DYNAMICS REVEALS STRUCTURE: CHALLENGING THE LINEAR PROPAGATION ASSUMPTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural networks adapt through first-order parameter updates, yet it remains unclear whether such updates preserve logical coherence. We investigate the geometric limits of the Linear Propagation Assumption (LPA), the premise that local updates coherently propagate to logical consequences. To formalize this, we adopt relation algebra and study three core operations on relations: negation flips truth values, converse swaps argument order, and composition chains relations. For negation and converse, we prove that guaranteeing direction-agnostic first-order propagation necessitates a tensor factorization separating entity-pair context from relation content. However, for composition, we identify a fundamental obstruction. We show that composition reduces to conjunction, and prove that any conjunction well-defined on linear features must be bilinear. Since bilinearity is incompatible with negation, this forces the feature map to collapse. These results suggest that failures in knowledge editing, the reversal curse, and multi-hop reasoning may stem from common structural limitations inherent to the LPA.

1 INTRODUCTION

Modern machine learning systems evolve through a long trajectory, spanning pretraining, continual learning (Wu et al., 2024), and unlearning (Jang et al., 2023). Throughout this lifecycle, the central operation for adapting to new information is the first-order parameter update. Ideally, this adaptation should be rational: when a model revises its belief about a fact, its logically related beliefs should update accordingly to maintain coherence. However, maximizing the likelihood of a target fact is fundamentally an optimization process, distinct from rational belief revision (Hase et al., 2024). Consequently, it remains unclear whether the local geometry of gradient-based updates can inherently guarantee such logical consistency without inducing contradictions.

Encouraged by the impressive capabilities of current Large Language Models (LLMs) in logical reasoning during inference (Kojima et al., 2022; Achiam et al., 2023), it is tempting to assume that such coherence is preserved under local, first-order parameter updates. This premise, which we term the **Linear Propagation Assumption (LPA)**, often serves as a foundational design principle in current techniques. For instance, prominent knowledge editing methods explicitly formulate the update as a constrained linear least-squares problem, treating network layers as linear associative memories (Bau et al., 2020; Meng et al., 2022a;b). Beyond editing, this implicit assumption also appears in continual learning strategies that aim to add tasks without forgetting (Lopez-Paz & Ranzato, 2017) and unlearning techniques designed to erase specific knowledge (Jang et al., 2023).

However, the validity of the LPA is questionable, as empirical failures of LLMs on logical coherence show. For instance, LLMs exhibit the "reversal curse," failing to generalize to reverse relationship (Berglund et al., 2023), and struggle with compositional reasoning tasks (Dziri et al., 2023). Since these representations are constructed through first-order updates, such persistent failures suggest that the update mechanism may not reliably imprint the necessary logical structure. This limitation is highlighted in knowledge editing, where even carefully targeted updates consistently fail to propagate to logical consequences such as negations or implications (Cohen et al., 2024; Liu et al., 2025). These phenomena suggest a shared structural issue: the geometry of first-order updates imposes structural constraints that are inherently ill-suited for systematic logical operations.

In this work, we rigorously investigate this hypothesis by asking: **What structural constraints are imposed on a model's representation if we demand that local linear updates respect the**

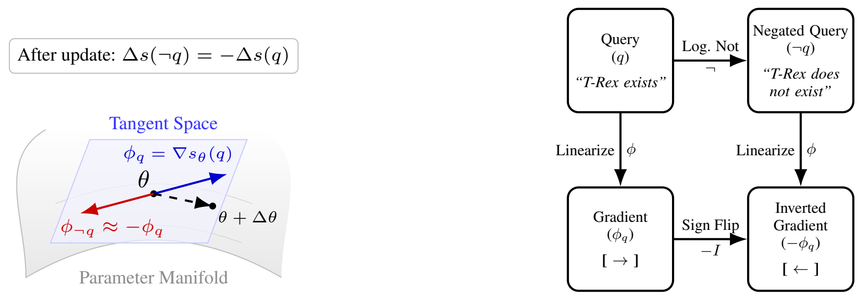

Figure 1: **Geometric interpretation of logical equivariance. (Left)** A query q is associated with a score $s_\theta(q)$, and its gradient $\phi_q = \nabla s_\theta(q)$ as a feature. Logical consistency under direction-agnostic first-order propagation requires that any local parameter change enhancing q suppresses $\neg q$, necessitating the gradient vectors be anti-aligned ($\phi_{\neg q} \approx -\phi_q$). **(Right)** This induces a commutative diagram where symbolic negation $\neg$ in the query space corresponds to a linear inversion $-I$ in the gradient feature space.

logical structure of relational knowledge? To answer this, we formalize relational knowledge using relation algebra (Givant, 2006), which builds relational knowledge via three fundamental operations: **negation** (flipping truth values), **converse** (swapping argument order), and **composition** (chaining relations). Following Tarski's invariance criterion for logical notions (Tarski & Corcoran, 1986; Sher, 2008), we treat entity renamings as symmetries of the query space and ask what constraints such symmetries impose on the linearized geometry. This provides a principled way to test whether first-order parameter updates can support systematic propagation of logical operations. Geometrically, satisfying these criteria requires navigating a fundamental trade-off between total superposition and perfect decoupling. Instead, we seek a *structured coupling*: for example, an update to p must automatically adjust $\neg p$ (Fig. 1), while remaining linearly independent from unrelated facts. We translate this requirement into the geometric structure of gradients to formalize **Systematic Linear Propagation (SLP)**. SLP imposes strict logical equivariance on coupled facts, ensuring that linear updates systematically track unary logical transformations (negation and converse).

Our theoretical analysis reveals necessary conditions on the required geometry of first-order updates. For negation, we prove that guaranteeing direction-agnostic first-order propagation necessitates a tensor factorization separating entity-pair context from relation content (Theorem 1), reminiscent of Smolensky (1990). In contrast, we identify a fundamental obstruction when extending this systematicity to composition (Theorem 2). We show that a minimal form of systematic composition reduces to conjunction, where a conjunction well-defined on linear features necessitates a bilinear structure. However, this bilinearity is incompatible with the geometry enforced by negation equivariance. This conflict forces the feature map to collapse, suggesting that the failure to propagate updates to compositional consequences in the first-order regime may stem from a fundamental geometric mismatch. More broadly, our results support the view that *dynamics reveals structure*: analyzing how representations transform under updates exposes structural necessities that are invisible to static function approximation. This contributes to the systematicity debate (Fodor & Pylyshyn, 1988) by deriving binding-compatible block structure as a geometric necessity for logical coherence under LPA. Ultimately, this opens a path toward **logical geometric deep learning** (Bronstein et al., 2021), treating logical operations as dynamic symmetries to be preserved throughout learning.

2 Problem Formulation

2.1 Preliminary: Relation Algebra

We first formalize the logical structure of relational knowledge using relation algebra. The NLP community has predominantly formalized factual knowledge as a collection of relational triples (h, r, t), where a head entity h and a tail entity t are connected by a binary relation r (e.g., `(T-Rex, HasA, FourLegs)`), interpreting LLM as an implicit knowledge graph (Petroni et al., 2019; Cohen et al., 2024). To analyze this setup, we adopt the formalism of relation algebra (Givant, 2006). Relation algebra provides a rigorous language for manipulating binary relations, which we use to organize the logical operations we study over relational knowledge modeled as a triplet.

Let E be a finite set of entities, and write $U := E \times E$ for the universe of ordered pairs. A relation on E is a subset $r \subseteq U$. Intuitively, $(h, t) \in r$ means that h stands in relation r to t. Now, we write the universe of relations $\mathsf{Rel} := 2^U$ for the set of all binary relations on E. Relation algebra consists of Boolean operations alongside relation-specific operations such as converse and composition. In this work, we study two regimes: we first focus on the unary operations negation ($\neg$) and converse ($(\cdot)^{\smile}$), and later turn to the binary operation of composition ($(\cdot;\cdot)$), which underlies multi-hop reasoning. For any relations $r, s \in \mathsf{Rel}$, these operations are defined set-theoretically as: $\neg r := U \setminus r$, $r^{\smile} := \{(t, h) : (h, t) \in r\}$, and $r; s := \{(h, t) : \exists b \in E, (h, b) \in r \wedge (b, t) \in s\}$, respectively. We focus on these three operations since recent studies identify them as recurring failure modes of LLMs: negation robustness (Kassner & Schütze, 2020; Liu et al., 2025), reversed queries (the "reversal curse") (Berglund et al., 2023), and multi-hop propagation (Cohen et al., 2024).

For our later analysis, let $\mathcal{R}$ be the closure of a base set of relations $\mathcal{R}_0 \subseteq \mathsf{Rel}$ under the two unary operations of negation and converse, i.e., the smallest set containing $\mathcal{R}_0$ and satisfying $r \in \mathcal{R} \Rightarrow \neg r \in \mathcal{R}$ and $r^{\smile} \in \mathcal{R}$. We call $(\mathcal{R}, \neg, (\cdot)^{\smile})$ our *unary relation algebra*. Now, given $h, t \in E$ and $r \in \mathcal{R}$, we write $r(h, t)$ for the atomic formula asserting that $(h, t) \in r$. Negation and converse act on these atomic formulas as $\neg r(h, t)$ and $r^{\smile}(t, h)$, respectively. In later sections, we will study how such operations should be represented and propagated in a linear feature space.

2.2 Queries, Scores, and Linearized Features

We now bridge the gap to the continuous geometry of neural networks. Our goal is to analyze how logical operations on symbols map to geometric transformations on features. To this end, we operate in the linearized regime of parameter updates, treating the gradient of a query as its feature vector.

Queries. We model each atomic formula as a triplet of a head entity, a relation, and a tail entity. Let E be a finite set of entities and let $\mathcal{R}$ be our unary relation algebra. A query is a triple $q = (h, r, t) \in Q := E \times \mathcal{R} \times E$, where a query $q = (h, r, t)$ can be read as an atomic formula $r(h, t)$.

Scores and Linearization. Fix a finite-dimensional real inner-product space $(\Theta, \langle \cdot, \cdot \rangle)$ representing the parameters of a differentiable model, and let $\theta_0 \in \Theta$ be a reference model state. We associate each query $q \in Q$ with a differentiable scalar score $s_\theta(q) \in \mathbb{R}$ (e.g., the logit of the correct tail entity). We assume that the model's decision about q (e.g., predicted truth or preference) is determined by a rule strictly monotone in $s_\theta(q)$. As we are primarily interested in how these scores change under the first-order regime, we define the feature of a query as follows:

Definition 1. *For each query $q \in Q$, we define its* linearized feature *at θ_0 as: $\phi_q := \nabla_\theta s_\theta(q)\big|_{\theta=\theta_0} \in \Theta$. Consequently, the first-order score change under a local parameter update $\Delta\theta$ is given by: $s_{\theta_0+\Delta\theta}(q) \approx s_{\theta_0}(q) + \langle \phi_q, \Delta\theta \rangle$.*

We restrict our analysis to the subspace $W := \operatorname{span}\{\phi_q : q \in Q\} \subseteq \Theta$, since components of an update $\Delta\theta$ orthogonal to W do not affect first-order score changes on the queries we study.

2.3 Systematic Linear Propagation

We now unify the algebraic structure of knowledge (Sec. 2.1) with the geometric view of features (Sec. 2.2). Our goal is to formalize when a linearized model supports the *systematic* propagation with respect to the unary relation algebra we consider. Qualitatively, systematicity imposes structured coupling in the feature space: an update to a query q should propagate to its converse and inversely to its negation, while logically unrelated facts remain independently controllable. A key choice in our formulation is that we do *not* define systematicity as the mere existence of a carefully selected update direction that satisfies these constraints. Instead, we ask for **automaticity**: logical coupling must be an intrinsic geometric property of the representation. This ensures that once an update is applied to enforce q, the induced effects on logically related queries follow *automatically*, without requiring the optimizer to solve a separate constraint-satisfaction problem. Formally, we require the logical constraints to hold for all $\Delta\theta \in W$ in the first-order regime.

