A quantum semantic framework for natural language processing

Christopher Agostino[*]

NPC Worldwide, Bloomington, Indiana 47403, USA

Quan Le Thien

Department of Physics, Indiana University, Bloomington, Indiana 47405, USA and
Quantum Science and Engineering Center (QSEC),
Indiana University, Bloomington, Indiana 47405, USA

Molly Apsel

Cognitive Science Program, Indiana University, Bloomington, Indiana 47405, USA and
Department of Psychological and Brain Sciences,
Indiana University, Bloomington, Indiana 47405, USA

Denizhan Pak

Cognitive Science Program, Indiana University, Bloomington, Indiana 47405, USA and
Luddy School of Informatics, Computing, and Engineering,
Indiana University, Bloomington, Indiana 47405, USA

Elina Lesyk

Independent Consultant, Munich 81549, Germany[†]

Ashabari Majumdar

Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA and
NPC Worldwide, Bloomington, Indiana 47403, USA

(Spontaneity Collaboration)

Semantic degeneracy represents a fundamental property of natural language that extends beyond simple polysemy to encompass the combinatorial explosion of potential interpretations that emerges as semantic expressions increase in complexity. In this work, we argue this property imposes fundamental limitations on Large Language Models (LLMs) and other modern NLP systems, precisely because they operate within natural language itself. Using Kolmogorov complexity, we demonstrate that as an expression's complexity grows, the amount of contextual information required to reliably resolve its ambiguity explodes combinatorially. The computational intractability of recovering a single intended meaning for complex or ambiguous text therefore suggests that the classical view that linguistic forms possess intrinsic meaning in and of themselves is conceptually inadequate. We argue instead that meaning is dynamically actualized through an observer-dependent interpretive act, a process whose non-deterministic nature is most appropriately described by a non-classical, quantum-like logic. To test this hypothesis, we conducted a semantic Bell inequality test using diverse LLM agents. Our experiments yielded average CHSH expectation values from 1.2 to 2.8, with several runs producing values (e.g., 2.3-2.4) in significant violation of the classical boundary ($|S| \leq 2$), demonstrating that linguistic interpretation under ambiguity can exhibit non-classical contextuality, consistent with results from human cognition experiments. These results inherently imply that classical frequentist-based analytical approaches for natural language are necessarily lossy. Instead, we propose that Bayesian-style repeated sampling approaches can provide more practically useful and appropriate characterizations of linguistic meaning in context.

I. INTRODUCTION

Prior to the deep-learning revolution of the 2010s, there was a prolonged AI winter spanning from the mid-1980s to the early 2000s. During this period, most researchers in natural language processing and artificial intelligence contended—based on the promises of AI that were unmet in the 1960s and 1970s—that computers on their own would never be capable of producing any kind of general intelligence or simulating the phenomenon that is described as natural language. AI researchers that tried often found themselves drowning in heuristics that broke down with edge cases, usually abandoning the tasks for more readily-solvable problems. See Dreyfus' *What Computers Still Can't Do* [1] to better understand the logical and semantic limitations of artificial reason perceived by researchers prior to the successes of tools making extensive use of large neural networks like Google Translate and others that followed (e.g., transformers,

---

[*] info@npcworldwi.de; http://www.cjagostino.com
[†] info@elinalesyk.com

large language models). Such large neural networks that have now become commonplace for consumer products are often termed Distributional Semantic Models (DSMs) [2–6]. DSMs, in effect, infer relevance and meaning from statistical co-occurrences, and they have enabled a variety of highly practical natural language processing (NLP) applications including topic models, sentiment classification, and large language models [7–10]. By construction, DSMs like BERT-style topic models presuppose that documents possess singular, intrinsic semantic compositions [e.g., early bag-of-words models; see 2, for contrast with dynamic views].

Because of the practical successes of DSMs in many domains, some researchers have forgotten or not considered many of the initial criticisms of artificial reason that to this day remain sound, instead assuming that these limitations too will be solved by more compute or more data. However, consider the following salient problems in the current field of Natural Language Processing:

1. DSM-powered approaches still exhibit limitations particularly when dealing with complex, ambiguous, or context-rich texts [11, 12].

2. The apparent lack of much progress seen in the latest generation of frontier large language models (LLMs)—compared to the leaps seen in prior model releases—appears to have revealed a fundamental barrier in the reliability of semantic problem solving that more compute/data have not alleviated (e.g., GPT-4.5, Llama 4, Gemini 2.5, Claude 3.7 Sonnet). Likewise, many frontier researchers have shifted away from prioritizing progress in LLMs in favor of more dynamically capable models (e.g. JEPA [13], MAMBA [14], CTM [15], or modifications to LLMs that allow them to self-improve [16, 17]).

These issues and the evolving research landscape suggest that the challenges faced by DSM-based approaches may stem from a more fundamental issue in linguistics: the problem of semantic degeneracy. This concept, extending beyond simple polysemy, refers to the inherent multiplicity of potential interpretations that arise when processing complex linguistic expressions [18–20].

Indeed, the concept intuitively makes sense: empirical observation reveals to any observer that natural language meaning is not fixed or absolute [2, 21] and interpretation itself is radically context-dependent [22–24]. Thus, any semantic meaning realized by an agent interpreting a natural language expression depends crucially on a combinatorially explosive set of potential factors. These factors include but are not limited to:

- the surrounding sentential and discourse context [25, 26]

- the agent's current attentional focus and task demands which can selectively highlight specific conceptual features, [27–32]

- the agent's background knowledge, cultural milieu [33–35],

- transient psychological states [36] [1]

- the specific language being used, as different languages carve up semantic space differently [37].

This profound context dependence implies that meaning is not merely decoded from the specific words in and of themselves independently but that meaning is actively constructed or realized by an interpretive agent within a specific situation—[38–40]. The same expression presented to different agents—or to the same agent under different conditions—can yield different interpretations [41, 42]. Critically, this highlights the *observer-dependent* nature of the actualization of semantic meaning through direct interpretation, providing a quantum mechanical analog to an observable acting upon a state which we will explore the implications of more thoroughly in Section III. This kind of co-creative process has been posited to be mediated by a process called Relevance Realization (RR)—a core cognitive capacity enabling individuals to efficiently navigate vast semantic spaces by employing context-sensitive attentional mechanisms to identify relevant information and filter out the irrelevant [43–45]. Crucially, RR is non-algorithmic in the standard computational sense [45], and as such is more adept at handling challenges like problem framing and indefinite search spaces that are intractable for purely formal systems operating in "small worlds" [46]. This embodied, observer-dependent, and contextually situated view of meaning construction challenges any notion of semantic expressions possessing intrinsic, context- or observer-independent meaning, aligning with dynamic models of semantic memory [22, 27, 38, 41, 47–50].

