Preferring Unified Theories

Robert J. Carroll*

March 2026

Abstract

What makes unification of scientific theories valuable? Merely conjoining two theories (the categorical coproduct) is conservative—it produces no new consequences for either domain. We argue that genuine unification is a *generative effect*: the unified theory proves things that no mere conjunction of its components could prove. Working within the categorical framework for scientific theories (Halvorson, Barrett) and the theory of generative effects (Fong–Spivak, Adam), we show that an observation can detect the value of unification if and only if it is not a left adjoint, and that preferences representable by single-domain (left-adjoint) observation families are structurally blind to interaction value. Two agents satisfying identical rationality axioms can therefore rationally disagree about whether unification is worth pursuing: the dividing line is whether they ask single-domain or cross-domain questions. We illustrate the framework with examples from propositional logic, electromagnetism, and political economy.

*Department of Political Science, University of Illinois at Urbana-Champaign, rjc@illinois.edu

Unification is supposed to be valuable. When Maxwell showed that electricity and magnetism are aspects of a single electromagnetic theory, or when the electroweak theory revealed a common origin for electromagnetism and the weak nuclear force, the result was not merely a tidier bookshelf. New phenomena emerged—electromagnetic waves, the W and Z bosons—that neither component theory could have predicted alone. The unified theory *bought something*.

But what, exactly? And how do we distinguish the cases where unification buys something from the cases where it doesn't? These questions are harder than they appear. Consider two candidate answers to "What does it mean to unify theories T_1 and T_2?":

(i) *Conjoin them.* Place T_1 and T_2 side by side in a single framework: here is our theory of electricity, and here is our theory of magnetism.

(ii) *Expand the vocabulary.* Embed both theories in a richer language—add sorts, function symbols, relation symbols—so that both can be expressed in a common formalism.

Neither answer captures what we value about unification. Option (i) is trivial: it produces no new consequences about either domain. Option (ii) is potentially misleading: expanding the vocabulary without adding new content is a conservative extension—it complicates the formalism without buying anything. Calling this "unification" would conflate notational expansion with substantive progress.

The philosophical literature has offered two influential accounts of what genuine unification buys. Kitcher [1981] argued that the value lies in explanatory connections: a unified theory derives diverse phenomena from a small stock of argument patterns. Myrvold [2003] gave a Bayesian account: unified theories receive higher posterior probability because they make the evidential support from one domain bear on predictions in another. Morrison [2000] argued, complementarily, that unification and explanation can come apart: a theory can unify without explaining, and vice versa. Our framework is consistent with Morrison's insight: generative effects are about new *consequences*, not explanatory structure, and a unified theory can produce novel cross-theoretic consequences without those consequences constituting explanations of either domain's phenomena. All three accounts identify something real. But none provides a *structural* characterization—a necessary and sufficient condition, stated in terms of the logical architecture of theories, for

when unification has bought something and for which agents are in a position to see the purchase.

This paper provides such a characterization. Each of its ingredients is established: the categorical framework for scientific theories developed by Halvorson [2019] and Barrett and Halvorson [2016]; the theory of generative effects from Fong and Spivak [2019] and Adam [2017]; and the decision-theoretic tradition of Savage [1954]. Our contribution is not to extend any of these programs individually, but to connect them: the category of theories has the preorder-with-joins structure that the generative effects framework requires, and the resulting characterization theorem has consequences for rational theory choice that none of these literatures has drawn out. We argue for three claims:

Unification is a generative effect. A unified theory is one whose combination of components produces consequences that no component—and no mere conjunction of components—could produce alone. The categorical framework makes this precise: mere conjunction is the coproduct (or pushout), which is conservative; genuine unification is a non-conservative extension of the coproduct, exhibiting what Fong and Spivak call a *generative effect*.

The adjunction theorem characterizes who can see it. Not all ways of evaluating a theory are sensitive to unification. The criterion is sharp: an observation (a monotone map from theories to some evaluation space) can detect unification if and only if it is *not* a left adjoint. Single-language consequence operators—"What does this theory tell me about *my* domain?"—are left adjoints, and are therefore blind to unification. Cross-domain observations are not, and can see the surplus.

Rational disagreement about unification is structural. Two agents who satisfy identical rationality axioms can disagree about whether unification is worth pursuing. The dividing line is not psychology but observation structure: an agent who asks only single-domain questions will rationally see no value in unification, while an agent who asks cross-domain questions will see strict value. The framework makes this disagreement precise, predictable, and non-pathological.

For readers who want the argumentative skeleton up front: the paper has three steps. First, it identifies the baseline for "mere combination" of theories

(coproduct/pushout) and proves conservativity at that baseline. Second, it defines unification as a non-conservative extension of that baseline and uses the adjunction theorem to characterize which observations can detect the resulting surplus. Third, it shows that this distinction induces a normative one: preferences built from adjoint observations are blind to interaction value, while preferences built from at least one non-adjoint observation can strictly favor unification.

The framework also subsumes *intertheoretic reduction* as a special case. Nagelian bridge laws [Nagel, 1961], refined by Schaffner [1967] and reconstructed by Dizadji-Bahmani et al. [2010], are formally the same operation as adding bridge axioms to a coproduct or pushout—a non-conservative extension producing generative effects. But reduction is asymmetric (one theory reduces *to* another), whereas the unification formalized here is symmetric: Maxwell's electromagnetism does not reduce electricity to magnetism; it unifies them. We return to this relationship in Section 6.

Sections 1–2 set up the categorical framework and establish the conservativity baseline. Section 3 identifies unification with generative effects and states the adjunction theorem. Section 4 derives the normative results. Section 5 works three examples at increasing levels of complexity, and Section 6 discusses limitations.

1 The Category of Theories

We work within the framework of multi-sorted first-order logic, following Halvorson [2019] and Barrett and Halvorson [2016]. This section establishes the category **Th** whose objects are theories and whose morphisms are translations, together with the key notions of conservative extension and theoretical equivalence that the rest of the paper relies on.

Definition 1.1 (Theory)**.** A *theory* is a pair $T = (\Sigma, \Gamma)$, where

- Σ is a multi-sorted first-order *signature*: a specification of sort symbols $s_1, s_2, \ldots$, function symbols with designated input and output sorts, and relation symbols with designated argument sorts; and
- Γ is a consistent, deductively closed set of Σ-sentences.

We write $L(\Sigma)$ or $L(T)$ for the set of all sentences in the language of Σ.

Multi-sorted signatures are the natural setting for scientific theories, which routinely quantify over several kinds of entity at once—spacetime points and field values in physics, agents and goods in economics, voters and policies in political science. The single-sorted case is recovered when Σ has exactly one sort.

Standing assumption. Throughout this paper, every theory is assumed to be consistent. This rules out the degenerate case of the inconsistent theory, which proves every sentence and would trivially dominate any comparison (see Section 3).

The morphisms of **Th** are maps that carry the content of one theory into the language of another.

Definition 1.2 (Translation)**.** A *translation* $F\colon T_1 \to T_2$ assigns

- to each sort s of Σ_1, a formula-in-context $\delta_s(x)$ of $L(\Sigma_2)$ (specifying a "translated sort"—a definable subset of a sort of T_2);
- to each n-ary relation symbol R of Σ_1, a formula $\varphi_R(x_1, \ldots, x_n)$ of $L(\Sigma_2)$;
- to each function symbol, a corresponding provably functional relation in T_2;

in such a way that provability is preserved: for every sentence σ,

$$T_1 \vdash \sigma \quad \Longrightarrow \quad T_2 \vdash F(\sigma).$$

Translations compose: if $F\colon T_1 \to T_2$ and $G\colon T_2 \to T_3$ are both provability-preserving, then so is $G \circ F$. The identity translation on any theory is trivially provability-preserving. Theories and translations therefore form a category, which we denote **Th**.