This direction-agnostic requirement is motivated by the potential fragility of relying on specific "safe" update directions in high-capacity models. In practical regimes where the feature space is highly superposed (Elhage et al., 2022; Hu et al., 2025), reliably identifying an update direction that satisfies logical constraints while avoiding interference with unrelated facts is often prohibitively difficult. Furthermore, in lifelong learning settings, any such transient "safe subspace" itself might be prone to drift as the representation evolves. Thus, by enforcing systematicity for all update directions, we isolate a notion in which logic is an intrinsic property of the local geometry rather than an artifact of a specific optimization trajectory or a transient subspace.

To formalize this, we first identify the closed sets of queries that must be logically coupled under any update. Recall that our unary relation algebra $\mathcal{R}$ is closed under negation $\neg$ and converse $(\cdot)^{\smile}$. These relational operations induce a corresponding action on the space of queries Q as: $\neg(h,r,t) := (h,\ \neg r,\ t)$, and $\mathrm{rev}(h,r,t) := (t,\ r^{\smile},\ h)$, respectively. Intuitively, $\neg$ flips the truth value by acting locally on the relation slot (e.g., $\neg$`ChildOf` $\to$ `NotChildOf` where `ChildOf` $\subseteq U$ and `NotChildOf` $:= U \setminus$ `ChildOf`), while keeping entities fixed. In contrast, rev swaps the head and tail entities while moving to the converse relation (e.g., `ChildOf`$^{\smile}$ $\to$ `ParentOf`), thereby preserving the truth value (i.e., $r(h,t) \Leftrightarrow r^{\smile}(t,h)$). Observe that both operations are involutions and commute: $\neg(\neg q) = q$, $\mathrm{rev}(\mathrm{rev}(q)) = q$, and $\mathrm{rev}(\neg q) = \neg(\mathrm{rev}(q))$. Hence, these two logical operations generate a group $G := \{\, \mathrm{id},\ \neg,\ \mathrm{rev},\ \neg\circ\mathrm{rev} \,\}$ acting on Q. We define the orbit of a query q under this group as its **logical family**: $G \cdot q := \{\, g(q) : g \in G \,\}$. In plain words, two queries lie in the same family iff one can be transformed into the other via the group operations.

We now translate the semantic requirements of these families into constraints on the feature space W. As discussed above, we require that the score changes are coordinated for *all* update directions $\Delta\theta$. For negation, the condition $\Delta s(\neg q) = -\Delta s(q)$ implies that the feature vectors are strictly anti-aligned: $\phi_{\neg q} = -\phi_q$. Applying the same logic to the converse operation yields $\phi_{\mathrm{rev}(q)} = \phi_q$. Therefore, we formalize this requirement as follows:

Definition 2. *A feature map $\phi : Q \to W$ is* logically equivariant *w.r.t.* $\neg$ *and* rev *if $\phi_{\neg q} = -\,\phi_q$ and $\phi_{\mathrm{rev}(q)} = \phi_q$ for all $q \in Q$.*

While Def. 2 ensures coupling *within* families, we must also ensure that logically unrelated facts do not interfere with each other. Therefore, to enable selective update, distinct logical families must be linearly independent in the feature space. We formulate these two requirements, intra-family coupling and inter-family decoupling, as **Systematic Linear Propagation (SLP)**.

Definition 3 (Systematic Linear Propagation (SLP))**.** *A linearized model $\{\phi_q\}_{q\in Q}$ is said to satisfy* Systematic Linear Propagation *with respect to the unary relation algebra if: (1) It is logically equivariant in the sense of Def. 2 (2) There exists a choice of one representative query from each family s.t. the corresponding feature vectors are linearly independent in W.*

3 SLP Induces Tensor-Factorized Features

To rigorously derive the geometric structure, we adopt Tarski's criterion (Tarski & Corcoran, 1986; Sher, 2008) as our guiding principle. It posits that logical notions are characterized by their invariance under all permutations of the domain's objects. In our context, this implies that logical operations must be invariant to the specific identities of the entities involved. For instance, the logic of negation should apply equally to `T-Rex` and `Chicken`. We formalize this requirement by identifying the symmetry group acting on the query space Q. Let $G_E := \mathrm{Sym}(E)$ be the permutation group acting on the set of entities E. Following Tarski's criterion, the system must remain consistent under any entity renaming $\sigma \in G_E$, which acts on queries by renaming entities uniformly: $\sigma \cdot (h,r,t) := (\sigma(h), r, \sigma(t))$. Simultaneously, the negation operation ($\neg$) is an involution ($\neg(\neg q) = q$), generating the cyclic group of order 2, denoted as $\mathbb{Z}_2 = \{1, -1\}$.

Crucially, **entity renaming commutes with logical operations:** renaming entities does not alter the logical relationship, and logical transformations do not affect entity identities, i.e., $\sigma\cdot(\neg q) = \neg(\sigma\cdot q)$. Consequently, we can define a product group $H := G_E \times \mathbb{Z}_2$, which acts on the set of queries Q via the simultaneous action of renaming and optional negation. We translate these symbolic symmetries into the geometry of the feature space W. As formally verified in Lemma 7, under SLP, it is guaranteed that the action of H on the query space induces a well-defined linear group representation of H on the feature space W.[1] Utilizing stanard results in representation theory, we prove the following factorization theorem.

[1] We provide a brief primer on group representation theory in Sec. C.

Theorem 1 (Proof in Sec. D). *Let $\phi : Q \to W$ be the feature map defined by $q \mapsto \phi_q$. If ϕ satisfies SLP, then there exist real vector spaces $\{C_i\}_i$, $\{R_i\}_i$ and an isomorphism $W \cong \bigoplus_i (C_i \otimes R_i)$ s.t.*

$$\phi(h, r, t) \;=\; \bigoplus_i \left(\sum_{k=1}^{m_i} u_{i,k}(h, t) \otimes v_{i,k}(r) \right),$$

where $u_{i,k} : E \times E \to C_i$ and $v_{i,k} : \mathcal{R} \to R_i$. Moreover, negation acts locally as a sign flip on each relation component, i.e., $v_{i,k}(\neg r) = -v_{i,k}(r)$.

Thus, to support systematic linear propagation with respect to negation, thse feature geometry must separate entity-pair context information from relation information. Specifically, logical negation is realized via negation on the relation factors $v_{i,k}(r)$. See Sec. B for additional results on converse.

4 The Collapse of Linear Conjunction

In this section, we seek systematic propagation to compositional consequences: a targeted update to an atomic formula on a relation $r \in \mathcal{R}$ must automatically adjust composed queries involving a composition $r; s$ for some $s \in \mathcal{R}$. That is, we require that the feature of $r; s$ be modeled *constructively* from its constituents to ensure propagation under arbitrary parameter updates that change r. Recall from Sec. 2.1 that composition, $(r; s)(h, t) \iff \exists b : \ r(h, b) \wedge s(b, t)$, links two queries via an intermediate entity. While the general case involves existential aggregation over possibly many witnesses, we isolate a minimal subclass that removes aggregation altogether. Specifically, in the *unique-witness* case where there exists a unique b^* satisfying the link, composition reduces to the conjunction $r(h, b^*) \wedge s(b^*, t)$. Thus, any mechanism that supports systematic composition must, at a minimum, support systematic conjunction. We therefore investigate whether conjunction can be realized in the linearized feature space while remaining compatible with negation.

We first characterize the constraints LPA imposes on feature geometry to ensure that conjunction is supported systematically. Central to this relationship is compositionality: the representation of a compound statement should be systematically determined by its constituents. That is, given a compound query $p \wedge q$ with $p, q \in Q$, its feature $\phi_{p\wedge q}$ should be determined solely by ϕ_p and ϕ_q. Along with the logical properties of sentential conjunction, we formalize this intuition as follows:

Definition 4. *Let $Q^\wedge$ be the closure of Q with respect to negation and conjunction, and let $W^\wedge := \mathrm{span}\{\phi_p : p \in Q^\wedge\}$. We say that ϕ is conjunction-faithful (under LPA) if there exists a binary operator $F : W^\wedge \times W^\wedge \to W^\wedge$ that governs the conjunction of features, satisfying the following properties for all $p, q \in Q^\wedge$ and $u, v \in W^\wedge$: **(i) Consistency:** $\phi_{p\wedge q} \;=\; F(\phi_p, \phi_q)$. **(ii) Symmetry:** $F(u, v) \;=\; F(v, u)$. **(iii) Idempotence:** $F(u, u) \;=\; u$.*

Below is a direct consequence of Def. 4.

Lemma 1. *If ϕ is conjunction-faithful, then for all $p, p', q \in Q^\wedge$, $\phi_p = \phi_{p'}$ implies $\phi_{p\wedge q} = F(\phi_p, \phi_q) = F(\phi_{p'}, \phi_q) = \phi_{p'\wedge q}$.*

While Lemma 1 guarantees that conjunction is a well-defined function of features, the linearity of LPA imposes a stronger constraint. Under the first-order regime, the editing dynamics are governed entirely by linear projections ($\Delta s(p) = \langle \phi_p, \Delta\theta \rangle$). This implies that the editor cannot distinguish any linear dependence: if a weighted sum of features is zero ($\sum_i a_i \phi_{p_i} = 0$), the collective score change is identically zero for any parameter update. Geometrically, such a zero-sum combination constitutes information that is operationally non-existent to the model's update mechanism. If a conjunction operator were to map a null signal to a non-zero feature, it would generate distinctions based on information invisible to the editor, thereby decoupling the logic of propagation from the physics of editing. To ensure that propagation remains predictable from first-order geometry alone, we adopt a strong notion of systematicity: logical operations must be consistent with the linear geometry of the editor, meaning they must preserve linear dependencies, formalized as below.

Assumption 1. *Let $V^\wedge := \mathrm{span}\{e_p : p \in Q^\wedge\}$ be the free real vector space on $Q^\wedge$, and let $\Phi : V^\wedge \to W^\wedge$ be the linear extension defined by $\Phi(e_p) := \phi_p$. For each fixed $q \in Q^\wedge$, define the linear map $T_q : V^\wedge \to V^\wedge$ by $T_q(e_p) := e_{p\wedge q}$. Assume that $T_q(\ker \Phi) \subseteq \ker \Phi$ for all $q \in Q^\wedge$.*

Under this assumption, we can prove that a feature-level conjunction of two queries is fully characterized by a bilinear operator.

Lemma 2 (Proof in Sec. F)**.** *Assume Assumption 1 and the symmetry of conjunction (Def. 4(ii)). Then, there exists a unique symmetric bilinear operator* $\tilde{F} : W^\wedge \times W^\wedge \to W^\wedge$ *such that* $\tilde{F}(\phi_p, \phi_q) = \phi_{p\wedge q}$ *for all* $p, q \in Q^\wedge$.

The theorem below demonstrates that no non-trivial bilinear feature map can satisfy idempotence.

Theorem 2. *Let* $\tilde{F}$ *be the bilinear map from Lemma 2. If* ϕ *satisfies idempotence (Def. 4) and negation equivariance (*$\phi_{\neg p} = -\phi_p$*), then* $\phi_q = 0$ *for all* $q \in Q^\wedge$.

Proof. Let $p \in Q$ and $u = \phi_p$. From idempotence, $\tilde{F}(u, u) = \phi_{p\wedge p} = \phi_p = u$. By negation equivariance, $\phi_{\neg p} = -u$. Since $\neg p \in Q^\wedge$, we may apply idempotence to $\neg p$ as well: $\tilde{F}(-u, -u) = \tilde{F}(\phi_{\neg p}, \phi_{\neg p}) = \phi_{\neg p\wedge\neg p} = \phi_{\neg p} = -u$. On the other hand, since $\tilde{F}$ is bilinear, $\tilde{F}(-u, -u) = (-1)(-1)\tilde{F}(u, u) = \tilde{F}(u, u) = u$. Comparing the two results, we have $-u = u$, which implies $u = 0$. Thus, $\phi_p = 0$ for all atomic queries. Since any compound query $q \in Q^\wedge$ is formed by finite conjunctions of atomic queries and $\tilde{F}(0, \cdot) = 0$, by induction, $\phi_q = 0$ for all $q \in Q^\wedge$. ☐

This result implies that, in the first-order regime, the geometries required for negation and conjunction are incompatible. Importantly, this obstruction is structural rather than an optimization failure.