The inherent uncertainties and deep context-dependencies of this dynamic process suggest that classical probabilistic and logical frameworks are insufficient. Consequently, researchers have turned to non-classical frameworks—such as those employing principles from quantum theory—to find more suitable mathematical tools. Such approaches have been adopted to model a wide range of cognitive phenomena, including concept combinations, decision-making, and memory [51–57]. The utility of these quantum-inspired models is not merely theoretical but is demonstrated

————

[1] While many consider LLMs simply statistical machines that are completely disconnected from the material reality of the physical world, it is certainly possible that real world physical variations in the environment that they inhabit or other random phenomena like cosmic rays could induce transient states. Additionally, when a model like Claude from Anthropic possesses knowledge of the time of the year, its having a French name and the training and system prompting of Claude reinforce a French mindset onto the model to the point that its response patterns and willingness to put in effort vary depending on whether the time of year coincides with French holidays.

in empirical studies. For instance, Aerts *et al.* [58] identified quantum-like contextuality effects when analyzing concept co-occurrence statistics for the "Pet-Fish problem" on the World-Wide Web, showing that meaning construction at this scale deviates from classical probabilistic assumptions. In a similar vein, other work has applied Bell's [59] theorem to human cognition, uncovering non-classical correlations that violate classical bounds in both information retrieval judgments [60] and cognitive decision-making tasks [61–63].

More recent work provides a crucial conceptual refinement by distinguishing between two types of contextual influence. In experiments on facial trait judgments, Bruza *et al.* [64] delineate between *context-sensitivity*, a standard causal influence of context, and true *contextuality*. The latter is defined as an acausal form of context dependence where a property may be genuinely indeterminate prior to measurement. They argue that if a cognitive phenomenon is found to be contextual, the underlying cognitive properties do not possess well-defined, pre-existing values. Instead, the property is actualized in the moment of judgment—a phenomenon that non-classical models are uniquely equipped to formalize.

Given the state of affairs, we aim in this work to accomplish two primary tasks: (1) identify the combinatorial problems that have stalled the apparent progress of frontier LLMs and (2) provide a practical path forward for understanding and studying natural language using a non-classical framework. To this end, in Section II, we provide an information-theoretic exploration of the role of semantic degeneracy in single-turn problem solving tasks. Then, in Section III, we formulate a quantum semantic theoretical framework for the act of interpretation. Following this, we detail our experimental methodology in Section IV to test whether natural language interpretation may exhibit non-classical behavior, leveraging LLMs as interpretative stand-ins. Section V presents our findings, and Section VI discusses their broader implications for the future of computational linguistics and our understanding of cognitive science.

II. KOLMOGOROV COMPLEXITY, SEMANTIC DEGENERACY, AND THE CHALLENGE OF INTERPRETATION

The concept of Kolmogorov complexity (KC) [65] provides a powerful lens through which to understand the fundamental limitations of natural language interpretation, particularly pertinent with respect to LLMs which are expected to reliably solve problems for users. Kolmogorov Complexity, $K(s)$, of a string s is formally defined as the length of the shortest computer program (in a fixed universal description language) that produces s as output. While KC strictly applies to finite strings, its underlying principle—the quantity of information required for minimal description—can be extended to the domain of semantics. For some semantic expression S_E We can

conceptualize $K(M(S_E))$ as the minimum number of bits required to unambiguously specify the *intended meaning* $M(S_E)$ of a given semantic expression S_E. This specification must capture not only the identities of the concepts involved but also their precise contextual nuances and the intricate web of relationships binding them into a coherent whole. For instance, the KC of specifying a single, coherent interpretation of 'I just went to the animal shelter and I brought a dog home' is low, whereas for a passage from a work as complicated and inter-connected as James Joyce's *Finnegans Wake* e.g.,

> And what sensitive coin I'd be possessed of at Latouche's, begor, I'd sink it sumtotal, every dolly farting, in vestments of subdominal poteen at prime cost and I bait you my chancey oldcoat against the whole ounce you half on your backboard (if madamaud strips mesdamines may cold strafe illglandsl) that I'm the gogetter that'd make it pay like cash registers as sure as there's a pot on a pole. And, what with one man's fish and a dozen men's poissons, sowing my wild plums to reap ripe plentihorns mead, lashings of erbole and hydromel and bragget, I'd come out with my magic fluke in close time, fair, free and frolicky, zooming tophole on the mart as a factor.

the KC would be extraordinarily high to the extent that the amount of constraints from context needed to disambiguate the 'intended meaning' of the passage might require a program that is orders of magnitude longer than the original expression, reflecting the vast amount of information needed to resolve its multifaceted ambiguities into one particular reading.

Thus, the informational burden, $K(M(S_E))$, necessary to constrain interpretations to those closely aligned with an author's original intention, scales dramatically with the number and interconnectedness of the concepts and relationships within an expression S_E. While the constituent concepts c_i form the foundation, each requires a number of bits for contextual instantiation. This linear relationship might be tractable on its own, but it is the relationships $r_{i,j}$ between concepts that contribute to actualizing the intended meaning, and specifying the inter-relational constraints requires specifying an $O(N_{Concepts}^2)$ number of bits to have a shot at disambiguating potential interpretations and reproducing the intended one. Because of this, the KC for an expression can be written as an inequality with a lower bound based on the number

$$K(M(S_E)) \geq \sum_i^N c_i + \sum_{i,j}^N \cdot c_{i,j}^2 \qquad (1)$$

As can be seen, $K(M(S_E))$ increases at least superlinearly as the complexity of an expression increases. Each of the $K(M(S_E))$ bits represents an informational specificity or a semantic choice point (e.g., deciding to interpret 'bat' as an animal instead of a wooden stick used for

baseball) that an interpreting agent must successfully reconstruct. Crucially, at each such decision point, there is a potential opportunity for degenerate solutions as many words have multiple meanings, and so we specify some degeneracy per bit d_b (Note that this essentially serves as an error rate, but we employ it in this way to maintan the conceptual notion). Consequently, the probability of an interpreter correctly actualizing *all* $K(M(S_E))$ bits to arrive at the precise intended meaning is the product of the probabilities of getting each bit right,

$$P(\text{perfect interpretation}) = \frac{1}{N!} \prod_{k=1}^{K(M(S_E))} (\frac{1}{d}) \quad (2)$$

If we assume an average degeneracy per bit, $\bar{d}_b$

$$P(\text{perfect interpretation}) \approx \frac{1}{N!}(\frac{1}{\bar{d}_b})^{K(M(S_E))} \quad (3)$$

a relationship which we illustrate graphically in Figure 1. As $K(M(S_E))$ grows in Figure 1, the probability of a perfect (or otherwise similar enough) interpretation diminishes exponentially, rapidly approaching zero for expressions of moderate complexity. This result provides a clear demonstration of semantic degeneracy in action: the combinatorial explosion of alternative, plausible interpretations surrounding any expression S_E. At this point, it is worth mentioning that this situation bears an analogy to concepts in statistical mechanics, wherein S_E is like an ensemble of micro-states. An error in inferring even one bit (a constraint on a degree of freedom) leads to a different microstate, and, with the high dimensionality ($K(M(S_E))$), it becomes overwhelmingly probable that an interpreter will rarely, if ever, reproduce the specific set of micro-states (which we will argue in Section III are themselves unknowable a priori) that make up the ensemble, resulting in high 'semantic entropy'. This KC-based analysis highlights a fundamental limitation for NLP systems and explains the persistent difficulties in LLM-assisted tasks requiring deep, unambiguous understanding or translation of semantically degenerate expressions: the LLM generates a *plausible* meaning—one of many accessible microstates—but almost never the *singularly intended* one. This result on its own highlights the need to move beyond training artificially intelligent systems that prioritize single-shot response success and to prioritize research on alternative models that can successfully simulate natural language in the way that LLMs have while also being able to dynamically update and adapt itself accordingly. It is our hope that these lessons here coupled with the quantum semantic approach can provide a clearer fundamental basis upon which future methods and models can be trained, tested, and evaluated.

