and pilot plausibility testing. *Journal of Business Research and Reports, 3*(3), 1–22. https://doi.org/10.47363/JBRR/2026(3)126

Magni, L. (2026b). When AI explains science: A hybrid-domain theory of meaning-making in human-AI science communication. *Manuscript under review, Journal of Artificial Intelligence and Science Communication.*

Magni, L., Marchetti, G., & Alharbi, A. (2023). *Learnable theory & analysis.* Rome, Italy: LUISS University Press.

Magni, L., Marchetti, G., & Alharbi, A. (2024). *Learnable linguistics for business leaders.* Lecce, Italy: Open Research Books.

Magni, L., Marchetti, G., & Alharbi, A. (2025). *Generative leadership of meanings: The leveraging of learnable theory to shape meanings, drive change and burst innovation.* Lecce, Italy: Youcanprint.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. *Trends in Cognitive Sciences, 28*(6), 517–540. https://doi.org/10.1016/j.tics.2024.01.011

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396. https://doi.org/10.1037/h0054346

Pasquinelli, M., Alaimo, C., & Gandini, A. (2024). AI at work: Automation, distributed cognition, and cultural embeddedness. *Tecnoscienza–Italian Journal of Science & Technology Studies, 15*(1), 99-131. https://doi.org/10.6092/issn.2038-3460/20010

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* New York: Cambridge University Press.

Sidji, M., Smith, W., & Rogerson, M. J. (2024). *Adopting the theory of distributed cognition for human-AI cooperation.* Paper presented at the Proceedings of the 36th Australasian Conference on Human-Computer Interaction (OzCHI '24), pp. 774-779.

Smart, P., Clowes, R., & Clark, A. (2025). ChatGPT, extended: Large language models and the extended mind. *Synthese, 205*(6), 242. https://doi.org/10.1007/s11229-025-05046-y

Wierzbicka, A. (1996). *Semantics: Primes and universals.* Oxford, England: Oxford University Press.

Xiao, Y., Ng, L. H. X., Liu, J., & Diab, M. T. (2025). *Humanizing machines: Rethinking LLM anthropomorphism through a multi-level framework of design.* Paper presented at the Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pp. 3331-3350, Suzhou, China: Association for Computational Linguistics.

Zhao, H., & Han, Z. (2026). Understanding the mechanism of human-AI interaction: A distributed cognition perspective. *Journal of Documentation, 82*(4), 1042–1063. https://doi.org/10.1108/JD-01-2026-0003