Remark 1.3. Halvorson and Tsementzis [2017] show that translations correspond to coherent functors between *syntactic categories*. The syntactic category $\mathcal{C}_T$ of a theory T has formulas-in-context as objects and provably functional relations as arrows; it carries a pretopos structure (finite limits and finite disjoint coproducts). Under this correspondence, **Th** is equivalent to the category **Pretop** of pretoposes and coherent functors. We will not need the pretopos machinery directly, but it guarantees the existence of the categorical constructions (Section 2) that we rely on.

The following notion captures when an extension of a theory is "merely notational"—adding vocabulary without adding content.

Definition 1.4 (Conservative translation)**.** A translation $F\colon T_1 \to T_2$ is *conservative* if T_2 proves nothing new about T_1's vocabulary: for every sentence $\varphi \in L(\Sigma_1)$,

$$T_2 \vdash F(\varphi) \quad \Longrightarrow \quad T_1 \vdash \varphi.$$

A conservative translation extends T_1 by possibly introducing new sorts, new relation symbols, and new axioms, but every consequence that can be stated in T_1's original language was already provable from T_1's own axioms. There is no explanatory gain, and no new empirical content, for the original domain.

Proposition 1.5. *Conservative translations are the monomorphisms of* **Th**.

See Halvorson [2019, Ch. 4] for the proof. This is the key structural fact for what follows. When we say that a unified theory "goes beyond mere conjunction," we will mean precisely that the translation from the conjunction into the unified theory is *not* conservative (Section 3).

Finally, we need a notion of when two theories should count as "the same." Several equivalence relations on theories have been studied; they form a strict hierarchy.

Definition 1.6 (Definitional equivalence)**.** Theories T_1 and T_2 are *definitionally equivalent* (DE) if there exist translations $F\colon T_1 \to T_2$ and $G\colon T_2 \to T_1$ such that the composites $G \circ F$ and $F \circ G$ are provably equivalent to the respective identity translations:

$$T_1 \underset{G}{\overset{F}{\rightleftarrows}} T_2 \qquad G \circ F \sim \mathrm{id}_{T_1}, \quad F \circ G \sim \mathrm{id}_{T_2}.$$

Definition 1.7 (Morita equivalence)**.** Theories T_1 and T_2 are *Morita equivalent* (ME) if they are definitionally equivalent up to the introduction of new sort symbols—product sorts, coproduct sorts, subsorts, and quotient sorts. Formally, T_1 and T_2 are ME if there exist *Morita extensions* $T_1' \supseteq T_1$ and $T_2' \supseteq T_2$ such that T_1' and T_2' are definitionally equivalent [see Barrett and Halvorson, 2016, for the full treatment].

For single-sorted theories, ME coincides with DE; for multi-sorted theories, ME is strictly weaker—there exist theories that are Morita equivalent but not definitionally equivalent [Halvorson, 2019, Ch. 5].

A third, purely semantic notion asks only that the *categories of models* be equivalent: $\mathrm{Mod}(T_1) \simeq \mathrm{Mod}(T_2)$. This *categorical equivalence* (CE) is strictly weaker than ME [Halvorson, 2019, Ch. 6]. The full hierarchy is:

$$\text{DE} \implies \text{ME} \implies \text{CE},$$

with both implications strict in the multi-sorted setting. The choice among these notions is an active area of debate; see Weatherall [2019] and Barrett [2020] for recent discussions of the criteria for theoretical equivalence and their physical significance.

Remark 1.8. The model functor $\mathrm{Mod}\colon \mathbf{Th} \to \mathbf{Cat}$ is neither full, nor faithful, nor essentially surjective [Halvorson and Tsementzis, 2017]. Working with models alone therefore loses information about translations between theories. This is a central motivation for the syntactic, category-theoretic approach taken here.

Throughout this paper, **Morita equivalence is our notion of theoretical identity**: ME-equivalent theories will be treated as interchangeable.

2 Coproducts, Pushouts, and Mere Conjunction

Having set up the category **Th**, we now describe the categorical constructions that formalize "putting theories together." The coproduct corresponds to placing two theories side by side with no interaction; the pushout generalizes this to theories that share a common sub-theory. Both constructions are *conservative*: they add no new consequences for either component's vocabulary. This is the baseline against which genuine unification will be measured in Section 3.

2.1 The coproduct as mere conjunction

Definition 2.1 (Coproduct of theories)**.** Let $T_1 = (\Sigma_1, \Gamma_1)$ and $T_2 = (\Sigma_2, \Gamma_2)$ be theories with disjoint signatures ($\Sigma_1 \cap \Sigma_2 = \varnothing$). Their *coproduct* $T_1 \sqcup T_2$

is the theory

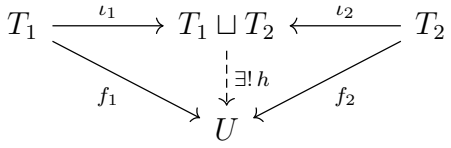
$$T_1 \sqcup T_2 \;=\; (\Sigma_1 \cup \Sigma_2,\;\; \overline{\Gamma_1 \cup \Gamma_2}),$$

where the bar denotes deductive closure. The *coprojections*

$$\iota_1 \colon T_1 \to T_1 \sqcup T_2 \qquad \text{and} \qquad \iota_2 \colon T_2 \to T_1 \sqcup T_2$$

are the inclusion translations: each sort, function symbol, and relation symbol is mapped to itself.

The coproduct satisfies a universal property: it is the "least committal" way to combine two theories. Given any theory U equipped with translations $f_1 \colon T_1 \to U$ and $f_2 \colon T_2 \to U$, there exists a unique translation $h \colon T_1 \sqcup T_2 \to U$ such that $h \circ \iota_1 = f_1$ and $h \circ \iota_2 = f_2$:



Any theory that receives translations from both T_1 and T_2 extends their coproduct. This means the coproduct is the weakest such theory—it asserts nothing beyond what T_1 and T_2 assert individually.

Proposition 2.2 (Conservativity of coprojections)**.** *The coprojections ι_1 and ι_2 are conservative translations. That is, for every sentence $\varphi \in L(\Sigma_1)$:*

$$T_1 \sqcup T_2 \vdash \varphi \quad \Longrightarrow \quad T_1 \vdash \varphi.$$

Proof sketch. Suppose $T_1 \sqcup T_2 \vdash \varphi$ for some $\varphi \in L(\Sigma_1)$. Since Σ_1 and Σ_2 are disjoint, any derivation of φ from $\Gamma_1 \cup \Gamma_2$ can use axioms from Γ_2 only vacuously—they share no vocabulary with φ or with Γ_1. By Robinson's joint consistency theorem (or equivalently, by Craig interpolation applied to the disjoint-signature case), the Γ_2-axioms can be eliminated from the proof, yielding $T_1 \vdash \varphi$. □

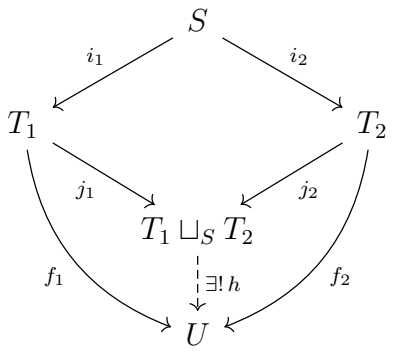
This is the formal content of *mere conjunction*. When we simply place two theories side by side—"here is our theory of X, and here is our theory of Y"—the result can say nothing about X that the original theory of X did not already say. The two theories coexist without interacting.