III. A THEORY OF QUANTUM SEMANTICS

The foundational premise of this work is that meaning is not an intrinsic, static property of a semantic ex-
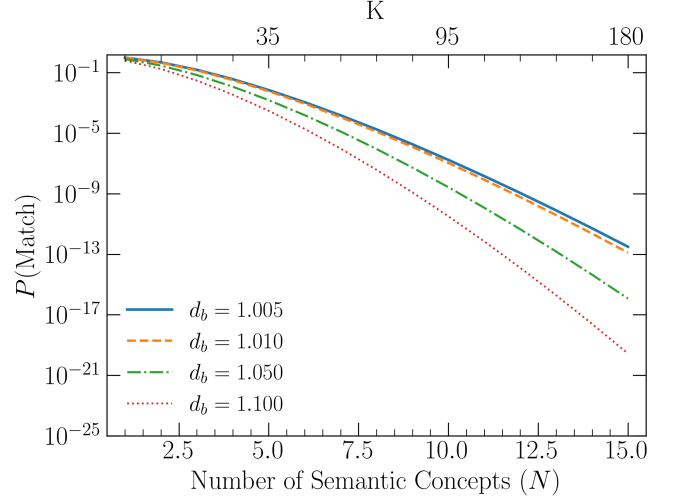

FIG. 1. Probability of perfect semantic interpretation versus the number of core semantic concepts ($N_{concepts}$) in an expression, illustrating the impact of superlinear growth in semantic complexity. The total semantic bits (K, shown on the top x-axis) are modeled as the sum of information required for individual concepts and their pairwise relationships, here assuming $c_{concept} = 5$ bits per concept and $c_{relationship} = 1$ bit per relationship. Different curves represent varying average probabilities of error per bit (p_e). The superlinear increase in K with $N_{concepts}$ leads to a dramatically faster exponential decrease in the probability of flawless interpretation, underscoring the profound challenge posed by semantic degeneracy as inter-concept dependencies proliferate.

pression, but rather an emergent phenomenon actualized through the dynamic interaction between the expression and an interpretive agent situated within a specific context. As described, this framework naturally challenges the assumption of realism which historically arose in classical physics. To formally model this observer-dependent and contextual nature of meaning, we propose a quantum semantic framework which mirrors the difference between quantum and classical physics. Semantic expressions are separated from syntax and instead treated as observables, mirroring how physical measurement outcomes in quantum systems are detached from the system as opposed to the realism as assumed in classical system. We hope that the processes of interpretation and meaning actualization are elucidated clearer under this quantum logical structure.

We begin by positing that a semantic expression, denoted S_E, does not possess a pre-defined, inherent meaning. Instead, it functions as a symbol that affords a spectrum of possible interpretations when engaged by an agent. To represent this capacity for interaction, S_E is associated with a state vector $|\psi_{S_E}\rangle$ in a complex Hilbert space $\mathcal{H}_S$, the semantic state space. This vector is a linear superposition of basis states, $|\psi_{S_E}\rangle = \sum_i c_i |e_i\rangle$. A crucial aspect of this proposal is that the basis states $\{|e_i\rangle\}$ themselves are not assumed to be known or fixed a

priori, nor are they necessarily interpretable as a universal set of predefined semantic primitives. Instead, they may represent abstract dimensions of potential semantic differentiation that only become operationally relevant or partially discernible through specific interpretive acts (measurements), where $\{|e_i\rangle\}$ is the set of eigenbasis of an observable $\hat{\text{も}}$—here denoted by a Japanese character which is pronounced 'mo'— as posited by the agent performing the measurement, i.e. interpreting the syntax. The Hilbert space $\mathcal{H}_S$ is thus a formal construct whose basis is defined by the set of all possible ways an expression can be distinctively engaged with, rather than by a pre-enumerated list of fundamental meanings. The coefficients c_i are complex numbers ($c_i \in \mathbb{C}$), determined by how the semantic expression S_E can be decomposed as a superposition of the semantic bases $|e_i\rangle$. While the squared moduli $|c_i|^2$ relate to probabilities of outcomes upon measurement with respect to a chosen observable, the complex phase of these coefficients holds additional information with no direct analogue in classical probabilistic models, and so it is impossible to derive any analogical particular meaning from these coefficients as one might consider weights in a language model. Indeed, the complex nature is fundamental for modeling potential interference and entanglement effects in semantic processing. The full interpretive significance of c_i is only realized through the associated interpretive axis $|e_i\rangle$ set by an agent, i.e. there is no notion of pre-existing and isolated probabilities of meaning.

An agent A engages S_E through an interpretive observable, $\hat{\text{も}}_A(t)$, which is dynamically constituted from its semantic memory (comprising goals, persona, knowledge, attentional state) as activated by the current context $C_A(t)$. The act of interpretation to ascertain a specific semantic aspect is represented by applying a Hermitian operator, denoted $\hat{\text{も}}_A(t)$ to $|\psi_{S_E}\rangle$. This operator $\hat{\text{も}}_A(t)$ embodies the specific semantic probe. The eigenvalues $\{m_i\}$ of $\hat{\text{も}}_A(t)$ encompass all possible outcomes of agent A's semantic measurement, specifically they represent all possible distinct interpretations actualized by the agent, even though they cannot be explicitly enumerated. The probability of actualizing interpretation m_i is thus given by

$$P(m_i) = |\langle c_i|\psi_{S_E}\rangle|^2 = \langle\psi_{S_E}|\hat{P}_i|\psi_{S_E}\rangle \quad (4)$$

where $\hat{P}_i = |e_i\rangle\langle e_i|$ is the projection operator onto the eigenspace of m_i. This interaction is analogous to a quantum measurement.

With such formulation, semantic measurements by different agents can now possess non-commutativity, similar to quantum observables. If two distinct interpretive operations, $\hat{\text{も}}_1$ and $\hat{\text{も}}_2$, do not commute, i.e. $[\hat{\text{も}}_1, \hat{\text{も}}_2] \neq 0$, then the semantic aspects they probe cannot generally possess simultaneous, definite, pre-existing values.