5 Discussion and Conclusion

A central insight of our work is that the structure of knowledge is best observed through its *dynamics*, i.e., how representations coordinate under first-order updates. While expressive networks may statically realize a logically coherent representation, our results (Sec. 3, Sec. 4) show that preserving such coherence under linearized updates imposes strict geometric constraints that are invisible to static analyses. This distinction between expressivity and systematic propagation provides a geometric explanation for when logical structure can be realized under first-order updates.

Implications for Systematicity. As discussed in Sec. A, tensor products were proposed as a sufficient mechanism for variable binding. Our results sharpen this connection in the linearized update regime: requiring systematic linear propagation forces a blockwise tensor factorization of the representation (Sec. 3). In this sense, our analysis suggests that such a mechanism is not only an architectural choice, but can be a geometric necessity for preserving logical structure under first-order dynamics. Moreover, our converse analysis (Sec. B) shows that systematic propagation constrains how argument order is encoded, requiring sufficient positional structure to remain consistent under reversal. Relatedly, even when a concept is linearly decodable at a fixed parameter state, static linearity alone does not guarantee systematic interaction under local updates.

Toward Logical Geometric Deep Learning. Our results align with geometric deep learning by casting systematic propagation as an equivariance constraint of logical operations. This motivates a direction toward *logical* geometric deep learning, where update-time behavior is constrained by these symmetries beyond static function approximation. That is, learning dynamics should remain near a constrained manifold defined by logical equivariance, instead of drifting into regions where logically related queries become entangled. Architecturally, this points to parameterizations and objectives that enforce the structures required by systematic propagation, for example, by regularizing gradient features or by building equivariant modules that preserve these symmetries across updates.

Implications for Model Adaptation and Editing. This geometric perspective has direct implications for practical model adaptation, most notably knowledge editing. Locate-and-edit methods assume that knowledge can be localized and manipulated through a single local update, but our results emphasize that the relevant structure is defined by how representations transform under that update. This shifts attention from *where* a fact is stored to *whether* the representation supports stable, structure-preserving update directions, i.e., *editability* as a geometric property (Sinitsin et al., 2020). At the same time, our conjunction result suggests that approaches relying on local linearity may face fundamental limits when asked to propagate edits to compositional consequences. This points to mechanisms beyond single-step locality, such as iterative nonlinear updates or memory-based updates that bypass single-update bottlenecks (Mitchell et al., 2022; Wang et al., 2024b).

REFERENCES

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

Fazl Barez, Tingchen Fu, Ameya Prabhu, Stephen Casper, Amartya Sanyal, Adel Bibi, Aidan O'Gara, Robert Kirk, Ben Bucknall, Tim Fist, Luke Ong, Philip H. S. Torr, Kwok-Yan Lam, Robert Trager, David Krueger, Sören Mindermann, José Hernández-Orallo, Mor Geva, and Yarin Gal. Open problems in machine unlearning for ai safety. *ArXiv*, abs/2501.04952, 2025. URL `https://api.semanticscholar.org/CorpusID:275405338`.

David Bau, Steven Liu, Tongzhou Wang, Jun-Yan Zhu, and Antonio Torralba. Rewriting a deep generative model. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

Lukas Berglund, Meg Tong, Max Kaufmann, Mikita Balesni, Asa Cooper Stickland, Tomasz Korbak, and Owain Evans. The reversal curse: Llms trained on" a is b" fail to learn" b is a". *arXiv preprint arXiv:2309.12288*, 2023.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Hoyeon Chang, Jinho Park, Seonghyeon Ye, Sohee Yang, Youngkyung Seo, Du-Seong Chang, and Minjoon Seo. How do large language models acquire factual knowledge during pretraining? In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL `https://openreview.net/forum?id=TYdzj1EvBP`.

Hoyeon Chang, Jinho Park, Hanseul Cho, Sohee Yang, Miyoung Ko, Hyeonbin Hwang, Seungpil Won, Dohaeng Lee, Youbin Ahn, and Minjoon Seo. Characterizing pattern matching and its limits on compositional task structures, 2025. URL `https://arxiv.org/abs/2505.20278`.

Roi Cohen, Eden Biran, Ori Yoran, Amir Globerson, and Mor Geva. Evaluating the ripple effects of knowledge editing in language models. *Transactions of the Association for Computational Linguistics*, 12:283–298, 2024.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL `https://proceedings.mlr.press/v48/cohenc16.html`.

David Steven Dummit, Richard M Foote, et al. *Abstract algebra*, volume 3. Wiley Hoboken, 2004.

Nouha Dziri, Ximing Lu, Melanie Sclar, Xiang Lorraine Li, Liwei Jiang, Bill Yuchen Lin, Sean Welleck, Peter West, Chandra Bhagavatula, Ronan Le Bras, Jena D. Hwang, Soumya Sanyal, Xiang Ren, Allyson Ettinger, Zaid Harchaoui, and Yejin Choi. Faith and fate: Limits of transformers on compositionality. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL `https://openreview.net/forum?id=Fkckkr3ya8`.

Nelson Elhage, Tristan Hume, Catherine Olsson, Nicholas Schiefer, Tom Henighan, Shauna Kravec, Zac Hatfield-Dodds, Robert Lasenby, Dawn Drain, Carol Chen, et al. Toy models of superposition. *arXiv preprint arXiv:2209.10652*, 2022.

Jerry A Fodor and Zenon W Pylyshyn. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3–71, 1988.

Mor Geva, Roei Schuster, Jonathan Berant, and Omer Levy. Transformer feed-forward layers are key-value memories. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 5484–5495, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.446. URL `https://aclanthology.org/2021.emnlp-main.446/`.

Steven Givant. The calculus of relations as a foundation for mathematics. *Journal of Automated Reasoning*, 37(4):277–322, 2006.

Klaus Greff, Sjoerd Van Steenkiste, and Jürgen Schmidhuber. On the binding problem in artificial neural networks. *arXiv preprint arXiv:2012.05208*, 2020.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. Don't stop pretraining: Adapt language models to domains and tasks. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8342–8360, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.740. URL `https://aclanthology.org/2020.acl-main.740/`.

Peter Hase, Thomas Hofweber, Xiang Zhou, Elias Stengel-Eskin, and Mohit Bansal. Fundamental problems with model editing: How should rational belief revision work in llms? *arXiv preprint arXiv:2406.19354*, 2024.

Chenhui Hu, Pengfei Cao, Yubo Chen, Kang Liu, and Jun Zhao. Knowledge in superposition: Unveiling the failures of lifelong knowledge editing for large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 24086–24094, 2025.

Arthur Jacot, Franck Gabriel, and Clement Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL `https://proceedings.neurips.cc/paper_files/paper/2018/file/5a4be1fa34e62bb8a6ec6b91d2462f5a-Paper.pdf`.

Joel Jang, Dongkeun Yoon, Sohee Yang, Sungmin Cha, Moontae Lee, Lajanugen Logeswaran, and Minjoon Seo. Knowledge unlearning for mitigating privacy risks in language models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 14389–14408, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.805. URL `https://aclanthology.org/2023.acl-long.805/`.

Nora Kassner and Hinrich Schütze. Negated and misprimed probes for pretrained language models: Birds can talk, but cannot fly. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, July 2020. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/2020.acl-main.698`.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In *ICML 2022 Workshop on Knowledge Retrieval and Language Models*, 2022. URL `https://openreview.net/forum?id=6p3AuaHAFiN`.

Brenden Lake and Marco Baroni. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. In *International conference on machine learning*, pp. 2873–2882. PMLR, 2018.

Jaehoon Lee, Lechao Xiao, Samuel Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. *Advances in neural information processing systems*, 32, 2019.

Wei Liu, Haomei Xu, Bingqing Liu, Zhiying Deng, Haozhao Wang, Jun Wang, Ruixuan Li, Yee Whye Teh, and Wee Sun Lee. Is model editing built on sand? revealing its illusory success and fragile foundation, 2025. URL `https://arxiv.org/abs/2510.00625`.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. Locating and editing factual associations in gpt. *Advances in neural information processing systems*, 35:17359–17372, 2022a.

Kevin Meng, Arnab Sen Sharma, Alex Andonian, Yonatan Belinkov, and David Bau. Mass-editing memory in a transformer. *arXiv preprint arXiv:2210.07229*, 2022b.

Eric Mitchell, Charles Lin, Antoine Bosselut, Christopher D Manning, and Chelsea Finn. Memory-based model editing at scale. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 15817–15831. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/mitchell22a.html.

Kiho Park, Yo Joong Choe, and Victor Veitch. The linear representation hypothesis and the geometry of large language models. In *Causal Representation Learning Workshop at NeurIPS 2023*, 2023. URL https://openreview.net/forum?id=T0PoOJg8cK.

Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2463–2473, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1250. URL https://aclanthology.org/D19-1250/.

Jean-Pierre Serre et al. *Linear representations of finite groups*, volume 42. Springer, 1977.

Gila Sher. Tarski's thesis. In *New Essays on Tarski and Philosophy*. Oxford University Press, 09 2008. ISBN 9780199296309. doi: 10.1093/acprof:oso/9780199296309.003.0012. URL https://doi.org/10.1093/acprof:oso/9780199296309.003.0012.

Anton Sinitsin, Vsevolod Plokhotnyuk, Dmitry Pyrkin, Sergei Popov, and Artem Babenko. Editable neural networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HJedXaEtvS.

Paul Smolensky. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, 46(1-2):159–216, 1990.

Alfred Tarski and John Corcoran. What are logical notions? *History and philosophy of logic*, 7(2): 143–154, 1986.

Boshi Wang and Huan Sun. Is the reversal curse a binding problem? uncovering limitations of transformers from a basic generalization failure, 2025. URL https://arxiv.org/abs/2504.01928.

Boshi Wang, Xiang Yue, Yu Su, and Huan Sun. Grokking of implicit reasoning in transformers: A mechanistic journey to the edge of generalization. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024a. URL https://openreview.net/forum?id=D4QgSWxiOb.

Peng Wang, Zexi Li, Ningyu Zhang, Ziwen Xu, Yunzhi Yao, Yong Jiang, Pengjun Xie, Fei Huang, and Huajun Chen. WISE: Rethinking the knowledge memory for lifelong model editing of large language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b. URL https://openreview.net/forum?id=VJMYOfJVC2.

Tongtong Wu, Linhao Luo, Yuan-Fang Li, Shirui Pan, Thuy-Trang Vu, and Gholamreza Haffari. Continual learning for large language models: A survey, 2024. URL https://arxiv.org/abs/2402.01364.

A Related Work

Linearized Dynamics and Model Adaptation. A common lens in modern deep learning is to approximate adaptation via local updates in a linearized feature space, a perspective theoretically grounded in Neural Tangent Kernel (NTK) analyses (Jacot et al., 2018; Lee et al., 2019). This viewpoint arises across settings that rely on gradient-based updates, including pretraining trajectories that accumulate factual associations (Chang et al., 2024), as well as domain adaptation and continual learning where updates interact with prior knowledge (Gururangan et al., 2020; Wu et al., 2024). Most explicitly, knowledge editing methods such as ROME and MEMIT treat Transformer MLPs as key-value memories (Geva et al., 2021), implementing edits as constrained least-squares updates (Meng et al., 2022a;b). Related first-order schemes also appear in unlearning approaches aimed at reducing the influence of specific data (Jang et al., 2023; Barez et al., 2025). A recurring empirical theme is that local updates do not reliably generalize to logically related or compositional variants of the target behavior (Cohen et al., 2024; Liu et al., 2025), echoing broader concerns about the gap between optimization and belief revision (Hase et al., 2024).

Systematicity and Variable Binding. The systematicity debate argues that connectionist models may lack the structural machinery for compositional generalization and variable binding (Fodor & Pylyshyn, 1988; Lake & Baroni, 2018). While Smolensky (1990) proposed Tensor Product Representations as a mechanism for variable binding, the binding problem remains an active challenge in modern deep learning (Greff et al., 2020). Recent interpretability work studies linear representations in neural activations (Park et al., 2023), yet empirical studies report persistent failures of compositional generalization in LLMs (Dziri et al., 2023; Wang et al., 2024a; Chang et al., 2025). In particular, Wang & Sun (2025) hypothesized that failures of variable binding underlie the reversal curse. These threads motivate analyzing what structures are required for systematic behavior.