Remark 2.3. The coproduct construction extends to any finite family of theories $T_1, \ldots, T_n$ with pairwise disjoint signatures. Nothing in our analysis is limited to pairs.

2.2 Pushouts and shared content

In practice, theories being combined usually share vocabulary. Rational-choice economics and rational-choice political science both embed the framework of rational choice theory. Two spacetime theories may both presuppose the apparatus of differential geometry. The coproduct, which demands disjoint signatures, cannot represent this situation directly. The right construction is the *pushout*.

Definition 2.4 (Pushout of theories)**.** Let S be a theory with conservative translations $i_1 \colon S \to T_1$ and $i_2 \colon S \to T_2$ (so that S is a common sub-theory of both T_1 and T_2). The *pushout* $T_1 \sqcup_S T_2$ is the colimit of the diagram



satisfying the universal property: for any theory U with translations $f_1 \colon T_1 \to U$ and $f_2 \colon T_2 \to U$ such that $f_1 \circ i_1 = f_2 \circ i_2$ (i.e., the shared content is treated consistently), there exists a unique $h \colon T_1 \sqcup_S T_2 \to U$ with $h \circ j_1 = f_1$ and $h \circ j_2 = f_2$.

Concretely, the pushout $T_1 \sqcup_S T_2$ can be described as follows. Begin with the signatures Σ_1 and Σ_2; identify the sorts and symbols that correspond under i_1 and i_2 (i.e., glue along the shared sub-signature coming from S). The axioms include both Γ_1 and Γ_2, together with the identifications. The non-shared parts of Σ_1 and Σ_2 remain disjoint.

The coproduct is the special case where S is the empty theory (the initial object of **Th**, with an empty signature and no non-logical axioms):

$$T_1 \sqcup T_2 \;=\; T_1 \sqcup_{\varnothing} T_2.$$

Remark 2.5. Pushouts exist in the category of pretoposes [Halvorson and Tsementzis, 2017], and hence in **Th** under the equivalence **Th** $\simeq$ **Pretop**

(Remark 1.3). For the cases we consider—where the non-shared vocabularies of T_1 and T_2 are disjoint—the concrete description above is straightforward.[1]

2.3 The conservativity baseline

The coproduct and pushout share a crucial feature: they are conservative over non-shared vocabulary.

Proposition 2.6 (Conservativity of pushout maps)**.** *Let $T_1 \sqcup_S T_2$ be the pushout along a common sub-theory S. The canonical map $j_1 \colon T_1 \to T_1 \sqcup_S T_2$ is conservative over the non-shared vocabulary: for any sentence $\varphi \in L(\Sigma_1 \setminus \Sigma_S)$,*

$$T_1 \sqcup_S T_2 \vdash \varphi \quad \Longrightarrow \quad T_1 \vdash \varphi,$$

and symmetrically for j_2 and T_2.

The argument is analogous to Proposition 2.2: the non-shared vocabularies are disjoint, so axioms from one side cannot contribute to proofs about the other side's proprietary language.

An important caveat: the pushout *can* produce new consequences in the shared vocabulary $L(\Sigma_S)$. This is not a failure but a feature—it reflects the interaction of T_1 and T_2 through their common core S. Consider: if T_1 implies that every S-entity has property P, and T_2 implies that every S-entity with property P also has property Q, then $T_1 \sqcup_S T_2$ implies that every S-entity has property Q—a consequence neither theory proved alone, but one that follows from their interaction through S.

The coproduct, where S is empty, is the extreme case: there is no shared vocabulary at all, so there is no interaction and no new consequences whatsoever.

To summarize: coproducts and pushouts formalize the *minimal* combination of theories—mere conjunction, possibly with shared content identified. They provide a baseline, a "zero point" of unification. The question that drives the rest of this paper is: what happens when we go *beyond* this baseline?

[1] The general existence of pushouts for theories with overlapping non-shared vocabulary requires additional care; see Halvorson [2019, Ch. 5] for the relevant Morita extension techniques.

3 Unification as Generative Effect

The coproduct $T_1 \sqcup T_2$ and the pushout $T_1 \sqcup_S T_2$ represent *mere conjunction*—the minimal combination of theories, with no surplus content. A genuinely unified theory goes further. This section makes this idea precise by importing the framework of *generative effects* from Fong and Spivak [2019] and Adam [2017], adapted to the setting of the category of theories.

The central identification is:

> *Genuine unification is a generative effect. A unified theory is one for which the whole is strictly more than the sum of its parts.*

3.1 The theory preorder and observations

We begin by equipping the category **Th** with a preorder structure and defining the maps that "observe" theories from the outside.

Definition 3.1 (Theory preorder)**.** Define a preorder $\leq$ on **Th** by

$$T_1 \leq T_2 \quad \overset{\text{def}}{\iff} \quad \text{there exists a translation } F \colon T_1 \to T_2.$$

This is a preorder (reflexive and transitive) but not a partial order: non-isomorphic theories can be mutually translatable. Quotienting by Morita equivalence yields a partial order. The coproduct $T_1 \sqcup T_2$ is the join of T_1 and T_2 in this preorder.[2]

Definition 3.2 (Observation)**.** An *observation* is a monotone map

$$\Phi \colon (\mathbf{Th}, \leq) \longrightarrow (Q, \leq_Q),$$

where $(Q, \leq_Q)$ is a preorder with joins (denoted $\vee_Q$).

Following Adam [2017], we think of Φ as a *quasi-veil*: it partially conceals the theory, revealing only what is visible through a particular observational lens. The codomain Q is a space of *phenomes*—what remains visible after the veil is applied. Two theories that are indistinguishable under Φ have the same phenome, even if they differ in their concealed content.

[2] More precisely, $T_1 \sqcup T_2$ is the join in the preorder restricted to theories receiving translations from both T_1 and T_2—the categorical coproduct.

Standing assumption. Throughout this paper, every observation depends only on the *deductive content* of theories: if T and T' have the same consequences in every language, then $\Phi(T) = \Phi(T')$ for every observation Φ. This rules out observations sensitive to, e.g., signature cardinality or syntactic presentation. All observations considered in this paper—consequence operators, cross-domain observations, and their aggregates—satisfy this condition.[3]

Example 3.3 (Consequence operator)**.** The canonical observation is the *consequence operator*. Fix a language L, and define

$$\mathrm{Cn}_L \colon \mathbf{Th} \longrightarrow (\mathcal{P}(L), \subseteq) \qquad \text{by} \qquad \mathrm{Cn}_L(T) \;=\; \{\, \varphi \in L \mid T \vdash \varphi \,\}.$$

This is monotone: if $T_1 \leq T_2$ (i.e., there is a translation $T_1 \to T_2$), then the translated L-consequences of T_1 are among the L-consequences of T_2. The observation Cn_L asks: "What does T tell us about L-sentences?"

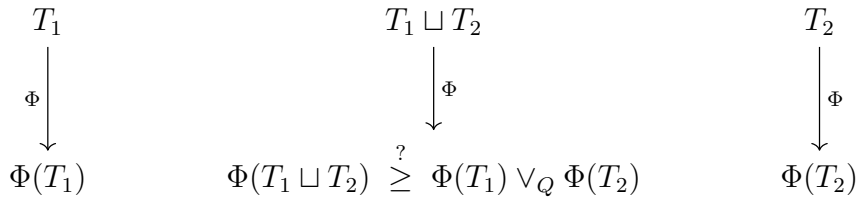


Figure 1: An observation Φ applied to two theories and their coproduct. A generative effect occurs when $\Phi(T_1 \sqcup T_2)$ strictly exceeds $\Phi(T_1) \vee_Q \Phi(T_2)$—the combination reveals more than the parts.

3.2 Generative effects

We now state the key definition, adapted from Fong and Spivak [2019, Def. 1.93] and Adam [2017, eq. 1.1].