The dynamics of meaning actualization extends to the time evolution of the interpretive process. The agent's in-

terpretive observable $\hat{\text{も}}(t)$ and the semantic expression $|\psi_{S_E}(t)\rangle$ are generally not static. While $\hat{\text{も}}(t)$ as an observable can vary explicitly in time, $|\psi_{S_E}(t)\rangle$'s dynamics can, in analogy to quantum mechanics, be governed by a semantic Schrödinger equation:

$$i\hbar_{sem}\frac{\partial}{\partial t}|\psi_{S_E}(t)\rangle = \hat{H}_{sem}(t)|\psi_{S_E}(t)\rangle \quad (5)$$

where the semantic Hamiltonian, $\hat{H}_{sem}(t)$, generates this evolution, encapsulating drivers such as shifts in context $C_A(t)$, sequential information processing from S_E, and the agent's internal cognitive dynamics. The constant $\hbar_{sem}$ sets the scale for these semantic dynamics where quantum coherence between the bases $|e\rangle_i$ is important. This formalism allows for agents' interpretive engagement evolution. The dependence of $\hat{H}_{sem}(t)$ on t allows for modeling changes in this evolution as the agent navigates different semantic loci. In this work, however, we set aside the time evolution component of our framework and focus more on certifying the quantum state nature of the semantic expression S_E through sampling exploration by way of a Semantic Bell Test, thus adapting a logic that has previously been employed in cognitive psychological experiments to reveal non-classical contextuality in human judgements across diverse domains, including decision-making, information retrieval, and assessments of concepts [60–63].

IV. EXPERIMENTAL DESIGN

This section outlines the experimental methodology designed to test for non-classical correlations in semantic interpretation, analogous to a CHSH-type Bell test in quantum physics [59, 66]. In particular, this experiment focuses on how Large Language Model (LLM) agents interpret ambiguous word pairs within simple sentence structures, under varying contextual (persona-based) settings.

A. LLMs as observers

In this work, LLM agents serve as the "observers" in this semantic Bell test. To mitigate potential model-specific biases and enhance the robustness of our findings, each agent instantiation is randomly selected from a predefined pool of diverse, state-of-the-art foundation models and providers. This pool includes models such as variants of Gemini (e.g., 'gemini-1.5-flash', 'gemini-2.0-flash-lite', 'gemini-2.0-flash'), Anthropic's Claude series (e.g., 'claude-3-5-sonnet-latest', 'claude-3-5-haiku-latest', and 'claude-3-7-sonnet-latest'), DeepSeek's 'deepseek-chat', as well as various models from OpenAI (e.g., 'gpt-4o', 'gpt-4o-mini', 'gpt-4.1-mini', 'gpt-4.1-nano', 'gpt-4.1-nano'). This approach aligns with recommendations for

multi-model triangulation to ensure fairness and generalizability of results [67–70] and to ensure that any potential correlations we find are not restricted only to a single model's weights. For each experimental trial, two primary base personas, "Alice" and "Bob" are generated. These personas are characterized by randomly assigned attributes such as age (e.g., 25-70 years) and location (e.g., Bloomington, IN; Detroit, MI), which implicitly defines their primary language (English, in this setup). These attributes inform the agent's base semantic memory profile for the trial.

It should be noted here that, paradoxically, while LLMs exhibit limitations on complex tasks because of semantic degeneracy, their internal mechanisms—specifically their attention architectures—function as a kind of black box which—like the brain—collapses the state of potential interpretations into a specific one that they use when responding to user inquiries. Although the underlying mechanism is of course distinct from the biological and cognitive underpinnings of human linguistic interpretation, the similarity of the two 'observing' a specific state suggests that LLMs do indeed effectively reproduce this function of language understanding and cognition. Thus, these models can serve as experimental interpretative probes in natural language tasks. In addition, it has also been demonstrated that LLMs can generate responses that can to first order mimic human linguistic behavior in various contexts (e.g. example surveys [71–73][2] Kitadai et al. [71] also note the power of personas in improving the verisimilitude of the responses, an important facet that we make use of as well. Thus, we argue, that one can and should use LLMs to probe the statistical patterns of semantic interpretations under diverse conditions (e.g. persona and context variations). By observing how LLMs grapple with semantic ambiguity and context-dependent meaning, we can gain insights into the mechanisms of interpretation and the types of computational strategies that are more or less successful in navigating tasks with high semantic degeneracy [74, 75].

B. Bell Test

The core of our semantic Bell test involves presenting sentences containing two ambiguous words to LLM agents. The agents are then tasked with choosing a singular, unambiguous meanings for each of the words in the pair. Stimuli are constructed by embedding ambiguous word pairs (e.g., "trunk" with meanings 'A' for storage/tree vs. "bow" with meanings 'A' for ship front/knot) into simple sentence templates (e.g., "The word1 was settled near the word2"). Four distinct interpretive settings are defined for agents Alice (A, A') and Bob (B, B'), created by providing their base personas with additional, distinct short textual prompts designed to prime semantically orthogonal different contextual perspectives (e.g., "You are a surgeon..." vs. "You are a bus driver..."). For each of these four settings, the corresponding LLM agent provides a singular simultaneous interpretation for both ambiguous words. Alice and Bob are not shown the definitions under consideration so as to avoid biasing them or limiting the semantic search space, and so a separate LLM call is required to then determine if each interpretation aligns with predefined meaning 'α' or 'β', triggering a re-interpretation if the choice is unclear (e.g. if the two options for meaning for the word 'chair' are 'leader of a group' or 'furniture to sit on' and the interpretation says 'furniture or leader') or outside these options (e.g. for the chair example if it decides the sentence is referring to an execution by 'electric chair', it would be considered outside of the definitional scope and a re-interpretation is carried out).[3] These classified interpretations are mapped to numerical values (A $\to$ +1, B $\to$ −1), yielding a 2-element outcome vector for each setting. Finally, the outcome vectors ($\hat{\mathfrak{t}}_A$, $\hat{\mathfrak{t}}_{A'}$, $\hat{\mathfrak{t}}_B$, $\hat{\mathfrak{t}}_{B'}$) are normalized, and expectation values $E(XY)$ are calculated as the average dot product of corresponding pairs of (α, β), which are then used to compute the CHSH S-value $S = E(A, B) - E(A, B') + E(A', B) + E(A', B')$. However, a critical assumption in this standard CHSH formulation is that of marginal consistency, also known as 'no-signaling' [76, 77], which posits that the marginal probability for one agent's outcome is independent of the measurement setting of the other agent. As Dzhafarov et al. [76] have extensively discussed, this assumption is often violated in cognitive and behavioral experiments. Such violations, termed 'inconsistent connectedness' can complicate the interpretation of Bell-type inequalities. Therefore, while a result of $|S| > 2$ suggests a violation of local realism, it must be interpreted with caution, as the excess correlation could potentially arise from these direct contextual influences rather than true contextuality.

All agent definitions and LLM response handling are carried out using the open-source `npcpy` package[4].

---

[2] The last of these references, Tjuatja et al. [73], actually describes the performance of LLMs in this regard as poor as they note they are subject to perturbations that humans in a similar survey would not be as strongly affected. We note here that their findings here fit neatly within the framework of this work and that, if one additionally considers the role of relevance realization for the human subject (context-rich by way of physical embeddedness) versus the LLM (context-poor and disembodied), it appears clear that the main 'problem' with the LLM responses in their experiment was due to a drastic contextual disadvantage.

---

[3] It is possible to imagine this experiment can be carried out with word choice sets with more than 2 meanings, and we plan to explore that in future work as this would allow us to further explore the qubit-style logic we might consider in more practical applications, but that is outside of the scope of this work.