Logical Structure as Geometric Invariance. To formalize logical structure in vector spaces, we adopt invariance-based views of logical notions (Tarski & Corcoran, 1986; Sher, 2008). This aligns with geometric deep learning, which characterizes representations by invariance and equivariance (Bronstein et al., 2021; Cohen & Welling, 2016). This perspective motivates treating logical operations as transformations on queries and studies the constraints they induce in a linearized feature geometry.

B Converse Equivariance Forces Positional Alignment

Having established that logical equivariance under negation forces a blockwise tensor factorization, we now investigate the structural implications of the converse operation. Recall that the symbolic converse operation swaps the head and tail entities and replaces r with its converse $r^{\smile}$, i.e., $\mathrm{rev}(h, r, t) = (t, r^{\smile}, h)$. SLP requires the feature map to be invariant under this operation, i.e., $\phi_{\mathrm{rev}(q)} = \phi_q$. This constraint forces a parity alignment between the context and relation components, as stated in the following theorem.

Theorem 3 (Symmetric-Antisymmetric Alignment (Proof in Sec. E))**.** *Assume the context-relation factorization from Theorem 1 and suppose ϕ is* converse-invariant*, i.e.*

$$\phi(h, r, t) = \phi(t, r^{\smile}, h) \qquad \textit{for all } (h, r, t) \in Q.$$

Then for each block i, the corresponding feature component ϕ_i admits a decomposition with matched-parity:

$$\phi_i(h, r, t) = \phi_i^{+}(h, r, t) + \phi_i^{-}(h, r, t),$$

where $\phi_i^{\pm}$ can be chosen as a sum of context-relation tensor terms $u(h, t) \otimes v(r)$ satisfying

$$u(t, h) = \pm u(h, t) \quad \textit{and} \quad v(r^{\smile}) = \pm v(r).$$

This result reveals a geometric account of how directionality can be represented under converse invariance. The first term represents symmetric components, where neither the relation nor the entity pair cares about order. In contrast, the second term $u^{-} \otimes v^{-}$ encodes directionality through a mechanism of **sign cancellation**: swapping the entities introduces a sign flip (-1) in the context factor $u(h, t)$, which is exactly compensated by the sign flip in the relation factor $v(r)$ induced by the converse operation.

Note that the constraints from negation and converse are compatible, as the corresponding symbolic operations commute. In conclusion, while SLP is theoretically realizable through this specific symmetry, such a structure is likely absent in the unconstrained feature spaces of practical LLMs.

C PRIMER ON FINITE GROUP REPRESENTATIONS

This section collects basic definitions from group theory (Dummit et al., 2004) and group representation theory (Serre et al., 1977) used in our proofs. We work over real vector spaces unless otherwise stated, and all groups considered in this paper are finite.

C.1 GROUPS

A *group* is a set G equipped with a binary operation $(g, h) \mapsto gh$ that encodes the structure of symmetries. Formally, it must satisfy three axioms:

- **Associativity:** $(gh)k = g(hk)$ for all $g, h, k \in G$.
- **Identity:** There exists an element $e \in G$ such that $eg = ge = g$ for all g.
- **Inverses:** For every $g \in G$, there exists an element $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$.

We introduce two standard families of groups:

- **Symmetric group (S_n):** The set of all permutations (bijections) of a set of n elements forms a group under function composition.
- **Cyclic group ($\mathbb{Z}_n$):** A group generated by a single element g such that $g^n = e$ (and $g^k \neq e$ for $k < n$). We specifically use the cyclic group of order 2, $\mathbb{Z}_2 \cong (\{+1, -1\}, \times)$, to model logical negation, where the non-identity element corresponds to the negation operator ($(-1)^2 = 1$).

C.2 GROUP ACTIONS

A (left) *action* of a group G on a set X formalizes the idea of transforming elements of X using the symmetries in G. It is a map $G \times X \to X$, denoted $(g, x) \mapsto g \cdot x$, that obeys two consistency rules:

- **Identity preservation:** $e \cdot x = x$ for all $x \in X$ (the identity transformation leaves everything unchanged).
- **Compatibility:** $g \cdot (h \cdot x) = (gh) \cdot x$ for all $g, h \in G$ and $x \in X$ (transforming by h then g is equivalent to transforming by the combined element gh).

Given an element $x \in X$, its *orbit* is the set of all points reachable from x under the group action: $G \cdot x := \{g \cdot x : g \in G\}$.

C.3 PRODUCT GROUPS

Let G and H be finite groups. The *direct product* $G \times H$ is the group of pairs (g, h) defined by the component-wise operation:

$$(g, h)(g', h') \;=\; (gg', hh').$$

This structure allows us to compose symmetries acting on different domains. Specifically, if G acts on a set X and H acts on a set Y, then the product group $G \times H$ naturally acts on the product set $X \times Y$ via

$$(g, h) \cdot (x, y) \;:=\; (g \cdot x, h \cdot y).$$

C.4 REPRESENTATIONS AND EQUIVARIANT MAPS

Let W be a real vector space. A (finite-dimensional) *representation* of G on W is a group homomorphism

$$\rho : G \to GL(W),$$

where $GL(W)$ denotes the group of invertible linear maps $W \to W$. This definition allows us to view G as *acting linearly* on the vector space W. We denote this action by $g \cdot w := \rho(g)w$. Since each $\rho(g)$ lies in $GL(W)$, this action inherently preserves the linear structure: $g \cdot (aw + bv) = a(g \cdot w) + b(g \cdot v)$ for all scalars a, b and vectors w, v.

A linear map $T : W \to W'$ between two G-representations (W, ρ) and (W', ρ') is called *G-equivariant* if it commutes with the group action:

$$T(g \cdot w) \;=\; g \cdot T(w) \qquad \text{for all } g \in G,\; w \in W.$$

In terms of the homomorphism ρ, this is equivalent to $T(\rho(g)w) = \rho'(g)T(w)$. The space of all such G-equivariant linear maps is denoted $\mathrm{Hom}_G(W, W')$.

C.5 Invariant Subspaces and Irreducibility

A subspace $U \subseteq W$ is called *G-invariant* if it is closed under the group action:

$$\rho(g)u \in U \quad \text{for all } g \in G, u \in U.$$

Intuitively, if a feature vector lies in a G-invariant subspace, applying any symmetry from G will keep the vector within that same subspace. In this case, the restriction $\rho|_U : G \to GL(U)$ defines a valid representation on U, called a *subrepresentation*.

A representation is *irreducible* if it contains no proper nontrivial invariant subspaces. That is, the only G-invariant subspaces are $\{0\}$ and W itself. A representation is *reducible* if it has a nontrivial invariant subspace.

A representation W is *completely reducible* if it can be decomposed entirely into these atomic blocks. Formally, this means W is isomorphic to a direct sum of irreducible subrepresentations:

$$W \;\cong\; \bigoplus_{i=1}^{\ell} W_i,$$

where each W_i is irreducible.

C.6 Maschke's Theorem

A key structural fact we use is that finite-dimensional representations of finite groups over $\mathbb{R}$ are completely reducible (Maschke's theorem). While the theorem holds for more general fields, we focus on the real field relevant to our work.

Theorem 4 (Maschke)**.** *Let G be a finite group and let W be a finite-dimensional representation of G over $\mathbb{R}$. Then W is completely reducible: every G-invariant subspace $U \subseteq W$ admits a G-invariant complement U' such that $W = U \oplus U'$.*

We note that by recursively applying this property, i.e., finding an invariant subspace, splitting it off, and repeating the process on the complement, one can guarantee that any finite-dimensional representation W can be fully decomposed into a direct sum of irreducible subrepresentations.

D Proof of Theorem 1

In this section, we provide the complete proof of Theorem 1. We proceed in four logical steps:

1. **Geometric Constraints:** First, we analyze the linear dependencies in the feature space, showing that the kernel of the feature map decomposes according to logical families under SLP (Lemma 3).
2. **Group Action Construction:** Building on this kernel structure, we translate the symbolic symmetries of entity renaming and negation to a well-defined linear representation of the group H on the feature space (Lemmas 6 and 7).
3. **Representation Theoretic Decomposition:** We invoke standard results from representation theory to decompose the feature space into a direct sum of tensor products (Lemma 8).
4. **Derivation of Factorization:** Finally, we utilize the isomorphism of equivariant maps on tensor products (Lemma 9) to derive the explicit context-relation factorization of the feature map (Theorem 1).

D.1 Logical Families and Controllability

Let V be the free real vector space with basis $\{e_q : q \in Q\}$, where e_q is a basis vector corresponding to the query q. Define a linear map

$$\Phi : V \longrightarrow W, \qquad \Phi(e_q) := \phi_q.$$

By construction, $\ker \Phi$ is the space of all linear dependencies among the feature vectors $\{\phi_q\}$. Moreover, recall from Sec. 2.3 that the two logical operations in the unary relation algebra generate a group $G := \{\, \mathrm{id},\ \neg,\ \mathrm{rev},\ \neg \circ \mathrm{rev} \,\}$ acting on Q, and logical family is defined as:

$$G \cdot q := \{\, g(q) : g \in G \,\}.$$

Intuitively, the second condition of SLP (Def. 3) states that no unintended collapse happens *across* different logical families. In other words, linear dependencies arise if and only if they are semantically forced *within* a family. The following lemma states this formally.

Lemma 3 (Family-wise kernel decomposition)**.** *Let F be a logical family and assume SLP (Def. 3). Let $V(F) := \mathrm{span}\{e_q : q \in F\} \subseteq V$ and let $\Phi|_{V(F)}$ denote the restriction of Φ to $V(F)$. Then any linear relation among features decomposes family-wise, i.e.,*

$$\ker \Phi \;=\; \bigoplus_F \ker \Phi|_{V(F)}.$$

Proof. The logical families $\{F\}$ form a partition of Q, so the basis $\{e_q : q \in Q\}$ of V splits accordingly into disjoint subsets $\{e_q : q \in F\}$. Hence

$$V \;=\; \bigoplus_F V(F),$$

and every $v \in V$ can be written uniquely as $v = \sum_F v_F$ with $v_F \in V(F)$.

By logical equivariance (Def. 2), for each family F, we can choose a representative $q_F^* \in F$ such that, for every $q \in F$, the feature vector ϕ_q is either $\phi_{q_F^*}$ or $-\phi_{q_F^*}$. Equivalently, for each $q \in F$ there exists a scalar $\lambda_{q,q^*} \in \{+1, -1\}$ with

$$\phi_q \;=\; \lambda_{q,q^*}\, \phi_{q_F^*}.$$

By the second condition of SLP (Def. 3), the chosen representatives can be taken so that the set $\{\phi_{q_F^*} : F \text{ is a logical family}\}$ is linearly independent in W.

Now take any $v \in \ker \Phi$ and decompose it as $v = \sum_F v_F$ with $v_F \in V(F)$. Write each component as

$$v_F \;=\; \sum_{q \in F} \alpha_q\, e_q.$$

Then

$$\Phi(v_F) \;=\; \sum_{q \in F} \alpha_q\, \phi_q \;=\; \sum_{q \in F} \alpha_q\, \lambda_{q,q^*}\, \phi_{q_F^*} \;=\; \beta_F\, \phi_{q_F^*},$$

where we define

$$\beta_F := \sum_{q \in F} \alpha_q\, \lambda_{q,q^*}.$$

Since $v \in \ker \Phi$, we have

$$0 \;=\; \Phi(v) \;=\; \sum_F \Phi(v_F) \;=\; \sum_F \beta_F\, \phi_{q_F^*}.$$

The vectors $\{\phi_{q_F^*}\}_F$ are linearly independent, so each coefficient must vanish:

$$\beta_F = 0 \quad \text{for all } F.$$

Hence $\Phi(v_F) = \beta_F \phi_{q_F^*} = 0$ for every family F, i.e. $v_F \in \ker \Phi|_{V(F)}$. Since $v = \sum_F v_F$ and the decomposition $V = \bigoplus_F V(F)$ is unique, this shows that every element of $\ker \Phi$ decomposes uniquely as a sum of elements from the subspaces $\ker \Phi|_{V(F)}$, and therefore

$$\ker \Phi \;=\; \bigoplus_F \ker \Phi|_{V(F)}.$$

□

D.2 DIAGONAL RENAMING OF ENTITIES

We now formalize the idea that logical notions should be insensitive to the names of entities.