[3]Formally, this amounts to requiring that Φ factors through the deductive closure operator on the relevant sub-language. This is a mild restriction: it excludes only observations that respond to syntax rather than content, which are not epistemically interesting for theory evaluation.

Definition 3.4 (Generative effect)**.** Let $\Phi\colon (\mathbf{Th}, \leq) \to (Q, \leq_Q)$ be an observation. We say Φ has a *generative effect* at the pair (T_1, T_2) if

$$\Phi(T_1 \sqcup T_2) \;>\; \Phi(T_1) \vee_Q \Phi(T_2).$$

In Adam's language: interactional effects emerge when the *phenome* of the interconnected system cannot be explained by interconnecting the phenomes of the separate systems. When $\Phi(T_1 \sqcup T_2) = \Phi(T_1) \vee_Q \Phi(T_2)$, the observation "decomposes"—what Φ sees in the whole is exactly what it sees in the parts, and there is nothing new to account for. When strict inequality holds, the combination has produced something that neither part, nor their simple union, predicted.

The conservativity of the coproduct (Proposition 2.2) can now be restated in this language:

Proposition 3.5 (Mere conjunction has no generative effects)**.** *For any language $L \subseteq L(\Sigma_1)$, the consequence operator Cn_L has no generative effect at (T_1, T_2) relative to their coproduct:*

$$\mathrm{Cn}_L(T_1 \sqcup T_2) \;=\; \mathrm{Cn}_L(T_1) \;\cup\; \mathrm{Cn}_L(T_2) \;=\; \mathrm{Cn}_L(T_1).$$

The second equality holds because $L \subseteq L(\Sigma_1)$ and the signatures are disjoint, so $\mathrm{Cn}_L(T_2) \subseteq \mathrm{Cn}_L(T_1 \sqcup T_2)$ contributes only logical tautologies already in $\mathrm{Cn}_L(T_1)$.

Proof. This is Proposition 2.2 restated: the coprojection ι_1 is conservative, so $T_1 \sqcup T_2$ has exactly the same $L(\Sigma_1)$-consequences as T_1. □

Mere conjunction is invisible to single-language observations. This is the formal content of the intuition that "putting two theories side by side" adds nothing to either one.

We can now define the central concept.

Definition 3.6 (Unified theory)**.** A theory U together with a translation $e\colon T_1 \sqcup T_2 \to U$ is a *unified theory* of T_1 and T_2 if there exists an observation Φ such that Φ has a generative effect at (T_1, T_2) relative to U:

$$\Phi(U) \;>\; \Phi(T_1) \vee_Q \Phi(T_2).$$

Equivalently, U is a unified theory precisely when the translation $e\colon T_1 \sqcup T_2 \to U$ is non-conservative: U proves something in the language of $T_1 \sqcup T_2$ that $T_1 \sqcup T_2$ itself cannot prove.

Remark 3.7 (Three types of new consequences)**.** The new consequences witnessed by a unified theory U fall into three types:

1. **Cross-theoretic consequences:** $U \vdash \varphi$ where φ contains vocabulary from both Σ_1 and Σ_2. These are impossible in $T_1 \sqcup T_2$, since disjoint-signature sentences have no non-trivial mixed consequences. This is the hallmark of genuine unification.

2. **Novel consequences for T_1:** $U \vdash \psi$ where $\psi \in L(\Sigma_1)$ but $T_1 \nvdash \psi$. The unified theory tells us new things about T_1's domain that T_1 alone could not derive.

3. **Novel consequences for T_2:** Symmetric to type (2).

Type (1) is the signature of genuine unification; types (2) and (3) show that the unified theory is also individually richer.

3.3 The adjunction theorem

Definition 3.4 tells us *what* a generative effect is. We now ask: *which observations can detect them?* The answer is given by a structural characterization theorem due to Fong and Spivak [2019, Thm. 1.115].

Definition 3.8 (Galois connection)**.** Let $(P, \leq)$ and $(Q, \leq)$ be preorders. A *Galois connection* between P and Q is a pair of monotone maps

$$P \underset{g}{\overset{f}{\rightleftarrows}} Q$$

satisfying the adjunction condition: for all $p \in P$ and $q \in Q$,

$$f(p) \leq_Q q \quad \iff \quad p \leq_P g(q).$$

We call f the *left adjoint* and g the *right adjoint*, and write $f \dashv g$.

Theorem 3.9 (Characterization of generative effects)**.** *Let $\Phi \colon (P, \leq) \to (Q, \leq)$ be a monotone map between preorders with joins. Then Φ has no generative effects—i.e., $\Phi(a \vee b) = \Phi(a) \vee \Phi(b)$ for all a, b—if and only if Φ is a left adjoint (part of a Galois connection).*

See Fong and Spivak [2019, Thm. 1.115] for the proof. We use only the ($\Leftarrow$) direction (left adjoints preserve joins) and only for binary joins (coproducts), which exist in **Th** (Remark 1.3).

Corollary 3.10 (Characterization of unification-detecting observations)**.** *An observation* $\Phi\colon \mathbf{Th} \to Q$ *can detect the difference between the coproduct* $T_1 \sqcup T_2$ *and a genuine unified theory* U only if Φ *is not a left adjoint. Observations that are left adjoints are insensitive to interaction value: they recover what the parts contribute, but not the surplus generated by their interaction.*

This gives a clean structural criterion. The observations that fail to detect unification are those that "decompose cleanly"—those possessing a right adjoint that reconstructs a theory from an observed phenome. Observations sensitive to unification lack this decomposition property.

Example 3.11. Consider the consequence operator Cn_L for $L \subseteq L(\Sigma_1)$. Its right adjoint

$$g\colon (\mathcal{P}(L), \subseteq) \longrightarrow (\mathbf{Th}, \leq)$$

sends a set of L-sentences S to the weakest theory whose L-consequences include S—namely, the deductive closure of S in L. Since Cn_L is a left adjoint, Theorem 3.9 confirms that it has no generative effects at the coproduct. This recovers Proposition 3.5 from the structural side.

But the point of the adjunction theorem goes further: it tells us that Cn_L is not designed to register interaction value, not just at the coproduct baseline but across unification cases more generally. More broadly, observations possessing a right adjoint preserve joins and therefore do not witness generative effects.

The adjunction theorem thus tells us which "questions you can ask of a theory" are sensitive to whether it is genuinely unified. Questions that decompose—that have the adjoint structure—are insensitive. The questions that detect unification are precisely those that resist decomposition.

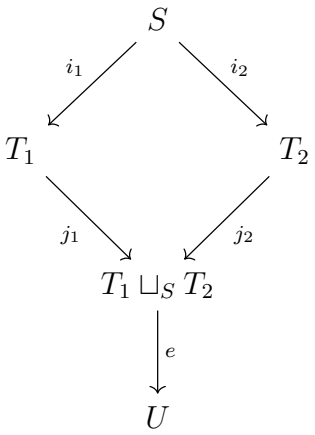
3.4 The pushout case

When theories share a common sub-theory S, the natural baseline is the pushout $T_1 \sqcup_S T_2$ rather than the coproduct.

Definition 3.12 (Unified theory over shared content)**.** A theory U with translation $e\colon T_1 \sqcup_S T_2 \to U$ is a *unified theory of* T_1 *and* T_2 *over* S if there exists an observation Φ such that

$$\Phi(U) \;>\; \Phi(T_1 \sqcup_S T_2).$$

The relationship between the pushout and genuine unification has a layered structure:



Remark 3.13 (Two levels of generativity)**.** It is important to distinguish two phenomena:

1. **Pushout-generativity.** The pushout $T_1 \sqcup_S T_2$ may already exhibit generative effects relative to T_1 and T_2 individually. This occurs when the interaction through the shared core S forces new consequences in $L(\Sigma_S)$—consequences that neither T_1 nor T_2 proved alone, but that follow from combining their respective extensions of S.