[4] https://github.com/NPC-Worldwide/npcpy

Experiment	N Trials	|S|
1	5	2.8
2	5	1.2
3	10	2.0
4	10	2.44
5	20	2.32
6	20	2.0
7	50	2.33
8	200	1.83

TABLE I. Calculated $|S|$ values for the CHSH inequality from different experimental runs, with slight variations between the persona configurations (age, location) as well as the total number of trials (N) per experiment. Values $|S| > 2$ indicate a violation of the classical Bell-CHSH inequality.

V. RESULTS

In this short section, we highlight and discuss the results of our Semantic Bell test.

In our experimentation, we conducted 8 experiments where we varied the number of trials in the individual experiments from 5-200 and instantiated different persona combinations, and we show the various $|S|$ values for the experiments in Table I. For illustrative purposes, we show in Figure 2 the evolution of the $|S|$ value's evolution across trial number for experiment #7 with $N = 50$ trials. To be clear, there is not any particular meaning to be derived from the variation in $|S|$ as it progresses with trial number as the $|S|$ value itself ought to be derived from an infinite number of trials. This illustration simply serves as a visual aid to show the stabilization of the $|S|$ value for a non-negligible number of trials.

Importantly in these results, we find a rich variety of correlation structure ranging from classical behavior to non-classical quantum logical. Theoretically, the upper bound for quantum systems in the Bell experiment is $|S| \leq 2\sqrt{2}$. It is interesting then that one of our experiments (albeit $N = 5$), the $|S|$ value reached 2.8, providing an empirical upper bound for these experiments that can be consistent with quantum computational frameworks.

VI. DISCUSSION

The arguments presented in this work, concerning the fundamental limitations imposed by semantic degeneracy and the potential of a quantum semantic framework, offer a new lens through which to re-evaluate several core debates and prevailing methodologies in natural language processing and cognitive science. This discussion will first address how the information-theoretic challenge of semantic degeneracy, as quantified by Kolmogorov complexity, impacts the capabilities of distributional semantic models (DSMs) and Large Language Models (LLMs) in tasks requiring deep contextual understanding (Section VI A). Subsequently, we will explore how a quantum semantic framework, emphasiz-
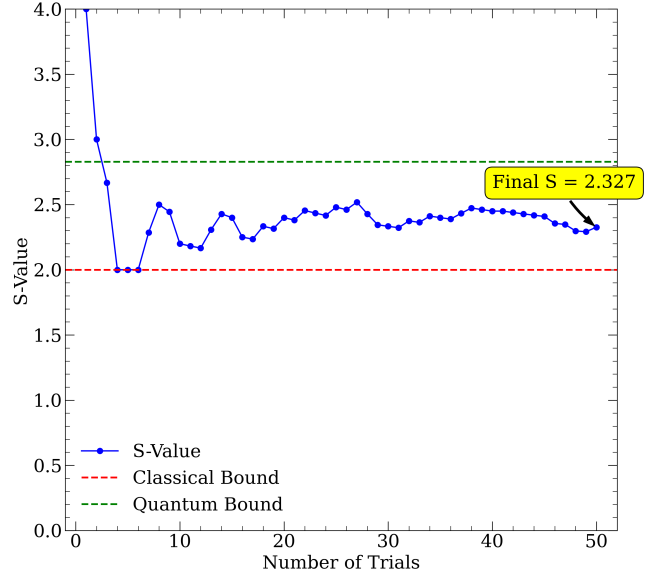

FIG. 2. Example evolution of the calculated CHSH S-value as the trial number progressed for an experiment run with a total of 50 trials, eventually settling at 2.327 (Experiment #7 from Table I). The dashed red line at $S = \pm2$ indicate the classical bounds of the CHSH inequality and the dashed green line indicates the proposed quantum bound at S=$2\sqrt{2}$.

ing observer-dependent meaning actualization, engages with and offers alternatives to classical theories of meaning and interpretation (Section VI B). We will then connect these ideas to specific cognitive phenomena and psycholinguistic findings (Section VI C), and finally, consider the broader implications for NLP research (Section VI D) and practical industrial applications (Section VI E).

A. Semantic Degeneracy: Implications for DSMs and LLMs

Distributional Semantic Models, including contemporary Large Language Models (LLMs), have achieved remarkable success by learning statistical co-occurrences from vast text corpora [2, 6, 7, 9]. These models implicitly operate under the assumption that sufficient statistical exposure can lead to robust representations of meaning. However, the principle of semantic degeneracy, particularly when analyzed through Kolmogorov complexity (KC) as detailed in Section II, reveals a fundamental challenge to this assumption for tasks demanding precise, contextually-specific interpretations, as we have shown that as the complexity of a semantic expression and its required contextual disambiguation increases, the KC of specifying the singularly intended meaning grows superlinearly. This implies that any DSM or LLM—trained on finite sets of data and with fixed weights—will more

likely than not provide solutions and interpretations that are not aligned with the 'intended' one on some critical aspects that results in a complete breakdown in understanding. These information-theoretic limits stem from the combinatorial explosion of potential interpretations and can adequately explain the observed plateaus in LLM performance on complex reasoning tasks and the persistent difficulties DSMs face with ambiguous or context-rich texts [11, 12] despite increasing model size and data. Indeed it appears that the limitations of DSMs at navigating tasks with high ambiguity likely precludes them from ever achieving the status of 'strong' AI. It appears likely that alternative methods or ensemble approaches (e.g. where LLMs are used as reasoning engines along with an alternative, more dynamic model) will prevail.

B. Quantum Semantics versus Classical Theories of Meaning

The limitations highlighted by semantic degeneracy motivate the exploration of alternative frameworks for meaning. Our proposed quantum semantic framework (Section III), which posits that meaning is not an intrinsic property of text but is actualized through an observer-dependent interpretive act, directly confronts the classical assumptions of semantic realism and locality often implicit in DSMs and traditional linguistic theories. Classical realism assumes pre-existing, definite meanings, while locality assumes semantic components can be independently determined. A quantum semantic approach, by contrast, treats expressions as affording a spectrum of potential interpretations (a superposition), with a specific meaning being 'collapsed' or actualized by an agent's interpretive 'measurement' within a given context. This aligns with constructivist theories in cognitive science that view understanding as an active, situated process [22, 26, 78]. Philosophical critiques of essentialist views of meaning [e.g., later 79, 80] also resonate with our results as meaning becomes tied to use and interaction rather than inhering statically in symbols. The non-commutativity of interpretive operations—a feature of the quantum semantic model—implies that the order of contextual probing can affect the realized meaning, a phenomenon difficult to capture in purely classical additive models but observed in human cognition, e.g. order effects in judgment [81] and quantum-like contextuality effects such as the 'Pet-Fish' problem [58].

C. Quantum Semantics: Connections to Cognitive and Psycholinguistic Phenomena

The observer-dependent actualization central to quantum semantics provides a compelling explanatory framework for a range of established cognitive and psycholinguistic findings. The profound influence of an agent's current context, goals, and primed semantic memory on

interpretation [38, 41, 42, 47, 48, 82] is naturally accommodated if meaning is not fixed but realized in interaction. This quantum concept of 'measurement contextuality' aligns with models of dynamic semantic memory [27, 49, 50], accounts of attentional focus [28–30], the non-algorithmic nature of Relevance Realization [45], and even cross-linguistic variations in semantic categorization [33, 34, 37].