Definition 5 (Entity renaming)**.** *Let $G_E := \mathrm{Sym}(E)$ be the permutation group of E. We let G_E act on Q by renaming entities uniformly in both argument slots:*

$$\sigma \cdot (h, r, t) \;:=\; (\sigma(h),\, r,\, \sigma(t)) \quad (\sigma \in G_E,\ (h, r, t) \in Q).$$

An important observation is that, since entity renaming permutes only entities and logical operations act only on relations, **the two operations commute**.

Lemma 4 (Symbolic renaming invariance)**.** *For any $\sigma \in G_E$ and any $q \in Q$ we have*

$$\neg(\sigma \cdot q) \;=\; \sigma \cdot \neg(q), \qquad \mathrm{rev}(\sigma \cdot q) \;=\; \sigma \cdot \mathrm{rev}(q).$$

Proof. This is the direct consequence of the definition of logical operations in the tiny relation algebra and Def. 5. ☐

In other words, if two queries are in the same logical family, then their renamings are in the same family.

Our goal is to show that this renaming symmetry *translates* to a linear symmetry of W. To this end, we first show that the kernel structure is preserved upon entity renaming. For each $\sigma \in G_E$ define a linear map $P_\sigma : V \to V$ on basis vectors by

$$P_\sigma e_q := e_{\sigma \cdot q}.$$

Lemma 5 (Kernel invariance under renaming)**.** *For every $\sigma \in G_E$, if $v \in \ker \Phi$, then $P_\sigma v \in \ker \Phi$.*

Proof. Let $v \in \ker \Phi$. By Lemma 3, it suffices to show the invariance for a component $v_F \in \ker \Phi|_{V(F)}$ entirely contained in a single logical family F. By the logical equivariance assumption, the feature vectors within a family F are all collinear. Specifically, fix a representative $q^* \in F$. Then for any $q \in F$, there exists $\lambda_{q,q^*} \in \{+1, -1\}$ such that

$$\phi_q \;=\; \lambda_{q,q^*}\, \phi_{q^*}.$$

Consequently, a vector $v_F = \sum_{q \in F} \alpha_q e_q$ lies in $\ker \Phi$ if and only if

$$\Phi(v_F) \;=\; \sum_{q \in F} \alpha_q \phi_q \;=\; \left(\sum_{q \in F} \alpha_q \lambda_{q,q^*} \right) \phi_{q^*} \;=\; 0.$$

Since $\phi_{q^*} \neq 0$ (by the second condition of SLP Def. 3), the condition for membership in the kernel is purely scalar:

$$\sum_{q \in F} \alpha_q\, \lambda_{q,q^*} \;=\; 0. \tag{1}$$

Now consider the transformed vector $P_\sigma v_F = \sum_{q \in F} \alpha_q e_{\sigma \cdot q}$. This vector is supported on the permuted family $\sigma(F)$. Let $p^* := \sigma \cdot q^*$ be the representative for $\sigma(F)$. For any $q \in F$, let $p := \sigma \cdot q$. By definition of a logical family (orbit under the logical-operation group G generated by $\neg$ and rev), for any $q \in F$ there exists $g \in G$ such that

$$q = g \cdot q^*.$$

We claim that entity renaming commutes with every $g \in G$:

$$\sigma \cdot (g \cdot x) \;=\; g \cdot (\sigma \cdot x) \qquad \forall\, \sigma \in G_E,\ \forall\, g \in G,\ \forall\, x \in Q. \tag{2}$$

Indeed, Lemma 4 gives this commutation for the generators $\neg$ and rev, and Eq. (2) follows for arbitrary g by closure under composition in G.

Applying Eq. (2) to $x = q^*$ yields

$$p := \sigma \cdot q = \sigma \cdot (g \cdot q^*) = g \cdot (\sigma \cdot q^*) =: g \cdot p^*,$$

so the same g relates p to p^*.

Next, by the logical equivariance condition, there is a sign homomorphism $\chi : G \to \{+1, -1\}$ such that for all $x \in Q$,

$$\phi_{g \cdot x} \;=\; \chi(g)\, \phi_x. \tag{3}$$

Using Eq. (3) with $x = q^*$ and $x = p^*$, we relate the scalar coefficients to χ:

$$\phi_q = \phi_{g \cdot q^*} = \chi(g)\phi_{q^*} \quad \Longrightarrow \quad \lambda_{q,q^*} = \chi(g),$$
$$\phi_p = \phi_{g \cdot p^*} = \chi(g)\phi_{p^*} \quad \Longrightarrow \quad \lambda_{p,p^*} = \chi(g).$$

Therefore, the relative signs are preserved:

$$\lambda_{p,p^*} \;=\; \chi(g) \;=\; \lambda_{q,q^*}.$$

Using the equality above, we finally show that $P_\sigma v_F \in \ker \Phi$:

$$\begin{aligned}\Phi(P_\sigma v_F) \;&=\; \sum_{q \in F} \alpha_q \phi_{\sigma \cdot q} \\ &=\; \sum_{q \in F} \alpha_q\, \lambda_{\sigma \cdot q, \sigma \cdot q^*}\, \phi_{p^*} \\ &=\; \left(\sum_{q \in F} \alpha_q\, \lambda_{q,q^*} \right) \phi_{p^*}.\end{aligned}$$

By equation Eq. (1), the coefficient sum is zero. Thus $\Phi(P_\sigma v_F) = 0$, proving that $P_\sigma v_F \in \ker \Phi$. □

Using Lemma 5, we can translate the entity renaming to a linear symmetry on the feature space.

Lemma 6. *Suppose that for each $\sigma \in G_E$ we have $P_\sigma(\ker \Phi) \subseteq \ker \Phi$. Then there exists a unique linear map $\rho_E(\sigma) : W \to W$ such that*

$$\phi_{\sigma \cdot q} \;=\; \rho_E(\sigma)\, \phi_q \quad \forall q \in Q.$$

Moreover, the assignment $\sigma \mapsto \rho_E(\sigma)$ defines a group representation $\rho_E : G_E \to GL(W)$.

Proof. Recall that $\Phi : V \to W$ is surjective by definition of $W = \operatorname{span}\{\phi_q : q \in Q\}$. For a fixed $\sigma \in G_E$, we define $\rho_E(\sigma)$ as follows: given any $w \in W$, choose $v \in V$ with $\Phi(v) = w$ and set

$$\rho_E(\sigma)\, w \;:=\; \Phi\big(P_\sigma v\big).$$

We first check that this does not depend on the choice of v. Suppose $v, v' \in V$ satisfy $\Phi(v) = \Phi(v')$, i.e. $\Phi(v - v') = 0$, so $v - v' \in \ker \Phi$. By assumption, $P_\sigma(\ker \Phi) \subseteq \ker \Phi$, hence

$$\Phi\big(P_\sigma(v - v')\big) = 0,$$

which implies

$$\Phi\big(P_\sigma v\big) = \Phi\big(P_\sigma v'\big).$$

Thus, $\rho_E(\sigma)\, w$ is well-defined. Linearity of $\rho_E(\sigma)$ follows immediately from the linearity of Φ and P_σ.

Now, we apply this definition to the basis elements. For $q \in Q$, we have $e_q \in V$ and $\Phi(e_q) = \phi_q$, so

$$\rho_E(\sigma)\, \phi_q \;=\; \rho_E(\sigma)\, \Phi(e_q) \;=\; \Phi\big(P_\sigma e_q\big) \;=\; \Phi(e_{\sigma \cdot q}) \;=\; \phi_{\sigma \cdot q}.$$

To check its uniqueness, observe that the vectors $\{\phi_q : q \in Q\}$ span W, so any linear map $T : W \to W$ satisfying $T\, \phi_q = \phi_{\sigma \cdot q}$ for all q must agree with $\rho_E(\sigma)$ on a spanning set, hence must be equal to $\rho_E(\sigma)$.

Finally, we verify the group property. For any $\sigma_1, \sigma_2 \in G_E$ and any, $w = \Phi(v)$ we have

$$\rho_E(\sigma_1)\, \rho_E(\sigma_2)\, w \;=\; \rho_E(\sigma_1)\, \Phi\big(P_{\sigma_2} v\big) \;=\; \Phi\big(P_{\sigma_1} P_{\sigma_2} v\big) \;=\; \Phi\big(P_{\sigma_1 \sigma_2} v\big) \;=\; \rho_E(\sigma_1 \sigma_2)\, w.$$

Thus, $\rho_E(\sigma_1)\rho_E(\sigma_2) = \rho_E(\sigma_1\sigma_2)$ on all of W, and $\rho_E : G_E \to GL(W)$ is a representation. □

Hence, SLP implies that the linearized features carry a compatible entity-renaming symmetry.

D.3 COMBINING RENAMING AND NEGATION

We now combine entity renaming with relation negation into a single symmetry structure. Recall from Def. 2 that logical equivariance of negation says

$$\phi_{\neg q} = -\,\phi_q \quad \forall q \in Q,$$

i.e., relation-level negation always flips the feature vector by a global sign -1, regardless of which entities appear.

Entity renaming acts only on the head and tail slots, while negation acts only on the relation slot. We want to treat these two operations together as a single family of joint transformations.

Definition 6 (Combined symmetry group). *Let $\mathbb{Z}_2 := \{+1, -1\}$ be the two-element group with multiplication. We define*

$$H := G_E \times \mathbb{Z}_2.$$

An element $(\sigma, \epsilon) \in H$ acts on a query by

$$(\sigma, \epsilon) \cdot (h, r, t) := \big(\sigma(h),\, \epsilon \cdot r,\, \sigma(t)\big),$$

where $\epsilon = +1$ means "keep the relation as it is" and $\epsilon = -1$ means "replace r by its negation $\neg r$".

On the feature space W we let (σ, ϵ) act linearly by

$$\rho(\sigma, +1) := \rho_E(\sigma), \qquad \rho(\sigma, -1) := -\,\rho_E(\sigma),$$

so that $\rho(\sigma, \epsilon)$ first applies the renaming operator $\rho_E(\sigma)$ and then, if $\epsilon = -1$, flips the sign of the feature vector. In other words, for any $w \in W$,

$$\rho(\sigma, \epsilon)\, w := \begin{cases} \rho_E(\sigma)\, w & \text{if } \epsilon = +1, \\ -\,\rho_E(\sigma)\, w & \text{if } \epsilon = -1. \end{cases}$$

Because ρ_E respects composition of renamings and $(-1)^2 = +1$, the maps $\rho(\sigma, \epsilon)$ also respect composition: applying (σ_1, ϵ_1) and then (σ_2, ϵ_2) has the same effect on features as applying $(\sigma_1\sigma_2, \epsilon_1\epsilon_2)$ once. This means H acts consistently and linearly on W. Hence, logical equivariance and renaming equivariance can now be combined into a single statement as below.

Lemma 7 (Equivariant feature map). *The map $\phi : Q \to W$ is compatible with the combined symmetry action of H in the sense that, for all $(\sigma, \epsilon) \in H$ and all $q \in Q$,*

$$\phi\big((\sigma, \epsilon) \cdot q\big) = \rho(\sigma, \epsilon)\, \phi_q.$$

Proof. When $\epsilon = +1$, we have

$$(\sigma, +1) \cdot (h, r, t) = (\sigma(h), r, \sigma(t)),$$

and this is represented on features by $\rho_E(\sigma)$, i.e. $\phi(\sigma \cdot q) = \rho_E(\sigma)\, \phi_q = \rho(\sigma, +1)\, \phi_q$.

When $\epsilon = -1$, we have

$$(\sigma, -1) \cdot (h, r, t) = (\sigma(h), \neg r, \sigma(t)).$$

By logical equivariance of negation, $\phi_{\neg q} = -\phi_q$ for every q, and by Lemma 4, negation commutes with renaming on the symbolic side. Combining these facts, we obtain

$$\phi\big((\sigma, -1) \cdot q\big) = \phi\big(\sigma \cdot (\neg q)\big) = \rho_E(\sigma)\, \phi_{\neg q} = \rho_E(\sigma)\, (-\phi_q) = -\,\rho_E(\sigma)\, \phi_q = \rho(\sigma, -1)\, \phi_q,$$

as claimed. □

In words, performing a renaming and optional negation on the symbolic query side has the same effect as applying the corresponding linear operator $\rho(\sigma, \epsilon)$ on the feature side.