2. **Unification-generativity.** U goes beyond what the pushout forces—it introduces bridge principles, identification axioms, or structural connections that go beyond the minimal gluing.

The coproduct (where $S = \varnothing$) has no pushout-generativity at all. The pushout has pushout-generativity but, by definition, no unification-generativity. A genuine unified theory U has both.

This distinction matters in practice. Consider rational-choice economics and rational-choice political science, sharing the sub-theory S of rational choice (utility maximization, preference orderings, strategic interaction). The

pushout $T_{\text{econ}} \sqcup_S T_{\text{pol}}$ is the minimal combination: it contains both theories' vocabularies with their shared decision-theoretic core identified. This pushout may already have pushout-generativity—for instance, if one theory constrains the preference orderings that the other takes as given.

But a substantive *political economy* goes further: it adds bridge principles ("electoral incentives shape fiscal policy," "market structure constrains regulatory capacity") that are not consequences of the pushout. These are unification-generativity—the genuinely new content that makes political economy more than the sum of its parts.

3.5 Toward measuring unification

The framework so far answers the binary question: does a given theory exhibit generative effects? But we may also want to compare unified theories—when is one "more unified" than another?

Adam [2017] develops exactly this kind of tool. His framework extracts algebraic objects—vector spaces, modules—from systems that encode their *generativity*: the potential to produce new phenomena upon interaction. These objects assemble into exact sequences from commutative algebra, and higher-order cohomology groups measure the *type* and *degree* of generativity. Translating this machinery into the setting of theories would yield a measure of how much new content a unified theory produces, and of what kind. We leave this as a direction for future work (Section 6).

What the present framework does establish is the identification of unification with generative effects, the characterization (via the adjunction theorem) of which observations are sensitive to unification, and the distinction between pushout-generativity and unification-generativity. With these tools in hand, we turn in the next section to the normative question: what do generative effects imply for rational preferences over theories?

4 Preferences over Theories

The previous section identified unification with generative effects. We now ask the normative question: what does this identification imply for rational preferences over theories? Drawing on the decision-theoretic tradition of Savage [1954], we develop a framework for theory choice and derive three results that depend essentially on the categorical machinery of Sections 1–3.

4.1 A preference-based framework for theory choice

We adapt the representational structure of Savage's (1954) framework to theory evaluation. An agent faces a *theory-choice problem*: she must evaluate candidate theories, but can only assess them through observations—partial views of what the theories say.

Definition 4.1 (Theory-choice problem)**.** A *theory-choice problem* is a tuple $(C, \mathcal{O}, \succeq)$ where:

- $C \subseteq \mathrm{Ob}(\mathbf{Th})$ is a finite set of candidate theories;
- $\mathcal{O} = \{\Phi_i\}_{i \in I}$ is a family of observations $\Phi_i \colon (\mathbf{Th}, \leq) \to (Q_i, \leq_i)$;
- $\succeq$ is a preference preorder on C.

The observations encode the agent's epistemic goals—the questions she cares about answering. A domain specialist might use only $\{\mathrm{Cn}_{L_1}\}$ (consequences in her field's language). An interdisciplinary researcher might include cross-domain observations $\Phi^{\times}$ that evaluate mixed-vocabulary consequences.

The correspondence to Savage's framework is:

$$\begin{aligned} \text{acts} &\longleftrightarrow \text{candidate theories } T \in C \\ \text{consequence function} &\longleftrightarrow \text{observation map } \Phi \\ \text{consequences} &\longleftrightarrow \text{phenomes } \Phi(T) \in Q \\ \text{expected utility} &\longleftrightarrow \text{aggregate observation value} \end{aligned}$$

The critical departure: in Savage's framework the consequence function is fixed. Here, the *choice of which observations to care about* is itself a parameter—and the adjunction theorem (Theorem 3.9) constrains this choice.

Remark 4.2 (Scope of the Savage analogy)**.** We borrow the *representational structure* of Savage's framework—acts, consequences, preferences—but not its probabilistic engine. Our axioms (A1)–(A4) are deliberately thinner: they concern which observations detect unification, not how unification should update beliefs. Combining the two—deriving subjective probabilities over theories that respect the adjunction structure—would connect our account to Myrvold's (2003) Bayesian treatment and is a direction for future work.

4.2 Axioms on preferences

Definition 4.3 (Admissible preference)**.** A preference $\succeq$ on C is *admissible* if it satisfies:

(A1) Completeness. For all $T, T' \in C$: $T \succeq T'$ or $T' \succeq T$.

(A2) Transitivity. For all $T, T', T'' \in C$: if $T \succeq T'$ and $T' \succeq T''$, then $T \succeq T''$.

(A3) Monotonicity. If there exists a translation $F\colon T \to T'$, then $T' \succeq T$.

(A4) ME-Invariance. If T and T' are Morita equivalent (Definition 1.7), then $T \sim T'$.

Axioms (A1) and (A2) are standard rationality requirements. Axiom (A3) says that theories with greater translational reach are weakly preferred: if T' can interpret everything T can, T' is at least as good. This is a weak form of "more content is better," but note it does *not* say *strictly* better—the question of when strict preference holds is exactly where generative effects enter.

Axiom (A4) is the genuinely new requirement, available only because theories are objects in a category with a principled identity notion. It says that theoretical identity (up to ME) implies evaluative identity. This axiom has no analogue in frameworks that treat theories as unstructured objects.

Remark 4.4. Naive consequence-maximization—preferring whichever theory proves more sentences—would trivially prefer inconsistent theories, which prove everything. Axiom (A3) avoids this by grounding preference in the translation preorder rather than consequence count. The standing consistency assumption (Section 1) provides an additional safeguard.

4.3 Observation-representable preferences

We now define a representation condition linking preferences to observations.

Definition 4.5 (Observation-representable preference)**.** A preference $\succeq$ on C is *representable* by an observation family $\mathcal{O} = \{\Phi_i\}_{i \in I}$ if there exists a strictly monotone aggregator $V\colon \prod_{i \in I} Q_i \to (\mathbb{R}, \leq)$ such that

$$T \succeq T' \quad \Longleftrightarrow \quad V\big(\Phi_i(T)\big)_{i \in I} \;\geq\; V\big(\Phi_i(T')\big)_{i \in I}.$$

Strict monotonicity of V means: increasing any $\Phi_i(T)$ in its own order $\leq_i$, while holding the others fixed, strictly increases V. This rules out degenerate aggregators that ignore some observations entirely.

The first bridge result connects the categorical notion of conservative translation to decision-theoretic indifference.

Proposition 4.6 (Conservativity–indifference bridge)**.** *Let $\succeq$ be representable by $\mathcal{O}$ via a strictly monotone aggregator V.*

(a) *If $e\colon T \to T'$ is conservative and every $\Phi_i \in \mathcal{O}$ evaluates only $L(\Sigma_T)$-consequences (i.e., Φ_i factors through Cn_{L_i} for some $L_i \subseteq L(\Sigma_T)$), then $\Phi_i(T') = \Phi_i(T)$ for every $\Phi_i \in \mathcal{O}$, and hence $T' \sim T$.*

(b) *If $e\colon T \to T'$ is non-conservative, then there exists an observation Φ such that $\Phi(T') > \Phi(T)$.*

Proof. (a) The translation $e\colon T \to T'$ gives $\Phi_i(T) \leq_i \Phi_i(T')$ by monotonicity of Φ_i. For the reverse: since e is conservative, for every $\varphi \in L(\Sigma_T)$ we have $T' \vdash e(\varphi)$ if and only if $T \vdash \varphi$. Each Φ_i factors through Cn_{L_i} for some $L_i \subseteq L(\Sigma_T)$, so $\mathrm{Cn}_{L_i}(T') = \mathrm{Cn}_{L_i}(T)$ and hence $\Phi_i(T') = \Phi_i(T)$ for all i.