This notion of measurement contextuality finds strong empirical parallels in human judgment research. For instance, work by Bruza *et al.* [55, 64] on facial judgments extends contextuality beyond language to perception, arguing that such properties are indeterminate and constructed in the moment of judgment. This reinforces our core premise: if concrete perceptual traits are indeterminate, then abstract semantic meaning is even more likely to require an interpretive act to become actualized. Our experimental findings provide direct quantitative support for this view. The violation of the CHSH inequality we observed is comparable in magnitude to results from cognitive experiments with humans, such as the work by Aerts *et al.* [62] on conceptual entanglement. This parallel suggests that LLMs tap into the same non-classical probabilistic structures inherent in human semantic processing.

This convergence of findings positions our LLM experiments as a novel methodology for investigating the fundamental nature of semantic contextuality itself. In this light, LLMs serve as sophisticated "computational cognitive systems" capable of bridging the gap between the macro-level statistical phenomena found in collective human data and the micro-level interpretive acts of an individual agent. The contextuality we observe, therefore, suggests this non-classical behavior is not a quirk of human psychology or a specific LLM architecture, but a pervasive feature of how semantic meaning is structured and processed in any complex, interconnected system.

D. Methodological Shifts: Bayesian Exploration and Quantum Cognition Analogies

The dual arguments—the limitations imposed by semantic degeneracy on classical systems and the potential of a quantum semantic framework—strongly suggest a paradigm shift in NLP methodologies towards non-classical, Bayesian-informed approaches. Instead of pursuing a single, definitive interpretation, techniques involving Monte Carlo sampling of interpretations under diverse contextual conditions, combined with dynamic explorations of semantic space (e.g., through Markovian random walks), may offer more practically useful and robust characterizations of text. This is particularly relevant for tasks such as nuanced translation, novel discovery, and complex single-turn completion where LLMs currently face inherent difficulties due to the Kolmogorov complexity challenge; a system designed to explore a wide distribution of plausible actualizations, effectively navi-

gating many potential paths from problem to solution, while slower, would likely prove more effective at approximating an understanding of the requirements of the semantic expression. Adopting such a Bayesian-informed perspective allows systems to treat ambiguity not as an error to be eliminated, but as an inherent and informative feature of the semantic landscape. This approach directly addresses the computational intractability posed by semantic degeneracy, offering a practical path toward building more resilient and contextually-aware language technologies that better reflect the probabilistic nature of meaning itself.

E. Practical Architectures and the Role of Human Oversight

Translating these conceptual shifts into practical applications, especially for complex scenarios such as multi-agent systems [83, 84], or large-scale enterprise environments demanding dynamic document understanding [85, 86], necessitates the development of novel and adaptable system architectures for models attempting to approach intelligence. These architectures must be capable of managing and leveraging contextual variability rather than attempting to eliminate it. Crucially, the observer-dependent nature of meaning actualization, as posited by the quantum semantic framework, underscores the enduring and fundamental importance of Human-in-the-Loop (HITL) systems. Far from being a temporary measure until AI achieves perfect autonomy, HITL will remain an integral component for navigating inherent semantic ambiguity, validating interpretations, and ensuring that system outputs align with human goals and ethical considerations, particularly in safety-critical domains [87–89] like healthcare, defense, and finance. The explicit recognition of both the fundamental limits of purely automated interpretation in open contexts and the potential of alternative, context-aware frameworks can guide the development of more realistic, robust, and ultimately more capable language technologies.

VII. CONCLUSIONS

In this work, we have used a theoretical framework based on Kolmogorov complexity and a novel experimental design using Large Language Model agents to inves-

tigate the fundamental, non-classical nature of semantic meaning. We have approached this analysis by identifying the information-theoretic limits inherent in any act of linguistic interpretation and have provided the first known test of contextuality in the interpretive acts of diverse LLM-powered AI agents. Our major conclusions are the following:

1. Semantic degeneracy is a fundamental property of natural language that imposes an information-theoretic limit on interpretation; our analysis using Kolmogorov complexity (Section II) formalizes how this makes the recovery of a single intended meaning from a complex expression computationally intractable for any system, thereby providing a clear explanation for the observed performance plateaus in LLMs.

2. Linguistic interpretation under ambiguity exhibits non-classical contextuality, as demonstrated by frequent and significant violations of the CHSH inequality ($|S| > 2$) in our semantic Bell test experiments with LLM agents (Sections IV, V).

3. The contextuality measured in the interpretive acts of LLMs is consistent with a broader pattern of non-classical findings across human cognitive science, indicating that observer-dependence and indeterminacy are general principles of information processing and not simply artifacts of human psychology.

4. The observer-dependent nature of meaning, confirmed by our experiment, reveals that there is no absolute, fundamental meaning to be found, only contextualized interpretations; consequently, the only viable scientific methodology is to shift from seeking any single "correct" answer and instead use repeated Bayesian-style sampling to characterize how these conditional interpretations interconnect within a possibility space.

5. The consistent emergence of non-classical contextuality across a diverse pool of non-biological LLM agents, when considered alongside similar findings in human cognition, indicates that these statistical properties are not artifacts of any specific interpretive system but are objective, structural features of natural language itself.

[1] H. L. Dreyfus, *What Computers Still Can?T Do: A Critique of Artificial Reason* (MIT Press, 1992).
[2] Z. S. Harris, Distributional Structure, WORD **10**, 146 (1954).
[3] S. C. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. A. Harshman, Indexing by latent semantic

analysis, Journal of the American Society for Information Science **41**, 391 (1990).
[4] T. K. Landauer and S. T. Dumais, A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge, Psychological Review **104**, 211 (1997).

[5] T. L. Griffiths, M. Steyvers, and J. B. Tenenbaum, Topics in semantic representation, Psychological Review **114**, 211 (2007).

[6] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, Distributed Representations of Words and Phrases and their Compositionality, arXiv preprint arXiv:1310.4546 (2013), arXiv:1310.4546 [cs.CL].

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv Preprint arXiv:1810.04805 (2018), arXiv:1810.04805.

[8] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, Language models are unsupervised multitask learners, OpenAI Blog (2019), openAI technical report, Version 1.

[9] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, Language models are few-shot learners, in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc., 2020) pp. 1877–1901.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention Is All You Need, arXiv preprint arXiv:1706.03762 (2017), arXiv:1706.03762 [cs.CL].

[11] M. N. Jones, When does abstraction occur in semantic memory: Insights from distributional models, Language, Cognition and Neuroscience **34**, 1338 (2018).

[12] P. Hoffman, M. A. Lambon Ralph, and T. T. Rogers, Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words, Behavior Research Methods **45**, 718 (2013).

[13] M. Assran, Q. Duval, I. Misra, P. Bojanowski, P. Vincent, M. Rabbat, Y. LeCun, and N. Ballas, Self-supervised learning from images with a joint-embedding predictive architecture, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023) pp. 15619–15629.

[14] A. Gu and T. Dao, Mamba: Linear-Time Sequence Modeling with Selective State Spaces, arXiv e-prints , arXiv:2312.00752 (2023), arXiv:2312.00752 [cs.LG].