D.4 Decomposition of H-representations

Using the Lemmas above, we prove Lemma 8, which will play a crucial role in proving Theorem 1.

Lemma 8 (Decomposition of H-representations)**.** *Let W be a finite-dimensional real representation of $H = G_E \times \mathbb{Z}_2$. Then W decomposes into a direct sum of tensor-product blocks:*

$$W \;\cong\; \bigoplus_{i=1}^{m} (C_i \otimes R_i),$$

where each C_i is an irreducible representation of the entity group G_E, and each R_i is an irreducible representation of the logical group $\mathbb{Z}_2$.

Proof. We explicitly construct the decomposition using the structure of $\mathbb{Z}_2 = \{1, -1\}$. Let $z := (\mathrm{id}, -1) \in H$. Since z commutes with every element in H (i.e., $zh = hz$ for all h), the linear map $\rho(z)$ must commute with the group action:

$$\rho(z)\rho(h) \;=\; \rho(zh) \;=\; \rho(hz) \;=\; \rho(h)\rho(z) \quad \text{for all } h \in H.$$

Also, since $z^2 = (\mathrm{id}, 1) = e$, we have $\rho(z)^2 = I$. Thus, $\rho(z)$ is an involution and W can be decomposed into eigenspaces corresponding to eigenvalues $+1$ and -1:

$$W \;=\; W^+ \oplus W^-, \quad \text{where } W^\pm := \{w \in W : \rho(z)w = \pm w\}.$$

Crucially, these eigenspaces are H-invariant. To see this, let $w \in W^\pm$ and $h \in H$. Using the commutativity shown above:

$$\rho(z)\big(\rho(h)w\big) \;=\; \rho(h)\big(\rho(z)w\big) \;=\; \rho(h)(\pm w) \;=\; \pm\big(\rho(h)w\big).$$

This shows that if $w \in W^\pm$, then the transformed vector $\rho(h)w$ also satisfies the condition to be in $W^\pm$. Thus, W^+ and W^- are subrepresentations of H.

We now determine the action of a general element $(\sigma, \epsilon) \in H$ on these subspaces. Note that, any element factors as $(\sigma, \epsilon) = (\sigma, 1)(\mathrm{id}, \epsilon)$. Then, the logical factor (id, ϵ) acts on the subspaces as:

$$\rho(\mathrm{id}, \epsilon)\big|_{W^+} = I, \qquad \rho(\mathrm{id}, \epsilon)\big|_{W^-} = \epsilon I.$$

Consequently, the full action is:

$$\rho(\sigma, \epsilon)w \;=\; \rho(\sigma, 1)\rho(\mathrm{id}, \epsilon)w \;=\; \begin{cases} \rho(\sigma, 1)w & \text{if } w \in W^+, \\ \epsilon \cdot \rho(\sigma, 1)w & \text{if } w \in W^-. \end{cases}$$

Since W is finite-dimensional, we can apply Maschke's Theorem (Theorem 4) to decompose W^+ and W^- into irreducible representations of G_E (via the map $\sigma \mapsto \rho(\sigma, 1)$):

$$W^+ \cong \bigoplus_{j \in J_+} C_j, \qquad W^- \cong \bigoplus_{j \in J_-} C_j.$$

Finally, we map this structure to the tensor product form claimed in the theorem. Let $\mathbf{1}$ and sgn denote the two irreducible real representations of $\mathbb{Z}_2$, respectively. Each irreducible G_E-summand $C \subseteq W^+$ is an H-subrepresentation on which (σ, ϵ) acts by $\rho(\sigma, 1)$ alone, hence it is H-isomorphic to $C \otimes \mathbf{1}$ (where $(\sigma, \epsilon) \cdot (c \otimes 1) = (\sigma c) \otimes 1$). Likewise, each irreducible G_E-summand $C \subseteq W^-$ is an H-subrepresentation on which (σ, ϵ) acts by $\epsilon\, \rho(\sigma, 1)$, hence it is H-isomorphic to $C \otimes \mathrm{sgn}$ (where $(\sigma, \epsilon) \cdot (c \otimes 1) = (\sigma c) \otimes \epsilon$). Therefore,

$$W \;\cong\; \left(\bigoplus_{j \in J_+} C_j \otimes \mathbf{1} \right) \;\oplus\; \left(\bigoplus_{j \in J_-} C_j \otimes \mathrm{sgn} \right).$$

Letting R_i represent either $\mathbf{1}$ or sgn as appropriate for each block, we obtain the claimed form $W \cong \bigoplus_i (C_i \otimes R_i)$. □

D.5 Equivariant Maps on Tensor Products

To prove the main theorem, we need to characterize equivariant maps between tensor product representations.

Lemma 9 (Equivariant maps on tensor products)**.** *Let G and K be groups. Let U, A be finite-dimensional real representations of G, and let V, B be finite-dimensional real representations of K. We regard the tensor products $U \otimes V$ and $A \otimes B$ as representations of the product group $G \times K$ via the component-wise action:*

$$(g,k)\cdot(u\otimes v) = (g\cdot u)\otimes(k\cdot v), \quad (g,k)\cdot(a\otimes b) = (g\cdot a)\otimes(k\cdot b).$$

Then there is a natural linear isomorphism

$$\mathrm{Hom}_G(U,A)\otimes\mathrm{Hom}_K(V,B) \;\cong\; \mathrm{Hom}_{G\times K}(U\otimes V,\; A\otimes B).$$

Proof. Consider first the full vector spaces of linear maps without equivariance constraints. For any linear maps $f \in \mathrm{Hom}(U,A)$ and $h \in \mathrm{Hom}(V,B)$, we define a map $f \boxtimes h : U\otimes V \to A\otimes B$ by its action on simple tensors:

$$(f\boxtimes h)(u\otimes v) := f(u)\otimes h(v) \quad \text{for all } u\in U, v\in V.$$

This construction induces a natural linear map

$$\Psi : \mathrm{Hom}(U,A)\otimes\mathrm{Hom}(V,B) \to \mathrm{Hom}(U\otimes V, A\otimes B), \qquad \Psi(f\otimes h) = f\boxtimes h.$$

We first claim that Ψ is an isomorphism. To see this, choose bases for the source and target vector spaces:

$$(u_i)_{i=1}^m \text{ of } U, \qquad (a_j)_{j=1}^p \text{ of } A,$$
$$(v_r)_{r=1}^n \text{ of } V, \qquad (b_s)_{s=1}^q \text{ of } B.$$

We define the elementary linear maps that form the bases for the hom-spaces. For each pair (j,i), let $E_{j,i} \in \mathrm{Hom}(U,A)$ be the unique linear map defined by:

$$E_{j,i}(u_k) = \begin{cases} a_j & \text{if } k = i, \\ 0 & \text{if } k \neq i. \end{cases}$$

The set $\{E_{j,i}\}_{j,i}$ forms a basis of $\mathrm{Hom}(U,A)$.

Similarly, for each (s,r), let $F_{s,r} \in \mathrm{Hom}(V,B)$ be defined by:

$$F_{s,r}(v_k) = \begin{cases} b_s & \text{if } k = r, \\ 0 & \text{if } k \neq r. \end{cases}$$

The set $\{F_{s,r}\}_{s,r}$ forms a basis of $\mathrm{Hom}(V,B)$. Consequently, the set of tensor products forms a basis for the domain of Ψ:

$$\{E_{j,i}\otimes F_{s,r}\}_{j,i,s,r} \quad \subset \quad \mathrm{Hom}(U,A)\otimes\mathrm{Hom}(V,B).$$

Now, consider the image of these basis vectors under Ψ. Let $M_{j,i,s,r} := \Psi(E_{j,i}\otimes F_{s,r})$. By the definition of Ψ (action on simple tensors), $M_{j,i,s,r}$ is the unique linear map $U\otimes V \to A\otimes B$ satisfying:

$$M_{j,i,s,r}(u_{i'}\otimes v_{r'}) = E_{j,i}(u_{i'})\otimes F_{s,r}(v_{r'}) = \begin{cases} a_j\otimes b_s & \text{if } (i',r') = (i,r), \\ 0 & \text{otherwise.} \end{cases}$$

The collection of maps $\{M_{j,i,s,r}\}_{j,i,s,r}$ is precisely the standard basis of the codomain space $\mathrm{Hom}(U\otimes V, A\otimes B)$. Since Ψ maps a basis of the domain bijectively onto a basis of the codomain, it is a linear isomorphism.

Our strategy is to identify equivariant maps as the fixed points of a group action on the function space. Specifically, we equip $\mathrm{Hom}(U,A)$ with a natural G-action via conjugation: for $g \in G$ and $f \in \mathrm{Hom}(U,A)$, the transformed map $g\cdot f$ is defined by

$$(g\cdot f)(u) := g\cdot f(g^{-1}\cdot u).$$

A map f is G-equivariant if and only if it is invariant under this action (i.e., $g \cdot f = f$ for all g), hence

$$\mathrm{Hom}_G(U, A) = \mathrm{Hom}(U, A)^G,$$

where the superscript G denotes the subspace of fixed points. Analogously, we define the K-action on $\mathrm{Hom}(V, B)$, and the $(G \times K)$-action on $\mathrm{Hom}(U \otimes V, A \otimes B)$: for any $T \in \mathrm{Hom}(U \otimes V, A \otimes B)$, the action is defined by

$$\big((g, k) \cdot T\big)(x) := (g, k) \cdot T\big((g, k)^{-1} \cdot x\big).$$

We now verify that the isomorphism Ψ respects these group structures. We equip the domain, $\mathrm{Hom}(U, A) \otimes \mathrm{Hom}(V, B)$, with the component-wise action of $G \times K$:

$$(g, k) \cdot (f \otimes h) := (g \cdot f) \otimes (k \cdot h).$$

For any $(g, k) \in G \times K$, a direct computation on simple tensors $u \otimes v$ shows:

$$\begin{aligned}
\big((g, k) \cdot (f \boxtimes h)\big)(u \otimes v) &= (g, k) \cdot (f \boxtimes h)\big((g, k)^{-1} \cdot (u \otimes v)\big) \\
&= (g, k) \cdot (f \boxtimes h)(g^{-1}u \otimes k^{-1}v) \\
&= (g, k) \cdot \big(f(g^{-1}u) \otimes h(k^{-1}v)\big) \\
&= (g \cdot f(g^{-1}u)) \otimes (k \cdot h(k^{-1}v)) \\
&= \big((g \cdot f) \boxtimes (k \cdot h)\big)(u \otimes v).
\end{aligned}$$

Since this equality holds for all simple tensors $u \otimes v$, which span $U \otimes V$, the linear maps are identical. Rewriting this using Ψ, we see that:

$$(g, k) \cdot \Psi(f \otimes h) = \Psi\big((g, k) \cdot (f \otimes h)\big).$$

Therefore, Ψ restricts to an isomorphism between the respective fixed-point subspaces:

$$\Psi : (\mathrm{Hom}(U, A) \otimes \mathrm{Hom}(V, B))^{G \times K} \xrightarrow{\sim} \mathrm{Hom}(U \otimes V, A \otimes B)^{G \times K}.$$

Finally, we identify the invariants of the tensor product. Let X be any G-representation and Y any K-representation, and consider $X \otimes Y$ as a $(G \times K)$-representation via $(g, k) \cdot (x \otimes y) = (g \cdot x) \otimes (k \cdot y)$. Then $(X \otimes Y)^{G \times K} = X^G \otimes Y^K$. Indeed, if $w \in (X \otimes Y)^{G \times K}$, then in particular $w \in (X \otimes Y)^{\{e\} \times K}$, so $w \in X \otimes Y^K$. Write $w = \sum_i x_i \otimes y_i$ with $y_i \in Y^K$. Now, invariance under $G \times \{e\}$ implies

$$\sum_i (g \cdot x_i) \otimes y_i = \sum_i x_i \otimes y_i \quad \text{for all } g \in G,$$

and since the y_i lie in Y^K, we may choose them to be linearly independent by regrouping terms. It follows that $g \cdot x_i = x_i$ for each i, hence $x_i \in X^G$ and therefore $w \in X^G \otimes Y^K$. The reverse inclusion $X^G \otimes Y^K \subseteq (X \otimes Y)^{G \times K}$ is immediate, proving the claim.