(b) Non-conservativity means $T' \vdash \varphi$ for some $\varphi \in L(T)$ with $T \nvdash \varphi$. Take $\Phi = \mathrm{Cn}_{L'}$ for any language L' containing φ; then $\Phi(T') \supsetneq \Phi(T)$. □

Conservative translation is thus decision-theoretically invisible to observations that evaluate the source theory's own language—the natural case for a domain specialist asking what a theory says about *her* domain. The source-language condition is essential: an observation evaluating vocabulary introduced by the extension can see a difference, but such an observation is not evaluating the original theory on its own terms. Part (b) provides the converse: a non-conservative extension is always detectable by some observation.

4.4 The adjunction impossibility theorem

We now prove the section's main result: preferences representable by left-adjoint observations are *structurally incapable* of seeing interaction value in combination.

Definition 4.7 (Adjoint observation family)**.** An observation family $\mathcal{O} = \{\Phi_i\}_{i \in I}$ is *adjoint* if every Φ_i is a left adjoint (part of a Galois connection, Definition 3.8).

Theorem 4.8 (Adjunction impossibility)**.** *Let $\succeq$ be an admissible preference representable by an adjoint observation family $\mathcal{O}$. Then for any theories T_1, T_2:*

$$V(T_1 \sqcup T_2) \;=\; V\big(\Phi_i(T_1) \vee_{Q_i} \Phi_i(T_2)\big)_{i \in I}.$$

*That is, the value of the combination equals the value of the component-wise join of the observations. The combination has **zero interaction value**: the agent sees no surplus from putting the theories together.*

Proof. Since each Φ_i is a left adjoint, the adjunction theorem (Theorem 3.9) gives $\Phi_i(T_1 \sqcup T_2) = \Phi_i(T_1) \vee_{Q_i} \Phi_i(T_2)$ for every $i \in I$. Substituting into the representation:

$$V(T_1 \sqcup T_2) = V\big(\Phi_i(T_1 \sqcup T_2)\big)_{i \in I} = V\big(\Phi_i(T_1) \vee_{Q_i} \Phi_i(T_2)\big)_{i \in I}. \qquad \square$$

Corollary 4.9. *An agent whose preference is representable solely by single-language consequence operators $\{\mathrm{Cn}_{L_1}, \mathrm{Cn}_{L_2}\}$—each a left adjoint (Example 3.11)—sees zero interaction value in any combination.*

Result. Theorem 4.8 and Corollary 4.9 yield a sharp claim: if all observations are adjoint, interaction value is forced to zero.

Interpretation. An adjoint agent *can* still prefer a unified theory U over the coproduct $T_1 \sqcup T_2$, since U may have more content (axiom (A3) yields $U \succeq T_1 \sqcup T_2$). But this extra value is registered only as "more theory," not as specifically unificatory surplus. Through an adjoint lens, bridge-driven gains are indistinguishable from any other non-conservative strengthening with the same single-language profile.

Limitation. The impossibility result is therefore not a charge of irrationality. It is a structural limit: an adjoint observation family cannot register interaction value by construction.

4.5 Observation-relativity and rational disagreement

Proposition 4.10 (Observation-relativity of unification value)**.** *Let T_1, T_2 be theories and U a unified theory (Definition 3.6) whose non-conservative content is purely cross-theoretic: U proves new sentences involving vocabulary from both Σ_1 and Σ_2, but proves no new sentences in $L(\Sigma_1)$ alone or in $L(\Sigma_2)$*

alone beyond what $T_1 \sqcup T_2$ already proves.[4] *There exist admissible preferences $\succeq_A$ and $\succeq_B$, both satisfying* (A1)–(A4)*, such that:*

- $\succeq_A$*:* $U \sim_A T_1 \sqcup T_2$ (indifference);
- $\succeq_B$*:* $U \succ_B T_1 \sqcup T_2$ (strict preference for unification).

Moreover, $\succeq_A$ is representable by an adjoint observation family, while $\succeq_B$ is representable by a family containing at least one non-adjoint observation.

Proof. For $\succeq_A$: take $\mathcal{O}_A = \{\mathrm{Cn}_{L(\Sigma_1)}, \mathrm{Cn}_{L(\Sigma_2)}\}$. By Theorem 4.8, the combination has zero interaction value. Since U's non-conservative content is purely cross-theoretic, U has the same single-language consequence profiles as $T_1 \sqcup T_2$, so $U \sim_A T_1 \sqcup T_2$.

For $\succeq_B$: include a cross-theoretic observation $\Phi^\times$ that evaluates consequences in the joint language $L(\Sigma_1 \cup \Sigma_2)$ restricted to mixed-vocabulary sentences. Since U is a genuine unified theory, there exists such a $\Phi^\times$ with $\Phi^\times(U) > \Phi^\times(T_1 \sqcup T_2)$ (by Definition 3.6 and Remark 3.7). This $\Phi^\times$ is not a left adjoint (since it witnesses a generative effect), and an aggregator giving it positive weight yields $U \succ_B T_1 \sqcup T_2$. □

Two scientists working on the same pair of theories can *rationally disagree* about whether unification is worth pursuing.[5] The framework makes this disagreement structural rather than psychological: the dividing line is adjoint vs. non-adjoint observations, which corresponds to single-domain vs. cross-domain research questions. A domain specialist who only asks "What does this theory tell me about *my* domain?" uses adjoint observations and will rationally see no interaction value. An interdisciplinary researcher who asks cross-domain questions uses non-adjoint observations may assign strict value to unification.

[4] This restriction isolates the effect of unification *per se*. If U also has type (2) or type (3) consequences (Remark 3.7), even adjoint agents may strictly prefer U via axiom (A3)—but they see this as "a stronger theory," not as a benefit of unification specifically.

[5] We are using "observation" in two registers: formally, a monotone map $\Phi\colon \mathbf{Th} \to Q$ (Definition 3.2); informally, a research question that a scientist cares about answering. The bridge between registers is that a research question determines a monotone map: "What does T say about L-sentences?" determines Cn_L; "Does T predict cross-domain phenomenon X?" determines an observation that checks for X in the joint language. The formal results apply to the maps; the interpretive payoff comes from reading the maps as questions.

This connects to a familiar tension in the organization of science. Kitcher [1981] argued that the value of unification lies in the explanatory connections it creates between previously unrelated phenomena. In our terms, these are cross-theoretic consequences (type (1) in Remark 3.7) that become decision-relevant under non-adjoint observations. Myrvold [2003] gave a Bayesian account, showing that unified theories receive higher posterior probability under certain priors. Our account is complementary: it is structural rather than probabilistic, and identifies the *structural conditions on the agent's goals* under which unification is valuable.

4.6 The generativity gap

We close with a quantitative measure of unification value.

Definition 4.11 (Generativity gap)**.** Given a theory-choice problem $(C, \mathcal{O}, \succeq)$ with $\succeq$ representable by $\mathcal{O}$ via aggregator V, the *generativity gap* of U relative to T_1, T_2 under $\mathcal{O}$ is

$$\Delta_{\mathcal{O}}(U; T_1, T_2) \;=\; V\big(\Phi_i(U)\big)_{i\in I} \;-\; V\big(\Phi_i(T_1) \vee_{Q_i} \Phi_i(T_2)\big)_{i\in I}.$$

The gap measures how much better U is than the "virtual conjunction" obtained by evaluating each component separately and combining the results.