[15] L. Darlow, C. Regan, S. Risi, J. Seely, and L. Jones, Continuous Thought Machines, arXiv e-prints , arXiv:2505.05522 (2025), arXiv:2505.05522 [stat.ML].

[16] Q. Sun, E. Cetin, and Y. Tang, Transformer-Squared: Self-adaptive LLMs, arXiv e-prints , arXiv:2501.06252 (2025), arXiv:2501.06252 [cs.LG].

[17] T. Simonds and A. Yoshiyama, LADDER: Self-Improving LLMs Through Recursive Problem Decomposition, arXiv e-prints , arXiv:2503.00735 (2025), arXiv:2503.00735 [cs.LG].

[18] P. Tabossi, L. Colombo, and R. Job, Accessing lexical ambiguity: Effects of context and dominance, Psychological Research **49**, 161 (1987).

[19] K. S. Binder and K. Rayner, Contextual strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading, Psychonomic Bulletin & Review **5**, 271 (1998).

[20] K. Rayner, A. E. Cook, B. J. Juhasz, and L. Frazier, Immediate disambiguation of lexically ambiguous words during reading: Evidence from eye movements, British Journal of Psychology **97**, 467 (2006).

[21] S. A. McDonald and R. C. Shillcock, Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing, Language and Speech **44**, 295 (2001).

[22] L. W. Barsalou, Situated simulation in the human conceptual system, Language and Cognitive Processes **18**, 513 (2003).

[23] W. Yeh and L. W. Barsalou, The Situated Nature of Concepts, The American Journal of Psychology **119**, 349 (2006).

[24] E. T. Higgins, Knowledge activation: Accessibility, applicability, and salience, in *Social Psychology: Handbook of Basic Principles*, edited by E. T. Higgins and A. W. Kruglanski (Guilford Press, 1996) pp. 133–168.

[25] K. Bicknell, J. L. Elman, M. Hare, K. McRae, and M. Kutas, Effects of event knowledge in processing verbal arguments, Journal of Memory and Language **63**, 489 (2010).

[26] W. Kintsch, Predication, Cognitive Science **25**, 173 (2001).

[27] L. A. M. Lebois, C. D. Wilson-Mendenhall, and L. W. Barsalou, Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects, Cognitive Science **39**, 1764 (2015).

[28] E. Yee, S. Z. Ahmed, and S. L. Thompson-Schill, Colorless green ideas (can) prime furiously, Psychological Science **23**, 364 (2012).

[29] K. Hoenig, E.-J. Sim, V. Bochev, B. Herrnberger, and M. Kiefer, Conceptual Flexibility in the Human Brain: Dynamic Recruitment of Semantic Maps from Visual, Motor, and Motion-related Areas, Journal of Cognitive Neuroscience **20**, 1799 (2008).

[30] S. Van Dantzig, D. Pecher, R. Zeelenberg, and L. W. Barsalou, Perceptual Processing Affects Conceptual Processing, Cognitive Science **32**, 579 (2008).

[31] C. Bermeitinger, D. Wentura, and C. Frings, How to switch on and switch off semantic priming effects: Activation processes in category memory depend on focusing specific feature dimensions, Psychonomic Bulletin & Review **18**, 579 (2011).

[32] N. S. Hsu, D. J. M. Kraemer, R. T. Oliver, M. L. Schlichting, and S. L. Thompson-Schill, Color, Context, and Cognitive Style: Variations in Color Knowledge Retrieval as a Function of Task and Subject Variables, Journal of Cognitive Neuroscience **23**, 2544 (2011).

[33] P. Athanasopoulos, Cognitive representation of colour in bilinguals: The case of Greek blues, Bilingualism: Language and Cognition **12**, 83 (2009).

[34] B. Thompson, S. G. Roberts, and G. Lupyan, Cultural influences on word meanings revealed through large-scale semantic alignment, Nature Human Behaviour **4**, 1029 (2020).

[35] B. T. Johns, Distributional social semantics: Inferring word meanings from communication patterns, Cognitive Psychology **131**, 101441 (2021).

[36] J. Cesario, J. E. Plaks, N. Hagiwara, C. D. Navarrete, and E. T. Higgins, The Ecology of Automaticity: How Situational Contingencies Shape Action Semantics and Social Behavior, Psychological Science **21**, 1311 (2010).

[37] V. Marian and M. Kaushanskaya, Language context guides memory content, Psychonomic Bulletin & Review **14**, 925 (2007).

[38] L. W. Barsalou, Simulation, situated conceptualization, and prediction, Philosophical Transactions of the Royal Society B: Biological Sciences **364**, 1281 (2009).

[39] D. Pecher, R. Zeelenberg, and J. G. W. Raaijmakers, Does Pizza Prime Coin? Perceptual Priming in Lexical Decision and Pronunciation, Journal of Memory and Language 38, 401 (1998).

[40] A. M. Borghi, A. M. Glenberg, and M. P. Kaschak, Putting words in perspective, Memory & Cognition 32, 863 (2004).

[41] L. Connell and D. Lynott, Principles of Representation: Why You Can't Represent the Same Concept Twice, Topics in Cognitive Science 6, 390 (2014).

[42] E. Musz and S. L. Thompson-Schill, Semantic variability predicts neural variability of object concepts, Neuropsychologia 76, 41 (2015).

[43] J. Vervaeke and L. Ferraro, Relevance, meaning and the cognitive science of wisdom (2013).

[44] B. P. Andersen, M. Miller, and J. Vervaeke, Predictive processing and relevance realization: Exploring convergent solutions to the frame problem, Phenomenology and the Cognitive Sciences , 1 (forthcoming).

[45] J. Jaeger, A. Riedl, A. Djedovic, J. Vervaeke, and D. Walsh, Naturalizing relevance realization: Why agency and cognition are fundamentally not computational (2023).

[46] J. Vervaeke and L. Ferraro, Relevance realization and the neurodynamics and neuroconnectivity of general intelligence, in SmartData, edited by I. Harvey, A. Cavoukian, G. Tomko, D. Borrett, H. Kwan, and D. Hatzinakos (Springer New York, New York, NY, 2013) pp. 57–68.

[47] D. L. Hintzman, MINERVA 2: A simulation model of human memory, Behavior Research Methods, Instruments, & Computers 16, 96 (1984).

[48] R. K. Jamieson, J. E. Avery, B. T. Johns, and M. N. Jones, An Instance Theory of Semantic Memory, Computational Brain & Behavior 1, 119 (2018).

[49] W. O. Van Dam, S.-A. Rueschemeyer, O. Lindemann, and H. Bekkering, Context Effects in Embodied Lexical-Semantic Processing, Frontiers in Psychology 1, 150 (2010).

[50] W. O. Van Dam, M. Van Dijk, H. Bekkering, and S.-A. Rueschemeyer, Flexibility in embodied lexical-semantic representations, Human Brain Mapping 33, 2322 (2012).

[51] D. Aerts, Quantum structure in cognition, Journal of Mathematical Psychology 53, 314 (2009), special Issue: Quantum Cognition.

[52] P. Bruza, J. R. Busemeyer, and L. Gabora, Introduction to the special issue on quantum cognition, Journal of Mathematical Psychology 53, 303 (2009), special Issue: Quantum Cognition.