Applying this with $X = \mathrm{Hom}(U, A)$ and $Y = \mathrm{Hom}(V, B)$ gives

$$(\mathrm{Hom}(U, A) \otimes \mathrm{Hom}(V, B))^{G \times K} \;=\; \mathrm{Hom}(U, A)^G \otimes \mathrm{Hom}(V, B)^K \;=\; \mathrm{Hom}_G(U, A) \otimes \mathrm{Hom}_K(V, B).$$

Combining this with the fact that Ψ restricts to an isomorphism on the invariant subspaces, we conclude:

$$\mathrm{Hom}_G(U, A) \otimes \mathrm{Hom}_K(V, B) \;\cong\; \mathrm{Hom}_{G \times K}(U \otimes V, A \otimes B).$$

□

D.6 Proof of Theorem 1

Lemma 8 implies that, with a proper choice of basis, any feature vector ϕ_q in a model satisfying SLP can be expressed as a direct sum of components, where each component is a tensor product of an entity-dependent factor (from C_i) and a relation-dependent factor (from R_i). We now translate this general decomposition into specific constraints on the query feature map $\phi(h, r, t)$, to prove Theorem 1.

Theorem 1 (Context-Relation Factorization)**.** *Let $\phi : Q \to W$ be the feature map defined by $q \mapsto \phi_q$. If ϕ satisfies SLP, then there exist real vector spaces $\{C_i\}_i$, $\{R_i\}_i$ and an isomorphism $W \cong \bigoplus_i (C_i \otimes R_i)$ s.t.*

$$\phi(h,r,t) \;=\; \bigoplus_i \left(\sum_{k=1}^{m_i} u_{i,k}(h,t) \otimes v_{i,k}(r) \right),$$

where $u_{i,k} : E \times E \to C_i$ and $v_{i,k} : \mathcal{R} \to R_i$. Moreover, negation acts locally as a sign flip on each relation component, i.e., $v_{i,k}(\neg r) = -v_{i,k}(r)$.

Proof. Let V with the free vector space with basis $\{e_q : q \in Q\}$. The group action of H on the set Q naturally induces a linear representation on V, defined by permuting the basis vectors:

$$(\sigma, \epsilon) \cdot e_q \;:=\; e_{(\sigma,\epsilon)\cdot q}.$$

Let $\Phi : V \to W$ be the unique linear extension of ϕ, defined by $\Phi(e_q) := \phi_q$. We identify V with the tensor product

$$V \;\cong\; V_{\mathrm{ctx}} \otimes V_{\mathrm{rel}}$$

via the basis mapping $e_{(h,r,t)} \;\mapsto\; e_{(h,t)} \otimes e_r$. Under this identification, the H-action on V acts component-wise:

$$(\sigma, \epsilon) \cdot (e_{(h,t)} \otimes e_r) \;:=\; e_{(\sigma(h),\sigma(t))} \otimes e_{\epsilon \cdot r}.$$

By Lemma 7, $\phi : Q \to W$ is H-equivariant with respect to the H-action on queries. Since Φ is defined linearly on the basis Q, this equivariance extends to the entire space V, making Φ an H-equivariant map. Moreover, by Lemma 8, one can fix an H-equivariant isomorphism $W \cong \bigoplus_i (C_i \otimes R_i)$, and let $\pi_i : W \to C_i \otimes R_i$ be the corresponding projection. Define the component map $\Phi_i := \pi_i \circ \Phi$. Since both Φ and π_i are equivariant, Φ_i belongs to the space

$$\mathrm{Hom}_H(V_{\mathrm{ctx}} \otimes V_{\mathrm{rel}}, C_i \otimes R_i) \;\cong\; \mathrm{Hom}_{G_E}(V_{\mathrm{ctx}}, C_i) \otimes \mathrm{Hom}_{\mathbb{Z}_2}(V_{\mathrm{rel}}, R_i),$$

where the isomorphism is established by Lemma 9. Thus, Φ_i is an element of a tensor product space, implying it can be written as a finite sum of pure tensors:

$$\Phi_i \;=\; \sum_{k=1}^{m_i} U_{i,k} \otimes S_{i,k},$$

where $U_{i,k} : V_{\mathrm{ctx}} \to C_i$ is G_E-equivariant and $S_{i,k} : V_{\mathrm{rel}} \to R_i$ is $\mathbb{Z}_2$-equivariant. Defining the embeddings $u_{i,k}(h,t) := U_{i,k}(e_{(h,t)})$ and $v_{i,k}(r) := S_{i,k}(e_r)$, the overall feature map decomposes as:

$$\phi(h,r,t) \;=\; \bigoplus_i \left(\sum_{k=1}^{m_i} u_{i,k}(h,t) \otimes v_{i,k}(r) \right).$$

It remains to show that negation acts by sign on the relation embeddings. Let $\epsilon \in \mathbb{Z}_2$ be the negation element. The H-action on the feature space decomposes block-wise. On the i-th block $C_i \otimes R_i$, the action is component-wise:

$$(\mathrm{id}, \epsilon) \cdot (u \otimes v) \;=\; (\mathrm{id} \cdot u) \otimes (\epsilon \cdot v).$$

Since R_i is a real irreducible representation of $\mathbb{Z}_2$, it is one-dimensional, so ϵ acts on R_i as a scalar $\eta_i \in \{+1, -1\}$. Crucially, because the first component of the group element is the identity, the embedding $u_{i,k}$ remains unchanged. Thus, the action on the sum is:

$$\begin{aligned} \phi_i(h, \neg r, t) \;&=\; (\mathrm{id}, \epsilon) \cdot \sum_{k=1}^{m_i} \big(u_{i,k}(h,t) \otimes v_{i,k}(r)\big) \\ &=\; \eta_i \sum_{k=1}^{m_i} \big(u_{i,k}(h,t) \otimes v_{i,k}(r)\big). \end{aligned}$$

The SLP condition requires $\phi(h, \neg r, t) = -\phi(h, r, t)$ for all queries. Comparing this with the equation above, we see that for any block i where the feature map is not identically zero, we must have $\eta_i = -1$. (If $\phi_i \equiv 0$, the SLP constraint is satisfied on this block for any η_i, hence the sign choice is immaterial.) Consequently, in all cases, the relation embeddings must satisfy the sign-flip property:

$$v_{i,k}(\neg r) \;=\; -v_{i,k}(r).$$

Thus, every contributing relation component transforms under negation by the sign representation. □

E PROOF OF THEOREM 3

In this section, we provide the proof of Theorem 3. We first prove the lemma below, which will play a crucial role in proving Theorem 3.

Lemma 10 (Converse alignment in Hom-spaces). *Fix i and an H-equivariant projection $\pi_i : W \to W_i \cong C_i \otimes R_i$. Let $\Phi_i := \pi_i \circ \Phi \in \mathrm{Hom}_H(V_{\mathrm{ctx}} \otimes V_{\mathrm{rel}},\, C_i \otimes R_i)$, and identify this space with $\mathcal{A}_i \otimes \mathcal{B}_i$ via Lemma 9, where $\mathcal{A}_i = \mathrm{Hom}_{G_E}(V_{\mathrm{ctx}}, C_i)$ and $\mathcal{B}_i = \mathrm{Hom}_{\mathbb{Z}_2}(V_{\mathrm{rel}}, R_i)$. Let $P_{\mathrm{pair}} : V_{\mathrm{ctx}} \to V_{\mathrm{ctx}}$ and $P_{\mathrm{rel}} : V_{\mathrm{rel}} \to V_{\mathrm{rel}}$ be the linear maps defined by their action on the basis vectors:*

$$P_{\mathrm{pair}}(e_{(h,t)}) := e_{(t,h)} \qquad \text{and} \qquad P_{\mathrm{rel}}(e_r) := e_{r^{\smile}}.$$

Note that the converse operation commutes with negation (i.e., $(\neg r)^{\smile} = \neg(r^{\smile})$), so P_{rel} is $\mathbb{Z}_2$-equivariant. Define involutions $\mathcal{P}_{\mathrm{pair}}(U) = U \circ P_{\mathrm{pair}}$ on $\mathcal{A}_i$ and $\mathcal{P}_{\mathrm{rel}}(S) = S \circ P_{\mathrm{rel}}$ on $\mathcal{B}_i$. If $\phi(t, r^{\smile}, h) = \phi(h, r, t)$ for all $(h, r, t) \in Q$, then

$$(\mathcal{P}_{\mathrm{pair}} \otimes \mathcal{P}_{\mathrm{rel}})(\Phi_i) = \Phi_i.$$

Proof. First, we verify that $\mathcal{P}_{\mathrm{pair}}$ and $\mathcal{P}_{\mathrm{rel}}$ are well-defined. For any $U \in \mathcal{A}_i$ and $g \in G_E$, since P_{pair} commutes with the G_E-action,

$$(U \circ P_{\mathrm{pair}})(g \cdot x) = U(P_{\mathrm{pair}}(g \cdot x)) = U(g \cdot P_{\mathrm{pair}}(x)) = g \cdot (U \circ P_{\mathrm{pair}})(x),$$

so $\mathcal{P}_{\mathrm{pair}}(U) \in \mathcal{A}_i$. Similarly, $\mathbb{Z}_2$-equivariance of P_{rel} ensures $\mathcal{P}_{\mathrm{rel}}$ is well-defined on $\mathcal{B}_i$.

Next, since $\phi(t, r^{\smile}, h) = \phi(h, r, t)$ for all $(h, r, t) \in Q$, $\Phi \circ (P_{\mathrm{pair}} \otimes P_{\mathrm{rel}}) = \Phi$. Applying π_i gives

$$\Phi_i \circ (P_{\mathrm{pair}} \otimes P_{\mathrm{rel}}) = (\pi_i \circ \Phi) \circ (P_{\mathrm{pair}} \otimes P_{\mathrm{rel}}) = \pi_i \circ \Phi = \Phi_i.$$

Under the identification $\mathcal{A}_i \otimes \mathcal{B}_i \cong \mathrm{Hom}_H(V_{\mathrm{ctx}} \otimes V_{\mathrm{rel}}, C_i \otimes R_i)$, a pure tensor $U \otimes S$ corresponds to the map $x \otimes y \mapsto U(x) \otimes S(y)$. Precomposing with $P_{\mathrm{pair}} \otimes P_{\mathrm{rel}}$ yields

$$(x \otimes y) \mapsto U(P_{\mathrm{pair}}(x)) \otimes S(P_{\mathrm{rel}}(y)) = (U \circ P_{\mathrm{pair}})(x) \otimes (S \circ P_{\mathrm{rel}})(y),$$

which is exactly the map corresponding to $(\mathcal{P}_{\mathrm{pair}} U) \otimes (\mathcal{P}_{\mathrm{rel}} S)$. Extending linearly shows that precomposition by $P_{\mathrm{pair}} \otimes P_{\mathrm{rel}}$ corresponds to $\mathcal{P}_{\mathrm{pair}} \otimes \mathcal{P}_{\mathrm{rel}}$, hence $(\mathcal{P}_{\mathrm{pair}} \otimes \mathcal{P}_{\mathrm{rel}})(\Phi_i) = \Phi_i$. ☐

To prove Theorem 3, we restate its general form.

Theorem 3 (Symmetric-Antisymmetric Alignment (Formal statement)). *Assume the context-relation factorization from Theorem 1 and $\phi(t, r^{\smile}, h) = \phi(h, r, t)$ for all $(h, r, t) \in Q$. Then, for each block i (as in Theorem 1), the projected feature component $\phi_i := \pi_i \circ \phi : Q \to W_i \cong C_i \otimes R_i$ admits a decomposition*

$$\phi_i(h, r, t) = \phi_i^{+}(h, r, t) + \phi_i^{-}(h, r, t),$$

where each part can be written as a finite sum of context–relation pure tensors

$$\phi_i^{\pm}(h, r, t) = \sum_{k \in I_i^{\pm}} u_{i,k}^{\pm}(h, t) \otimes v_{i,k}^{\pm}(r),$$

such that every summand satisfies the aligned symmetry relations

$$u_{i,k}^{\pm}(t, h) = \pm\, u_{i,k}^{\pm}(h, t), \qquad v_{i,k}^{\pm}(r^{\smile}) = \pm\, v_{i,k}^{\pm}(r).$$

Consequently,

$$\phi(h, r, t) = \bigoplus_i \big(\phi_i^{+}(h, r, t) + \phi_i^{-}(h, r, t)\big)$$

with the same componentwise symmetry properties on each block.