Proposition 4.12. $\Delta_{\mathcal{O}}(U; T_1, T_2) = 0$ *for all unified theories* U *if and only if* $\mathcal{O}$ *is adjoint.*

Proof. ($\Leftarrow$) Theorem 4.8: adjoint observations yield zero interaction value, so the gap collapses.

($\Rightarrow$) If some $\Phi_j \in \mathcal{O}$ is non-adjoint, then by the adjunction theorem there exist theories T_1, T_2 with $\Phi_j(T_1 \sqcup T_2) > \Phi_j(T_1) \vee_{Q_j} \Phi_j(T_2)$. Any non-conservative extension U of $T_1 \sqcup T_2$ satisfies $\Phi_j(U) \geq \Phi_j(T_1 \sqcup T_2) > \Phi_j(T_1) \vee_{Q_j} \Phi_j(T_2)$ by monotonicity, so the gap is positive under the strict monotonicity of V. □

The generativity gap is the decision-theoretic analog of Adam's algebraic generativity objects. Where Adam [2017] extracts vector spaces measuring the *type and degree* of generativity, we extract a scalar measuring the *decision-theoretic value* of generativity relative to a particular agent's goals. The gap factors through the non-adjoint components of $\mathcal{O}$: adjoint observations contribute zero, and only non-adjoint observations contribute positively.

5 Examples

We now apply the framework to three examples of increasing complexity: a fully formal propositional case, a semi-formal account of electromagnetism, and a semi-formal treatment of political economy as a pushout. Each example illustrates the machinery of Sections 2–4 and, in particular, the distinction between adjoint and non-adjoint observations.

5.1 Propositional bridge laws

This example is entirely verifiable by truth table. It shows how a single bridge axiom produces a generative effect.

What this example shows. Even the simplest bridge axiom can create cross-domain consequences that adjoint observations cannot see.

Theories. Let $T_1 = (\{p, q\},\ \overline{\{p \to q\}})$ and $T_2 = (\{r, s\},\ \overline{\{r \to s\}})$ be propositional theories with disjoint signatures. The coproduct (Definition 2.1) is

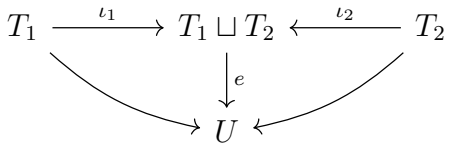
$$T_1 \sqcup T_2 \ = \ \big(\{p, q, r, s\},\ \overline{\{p \to q,\ r \to s\}}\big).$$

By Proposition 2.2, the coprojections are conservative: $T_1 \sqcup T_2$ proves nothing new about $\{p, q\}$ and nothing new about $\{r, s\}$.

Unified theory. Add the bridge axiom $q \to r$:

$$U \ = \ \big(\{p, q, r, s\},\ \overline{\{p \to q,\ r \to s,\ q \to r\}}\big).$$

The inclusion $e\colon T_1 \sqcup T_2 \to U$ is non-conservative, since $U \vdash p \to s$ while $T_1 \sqcup T_2 \nvdash p \to s$. Hence U is a unified theory (Definition 3.6).



Generative effects. The new consequence $p \to s$ is a type (1) cross-theoretic consequence (Remark 3.7): it mixes the T_1-variable p with the T_2-variable s, and is impossible in the coproduct. Types (2) and (3) are absent:

the bridge axiom $q \to r$ is one-directional, and the propositional setting with disjoint variables produces no novel single-language consequences.[6]

Adjoint vs. non-adjoint observations. The consequence operator $\mathrm{Cn}_{\{p,q\}}$ is a left adjoint (Example 3.11). It sees:

$$\mathrm{Cn}_{\{p,q\}}(U) \;=\; \mathrm{Cn}_{\{p,q\}}(T_1 \sqcup T_2) \;=\; \mathrm{Cn}_{\{p,q\}}(T_1).$$

The bridge axiom is invisible: the $\{p, q\}$-observer is indifferent between U and the mere conjunction (Theorem 4.8). Now consider $\Phi^{\times} = \mathrm{Cn}_{\{p,q,r,s\}}$ restricted to mixed-vocabulary sentences. This observation detects $p \to s$ in U but not in $T_1 \sqcup T_2$. The framework predicts exactly this: $\Phi^{\times}$ is not a left adjoint, so it can witness the generative effect (Corollary 3.10).

5.2 Electromagnetism

We sketch, in standard physics notation, how Maxwell's unification of electricity and magnetism exemplifies the framework.[7]

What this example shows. Classical unification can generate all three consequence types in Remark 3.7, and the perceived gain depends on the observation family.

Theories. Let T_E (electrostatics) and T_M (magnetostatics) have the following schematic signatures and axioms:

$$T_E\colon \quad \text{sorts: } \mathrm{Pt}_E, \mathrm{Ch}, \mathbf{E}\text{-fields}; \quad \text{axioms: } \nabla \cdot \mathbf{E} = \rho/\varepsilon_0, \quad \nabla \times \mathbf{E} = \mathbf{0}.$$

$$T_M\colon \quad \text{sorts: } \mathrm{Pt}_M, \mathrm{Cur}, \mathbf{B}\text{-fields}; \quad \text{axioms: } \nabla \times \mathbf{B} = \mu_0 \mathbf{J}, \quad \nabla \cdot \mathbf{B} = 0.$$

Historically, physicists shared spatial vocabulary between these theories. But as a thought experiment, we treat them as having disjoint signatures—separate spatial domains, separate ontologies—so that the coproduct baseline isolates the logical content of unification. The coproduct $T_E \sqcup T_M$ is then

[6] A two-directional bridge $q \leftrightarrow r$ would still produce no type (2) or type (3) consequences in this propositional setting, since neither T_1 nor T_2 has axioms whose conclusions gain new force from the identification. The richer examples below show how types (2) and (3) arise.

[7] Full formalization would require handling differential and geometric structure beyond standard first-order logic. The semi-formal treatment captures the conceptual structure faithfully.

two non-interacting theories: electric fields in one world, magnetic fields in another.

Unified theory. Maxwell's unification identifies the spatial domains ($\mathrm{Pt}_E = \mathrm{Pt}_M$) and replaces the static constraints with dynamical coupling:

$$\begin{aligned} \text{Faraday's law:} \quad & \nabla \times \mathbf{E} = -\,\partial\mathbf{B}/\partial t \qquad (\text{replacing } \nabla \times \mathbf{E} = \mathbf{0}), \\ \text{Displacement current:} \quad & \nabla \times \mathbf{B} = \mu_0\mathbf{J} + \mu_0\varepsilon_0\,\partial\mathbf{E}/\partial t \qquad (\text{extending Ampère's law}). \end{aligned}$$

Write U_{EM} for the resulting theory (Maxwell's equations). The inclusion $e\colon T_E \sqcup T_M \to U_{\mathrm{EM}}$ is non-conservative: U_{EM} proves sentences that $T_E \sqcup T_M$ cannot.

Generative effects—all three types.

1. **Cross-theoretic:** Electromagnetic waves. The coupled equations yield a wave equation mixing $\mathbf{E}$ and $\mathbf{B}$, with propagation speed $c = 1/\sqrt{\mu_0\varepsilon_0}$—the speed of light, derived from purely electromagnetic constants. This consequence involves vocabulary from both T_E and T_M and is impossible in the coproduct.

2. **Novel for T_E:** The displacement current $\mu_0\varepsilon_0\,\partial\mathbf{E}/\partial t$ implies that changing electric fields produce magnetic effects—a phenomenon statable in T_E's extended vocabulary but derivable only in the unified theory.