[53] L. Gabora and D. A. and, Contextualizing concepts using a mathematical generalization of the quantum formalism, Journal of Experimental & Theoretical Artificial Intelligence 14, 327 (2002), https://doi.org/10.1080/09528130210162253.

[54] P. Bruza, K. Kitto, B. Ramm, L. Sitbon, and D. Song, Quantum-like non-separability of concept combinations, emergent associates and abduction, Logic Journal of the IGPL 20, 445 (2012).

[55] P. Bruza, K. Kitto, B. Ramm, and L. Sitbon, A probabilistic framework for analysing the compositionality of conceptual combinations, Journal of Mathematical Psychology 67, 26 (2015).

[56] J. M. Yearsley and J. R. Busemeyer, Quantum cognition and decision theories: A tutorial, Journal of Mathematical Psychology 74, 99 (2016), foundations of Probability Theory in Psychology and Beyond.

[57] E. M. Pothos and J. R. Busemeyer, Quantum cognition, Annual Review of Psychology 73, 749 (2022).

[58] D. Aerts, M. Czachor, B. D'Hooghe, S. Sozzo, et al., The pet-fish problem on the world-wide web., in AAAI Fall Symposium: Quantum Informatics for Cognitive, Social, and Semantic Processes (2010).

[59] J. S. Bell, On the einstein podolsky rosen paradox, Physics Physique Fizika 1, 195 (1964).

[60] S. Uprety, Investigation and modelling of quantum-like user cognitive behaviour in information access and retrieval (2020).

[61] D. Aerts, S. Aerts, J. Broekaert, and L. Gabora, The violation of bell inequalities in the macroworld, Foundations of Physics 30, 1387 (2000).

[62] D. Aerts, J. A. Arguëlles, L. Beltran, S. Geriente, M. S. de Bianchi, S. Sozzo, and T. Veloz, Spin and wind directions i: Identifying entanglement in nature and cognition, Foundations of Science 23, 323 (2018).

[63] D. Aerts, J. A. Arguëlles, L. Beltran, S. Geriente, M. Sassoli de Bianchi, S. Sozzo, and T. Veloz, Spin and wind directions ii: A bell state quantum model, Foundations of Science 23, 337 (2018).

[64] P. Bruza, L. Fell, P. Hoyte, S. Dehdashti, A. Obeid, A. Gibson, and C. Moreira, Contextuality and context-sensitivity in probabilistic models of cognition, Cognitive Psychology 140, 101529 (2023).

[65] A. N. Kolmogorov, Three approaches to the quantitative definition of information, Problems of Information Transmission 1, 1 (1965).

[66] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed experiment to test local hidden-variable theories, Phys. Rev. Lett. 23, 880 (1969).

[67] W. Guo and A. Caliskan, Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases, in Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (2021) pp. 122–133.

[68] B. K. Payne, H. A. Vuletich, and K. B. Lundberg, The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice, Psychological Inquiry 28, 233 (2017).

[69] Z. Siddique, I. Khalid, L. D. Turner, and L. Espinosa-Anke, Shifting perspectives: Steering vector ensembles for robust bias mitigation in llms, arXiv preprint arXiv:2503.05371 (2025).

[70] I. O. Gallegos, R. A. Rossi, J. Barrow, M. M. Tanjim, S. Kim, F. Dernoncourt, T. Yu, R. Zhang, and N. K. Ahmed, Bias and fairness in large language models: A survey, arXiv preprint arXiv:2309.00770 (2023).

[71] A. Kitadai, K. Ogawa, and N. Nishino, Examining the feasibility of large language models as survey respondents, in 2024 IEEE International Conference on Big Data (BigData) (2024) pp. 3858–3864.

[72] A. Salecha, M. E. Ireland, S. Subrahmanya, J. Sedoc, L. H. Ungar, and J. C. Eichstaedt, Large language models display human-like social desirability biases in big five personality surveys, PNAS Nexus 3, pgae533 (2024), https://academic.oup.com/pnasnexus/article-pdf/3/12/pgae533/61188312/pgae533.pdf.

[73] L. Tjuatja, V. Chen, T. Wu, A. Talwalkar, and G. Neubig, Do llms exhibit human-like response biases? a case study in survey design, Transactions of the Association for Computational Linguistics

12, 1011 (2024), https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00685/2468689/tacl_a_00685.pdf.

[74] S. T. Piantadosi and F. Hill, Meaning without reference in large language models, arXiv preprint arXiv:2208.02957 (2022).

[75] A. Lampinen, I. Dasgupta, S. Chan, K. Mathewson, M. Tessler, A. Creswell, J. McClelland, J. Wang, and F. Hill, Can language models learn from explanations in context?, in *Findings of the Association for Computational Linguistics: EMNLP 2022*, edited by Y. Goldberg, Z. Kozareva, and Y. Zhang (Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022) pp. 537–563.

[76] E. N. Dzhafarov, J. V. Kujala, and V. H. Cervantes, Contextuality-by-default: a brief overview of ideas, concepts, and terminology, in *Quantum Interaction: 9th International Conference, QI 2015, Filzbach, Switzerland, July 15-17, 2015, Revised Selected Papers 9* (Springer, 2016) pp. 12–23.

[77] V. H. Cervantes and E. N. Dzhafarov, Advanced analysis of quantum contextuality in a psychophysical double-detection experiment, Journal of Mathematical Psychology **79**, 77 (2017).

[78] J. L. Elman, Finding Structure in Time, Cognitive Science **14**, 179 (1990).

[79] L. Wittgenstein, *Philosophical Investigations* (Basil Blackwell, Oxford, 1953).

[80] W. V. O. Quine, *Word & Object* (MIT Press, 1960).

[81] R. M. Hogarth and H. J. Einhorn, Order effects in belief updating: The belief-adjustment model, Cognitive Psychology **24**, 1 (1992).

[82] J. R. Anderson and R. Milson, Human memory: An adaptive perspective, Psychological Review **96**, 703 (1989).

[83] Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. H. Awadallah, R. W. White, D. Burger, and C. Wang, AutoGen: Enabling next-gen large language model applications via multi-agent conversation, arXiv preprint arXiv:2308.08155 (2023).

[84] LlamaIndex Contributors, *Multi-Agent Workflows — LlamaIndex Documentation* (2025).

[85] P. Herzig, Sap's new ai innovations and partnerships deliver real-world results (2024).

[86] J. Srivastava, J. Schuurmans, K. Ray, L. Mesnage, N. Butler, S. Dutta, T. Kubit, and T. Janner, Genai can revolutionize erp transformations, Boston Consulting Group Insight (2025).

[87] V. Quach, A. Fisch, T. Schuster, A. Yala, J. H. Sohn, T. S. Jaakkola, and R. Barzilay, Conformal language modeling, arXiv preprint arXiv:2306.10193 (2023).

[88] Z. Li, W. Wu, Y. Wang, Y. Xu, W. Hunt, and S. Stein, HMCF: A human-in-the-loop multi-robot collaboration framework based on large language models, arXiv preprint arXiv:2505.00820 (2025).

[89] K. Naminas, Human in the loop machine learning: The key to better models (2025).