Proof. Fix a block index i and an H-equivariant projection $\pi_i : W \to W_i \cong C_i \otimes R_i$, and write $\Phi_i := \pi_i \circ \Phi \in \mathrm{Hom}_H(V_{\mathrm{ctx}} \otimes V_{\mathrm{rel}},\, C_i \otimes R_i)$. By Lemma 9, we identify

$$\mathrm{Hom}_H(V_{\mathrm{ctx}} \otimes V_{\mathrm{rel}},\, C_i \otimes R_i) \;\cong\; \mathcal{A}_i \otimes \mathcal{B}_i,$$

where $\mathcal{A}_i := \mathrm{Hom}_{G_E}(V_{\mathrm{ctx}}, C_i)$ and $\mathcal{B}_i := \mathrm{Hom}_{\mathbb{Z}_2}(V_{\mathrm{rel}}, R_i)$.

Let $\mathcal{P}_{\mathrm{pair}}$ and $\mathcal{P}_{\mathrm{rel}}$ be the involutions on $\mathcal{A}_i$ and $\mathcal{B}_i$ defined in Lemma 10. Since they are involutions, their eigenvalues are either $+1$ or -1. We write their eigenspace decompositions as

$$\mathcal{A}_i = \mathcal{A}_i^+ \oplus \mathcal{A}_i^- \qquad \text{and} \qquad \mathcal{B}_i = \mathcal{B}_i^+ \oplus \mathcal{B}_i^-,$$

where the superscripts $\pm$ denote the eigenspaces corresponding to eigenvalues ± 1, respectively.

Recall from Lemma 10 that $(\mathcal{P}_{\mathrm{pair}} \otimes \mathcal{P}_{\mathrm{rel}})(\Phi_i) = \Phi_i$. Note that the operator $\mathcal{P}_{\mathrm{pair}} \otimes \mathcal{P}_{\mathrm{rel}}$ acts on a pure tensor $U \otimes S \in \mathcal{A}_i^\sigma \otimes \mathcal{B}_i^\tau$ (where $\sigma, \tau \in \{+, -\}$) by scalar multiplication with $\sigma \cdot \tau$. Since Φ_i is invariant (eigenvalue $+1$), it must lie in the subspace where this product is positive:

$$\Phi_i \in (\mathcal{A}_i^+ \otimes \mathcal{B}_i^+) \oplus (\mathcal{A}_i^- \otimes \mathcal{B}_i^-).$$

Consequently, Φ_i uniquely decomposes as $\Phi_i = \Phi_i^+ + \Phi_i^-$ with terms in these respective subspaces.

We now expand each $\Phi_i^\pm$ as a finite sum of pure tensors with factors in the corresponding eigenspaces. Since V_{ctx} and V_{rel} are finite-dimensional (in particular, $|E|, |R| < \infty$), the spaces $\mathcal{A}_i^\pm$ and $\mathcal{B}_i^\pm$ are finite-dimensional. Let $d_+ := \dim \mathcal{A}_i^+$ and choose a basis $\{U_{i,1}^+, \dots, U_{i,d_+}^+\}$ of $\mathcal{A}_i^+$. Consider the linear map

$$T^+ : (\mathcal{B}_i^+)^{d_+} \to \mathcal{A}_i^+ \otimes \mathcal{B}_i^+, \qquad T^+(S_1, \dots, S_{d_+}) := \sum_{k=1}^{d_+} U_{i,k}^+ \otimes S_k.$$

This map is surjective. Indeed, any element in the tensor product is a sum of pure tensors $U \otimes S$. Since $\{U_{i,k}^+\}$ is a basis, any $U \in \mathcal{A}_i^+$ can be written as $\sum_k c_k U_{i,k}^+$. Substituting this expansion into the pure tensors and regrouping terms by $U_{i,k}^+$ shows that any element takes the form $\sum_k U_{i,k}^+ \otimes S_k'$ for some $S_k' \in \mathcal{B}_i^+$.

Since the dimensions of the domain and codomain coincide:

$$\dim(\mathcal{B}_i^+)^{d_+} = d_+ \cdot \dim \mathcal{B}_i^+ = (\dim \mathcal{A}_i^+) \cdot (\dim \mathcal{B}_i^+) = \dim(\mathcal{A}_i^+ \otimes \mathcal{B}_i^+),$$

the surjective linear map T^+ is an isomorphism. Therefore, there exist unique maps $S_{i,1}^+, \dots, S_{i,d_+}^+ \in \mathcal{B}_i^+$ such that

$$\Phi_i^+ = \sum_{k=1}^{d_+} U_{i,k}^+ \otimes S_{i,k}^+.$$

Set $I_i^+ := \{1, \dots, d_+\}$. Analogously, we obtain unique maps $S_{i,1}^-, \dots, S_{i,d_-}^- \in \mathcal{B}_i^-$ such that

$$\Phi_i^- = \sum_{k=1}^{d_-} U_{i,k}^- \otimes S_{i,k}^-.$$

Set $I_i^- := \{1, \dots, d_-\}$.

Now, define the projected feature maps by evaluation on basis tensors:

$$\phi_i^\pm(h, r, t) := \Phi_i^\pm\big(e_{(h,t)} \otimes e_r\big), \qquad u_{i,k}^\pm(h, t) := U_{i,k}^\pm(e_{(h,t)}), \qquad v_{i,k}^\pm(r) := S_{i,k}^\pm(e_r).$$

Under the identification of Lemma 9, each pure tensor $U_{i,k}^\pm \otimes S_{i,k}^\pm$ acts by $(x \otimes y) \mapsto U_{i,k}^\pm(x) \otimes S_{i,k}^\pm(y)$, hence

$$\phi_i^\pm(h, r, t) = \sum_{k \in I_i^\pm} u_{i,k}^\pm(h, t) \otimes v_{i,k}^\pm(r),$$

and $\phi_i = \phi_i^+ + \phi_i^-$ since $\Phi_i = \Phi_i^+ + \Phi_i^-$. Finally, since $U_{i,k}^\pm \in \mathcal{A}_i^\pm$ means $U_{i,k}^\pm \circ P_{\mathrm{pair}} = \pm U_{i,k}^\pm$, evaluating at $e_{(h,t)}$ gives $u_{i,k}^\pm(t, h) = \pm u_{i,k}^\pm(h, t)$. Likewise $S_{i,k}^\pm \in \mathcal{B}_i^\pm$ means $S_{i,k}^\pm \circ P_{\mathrm{rel}} = \pm S_{i,k}^\pm$, and evaluating at e_r gives $v_{i,k}^\pm(r^\smile) = \pm v_{i,k}^\pm(r)$. Summing over i yields the global decomposition. $\square$

F PROOF OF LEMMA 2

In this section, we provide the proof of Lemma 2.

Lemma 2 (Kernel Stability Yields Bilinearity). *Assume Assumption 1 and the symmetry of conjunction (Def. 4(ii)). Then, there exists a unique symmetric bilinear operator $\tilde{F} : W^\wedge \times W^\wedge \to W^\wedge$ such that $\tilde{F}(\phi_p, \phi_q) = \phi_{p\wedge q}$ for all $p, q \in Q^\wedge$.*

Proof. Fix $q \in Q^\wedge$. Let $V^\wedge$ be the free real vector space with basis $\{e_p : p \in Q^\wedge\}$, and let $\Phi : V^\wedge \to W^\wedge$ be the linear map defined on basis vectors by $\Phi(e_p) = \phi_p$. Define a linear map $P_{\wedge,q} : V^\wedge \to V^\wedge$ by its action on basis:

$$P_{\wedge,q}(e_p) := e_{p\wedge q} \qquad (p \in Q^\wedge).$$

First, we claim that $P_{\wedge,q}$ induces a well-defined linear operator $L_\wedge(q) \in \mathrm{End}(W^\wedge)$ satisfying

$$L_\wedge(q)\,\phi_p = \phi_{p\wedge q} \qquad \forall\, p \in Q^\wedge.$$

For any $w \in W^\wedge$, since $W^\wedge = \mathrm{span}\{\phi_p : p \in Q^\wedge\}$, we may choose $x \in V^\wedge$ with $\Phi(x) = w$, and define

$$L_\wedge(q)\,w := \Phi\big(P_{\wedge,q}x\big).$$

It remains to check that this does not depend on the choice of x. If x' is another preimage of w, then $x - x' \in \ker\Phi$. By Assumption 1, we have $P_{\wedge,q}(\ker\Phi) \subseteq \ker\Phi$, hence $\Phi(P_{\wedge,q}(x - x')) = 0$, i.e., $\Phi(P_{\wedge,q}x) = \Phi(P_{\wedge,q}x')$. Thus $L_\wedge(q)$ is well-defined. Linearity follows from linearity of Φ and $P_{\wedge,q}$. Applying the definition to $w = \phi_p = \Phi(e_p)$ gives

$$L_\wedge(q)\,\phi_p = \Phi\big(P_{\wedge,q}e_p\big) = \Phi(e_{p\wedge q}) = \phi_{p\wedge q},$$

as claimed.

We now construct the map $\tilde{F} : W^\wedge \times W^\wedge \to W^\wedge$. Since $\{\phi_p : p \in Q^\wedge\}$ spans $W^\wedge$, any vector $v \in W^\wedge$ can be decomposed as a finite linear combination $v = \sum_i c_i\phi_{q_i}$. For any $u \in W^\wedge$, we define $\tilde{F}$ linear in the second argument by using the operators $L_\wedge(q_i)$:

$$\tilde{F}(u, v) \;:=\; \sum_i c_i L_\wedge(q_i)\,u.$$

To ensure the well-definedness of $\tilde{F}(u, v)$, we must check that this definition is independent of the decomposition of v. Suppose $\sum_i c_i\phi_{q_i} = 0$, i.e., the vector $y = \sum_i c_i e_{q_i}$ lies in $\ker\Phi$. We check the value of the map for any basis vector $u = \phi_p$ ($p \in Q^\wedge$):

$$\sum_i c_i L_\wedge(q_i)\,\phi_p \;=\; \sum_i c_i\phi_{p\wedge q_i}.$$

By commutativity of conjunction on realized inputs (equivalently, by the assumed symmetry of F), we have $\phi_{p\wedge q_i} = \phi_{q_i\wedge p}$ for all i. Thus,

$$\sum_i c_i\phi_{p\wedge q_i} \;=\; \sum_i c_i\phi_{q_i\wedge p} \;=\; \Phi\left(\sum_i c_i e_{q_i\wedge p}\right) \;=\; \Phi\big(P_{\wedge,p}(y)\big).$$

Since $y \in \ker\Phi$, by Assumption 1, we have $P_{\wedge,p}(y) \in \ker\Phi$, so $\Phi(P_{\wedge,p}(y)) = 0$. Thus, $\sum_i c_i L_\wedge(q_i)\phi_p = 0$ for all $p \in Q^\wedge$. Since $\{\phi_p : p \in Q^\wedge\}$ spans $W^\wedge$, this implies $\sum_i c_i L_\wedge(q_i)u = 0$ for all $u \in W^\wedge$. Therefore, if $v = \sum_i c_i\phi_{q_i} = \sum_j d_j\phi_{r_j}$ are two decompositions, then $\sum_i c_i\phi_{q_i} - \sum_j d_j\phi_{r_j} = 0$ implies

$$\sum_i c_i L_\wedge(q_i)u \;=\; \sum_j d_j L_\wedge(r_j)u \qquad \forall\, u \in W^\wedge,$$

so $\tilde{F}(u, v)$ is independent of the chosen decomposition of v.

By construction, $\tilde{F}$ is linear in the second argument. Also, since each $L_\wedge(q_i)$ is linear, $\tilde{F}$ is linear in the first argument. Finally, for realized inputs, $\tilde{F}(\phi_p, \phi_q) = L_\wedge(q)\phi_p = \phi_{p\wedge q}$. Thus, $\tilde{F}$ is the unique bilinear extension. ☐