3. **Novel for T_M:** Faraday induction. Changing magnetic fields produce electric fields. This is a sentence in T_M's extended vocabulary (involving $\mathbf{B}$ and its time derivative) but is derivable only in U_{EM}.

Adjoint vs. non-adjoint observations. An electrostatics specialist whose observation is $\mathrm{Cn}_{L(T_E)}$ uses a left adjoint. In the static limit ($\partial/\partial t = 0$), U_{EM} reduces to T_E: the specialist sees zero interaction value (Theorem 4.8). An electromagnetic theorist who evaluates cross-domain consequences—"Does the theory predict electromagnetic waves?"—uses a non-adjoint observation and sees the relevant generative surplus. The adjunction impossibility theorem helps formalize the old complaint that an electrostatics specialist "wouldn't see the point" of Maxwell's unification.

Remark 5.1 (Iterative unification)**.** The framework composes: U_{EM} can itself be unified with T_{weak} via a pushout over their shared QFT core, yielding

the electroweak theory with its own generative effects (W and Z boson predictions). Each step in the chain toward the Standard Model is a pushout followed by a non-conservative extension, and the generativity gap (Definition 4.11) at each level is independent.

5.3 Political economy as pushout

This example showcases the pushout construction and both levels of generativity (Remark 3.13). It is the most complex example and illustrates the full machinery of Section 4.

What this example shows. Pushout-generativity and unification-generativity come apart, and rational disagreement follows from the structure of observations rather than from failures of rationality.

Theories. Let S (rational choice theory) be a theory with the following shared structure:

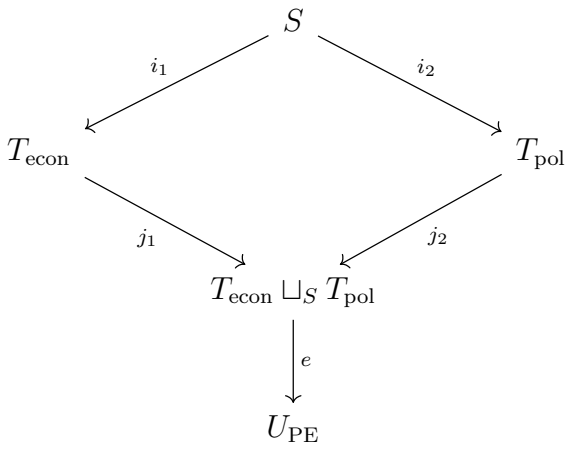
- **Sorts:** X (alternatives), N (individuals), Out (outcomes).
- **Relations:** For each $i \in N$, a binary relation $\succeq_i$ on X.
- **Axioms:** Completeness and transitivity of each $\succeq_i$; existence of a utility representation $u_i \colon X \to \mathbb{R}$ with $x \succeq_i y \iff u_i(x) \geq u_i(y)$; optimization (agents choose $\succeq_i$-maximal alternatives from feasible sets).

Both economics and political science begin here. The difference is what the alternatives *are* and what constrains the feasible set. Two theories extend S:

- T_{econ} (rational-choice economics): alternatives are consumption bundles over a set of goods; additional sorts Good, Price, Firm; additional axioms for budget constraints, profit maximization, and Walrasian equilibrium.
- T_{pol} (rational-choice political science): alternatives are policies or candidates; additional sorts Voter, Policy, Office, Candidate; additional axioms for strategic voting, Downsian convergence [Downs, 1957], and electoral competition.

The embeddings $i_1 \colon S \to T_{\text{econ}}$ and $i_2 \colon S \to T_{\text{pol}}$ are conservative: each theory extends the rational-choice core without weakening it.

Pushout. The pushout $T_{\text{econ}} \sqcup_S T_{\text{pol}}$ (Definition 2.4) identifies the rational-choice core:



By Proposition 2.6, the pushout proves nothing new in proprietary vocabulary: market-clearing remains a theorem of T_{econ} alone; Downsian convergence remains a theorem of T_{pol} alone.

Pushout-generativity (Level 1). The pushout can, however, produce new consequences in the shared vocabulary $L(\Sigma_S)$. If T_{econ} requires convex preferences (for the existence of Walrasian equilibria) and T_{pol} requires single-peaked preferences (for Downsian convergence), then the pushout entails both—a strictly stronger constraint on the shared preference structure than either theory imposed alone. This is pushout-generativity: new content forced by the interaction through S.

Unified theory. A genuine political economy U_{PE} adds bridge principles that go beyond what the pushout forces:

- *Fiscal-electoral bridge:* Incumbent reelection probability is monotone in economic growth rate [Nordhaus, 1975].
- *Regulatory-market bridge:* Concentrated industries have greater lobbying resources, shifting candidate platforms toward industry-favorable regulation.
- *Commercial peace bridge:* Bilateral trade interdependence is monotonically decreasing in interstate conflict probability [Oneal and Russett, 1999, Gartzke, 2007]. Neither trade theory nor international relations theory alone states this sentence—it connects trade flows (economic vocabulary) to militarized dispute probability (political vocabulary).

- *Credible commitment bridge:* Optimal monetary policy requires delegation to an institutionally independent central bank [Rogoff, 1985], because governments face a time-inconsistency problem [Kydland and Prescott, 1977]. Economic performance depends on political institutional design.

Each of these is a unification axiom: a sentence connecting economic and political vocabulary that neither the pushout nor either component theory implies.

Unification-generativity (Level 2)—all three types.

1. **Cross-theoretic:** Political business cycles—economic variables (GDP, unemployment) fluctuate with the electoral calendar—mix economic and political vocabulary and are impossible in the pushout. So does Rodrik's trilemma: deep economic integration, national sovereignty, and democratic politics are jointly inconsistent [Rodrik, 2011]. This is a cross-theoretic impossibility result, structurally analogous to Arrow's theorem, derivable only in the unified theory.

2. **Novel for T_{econ}:** Endogenous policy. Tax rates and regulations are not exogenous parameters but are determined by electoral competition: "In a two-party system, the equilibrium tax rate converges to the median voter's preferred rate." Similarly, the sentence "optimal monetary policy is achievable" is statable in $L(\Sigma_{\text{econ}})$ but derivable only via institutional analysis in U_{PE}—it requires the credible commitment bridge.

3. **Novel for T_{pol}:** Economic feasibility constraints. "A platform promising redistribution beyond the Laffer curve maximum is economically infeasible and electorally unsustainable." A sentence in $L(\Sigma_{\text{pol}})$ derivable only via U_{PE}.

Remark 5.2 (Chichilnisky's topological equivalences)**.** Chichilnisky [1982] showed that the Pareto condition and the existence of a dictator are *topologically equivalent*: any continuous social choice rule satisfying Pareto is homotopic to a dictatorial rule. This connects the First Welfare Theorem (economic vocabulary) to Arrow's Impossibility Theorem (political vocabulary) via the topology of preference spaces [Chichilnisky, 1980, Chichilnisky and Heal, 1983]. In our framework, this is a paradigmatic cross-theoretic

consequence: a sentence in the joint language invisible from within either T_{econ} or T_{pol} alone, detectable only by a non-adjoint observation.

Adjoint vs. non-adjoint observations. This example instantiates Proposition 4.10 concretely.

A pure economist whose observation family is $\mathcal{O}_A = \{\text{Cn}_{L(\Sigma_{\text{econ}})}\}$ employs a left adjoint. By Theorem 4.8, this agent sees zero interaction value: the combination has no surplus that a single-language observation can detect. The economist *can* see type (2) consequences—endogenous policy is visible in the economic language—but cannot distinguish their source. Endogenous policy looks no different from any other non-conservative extension of T_{econ}; the specifically *unificatory character* (that endogenous policy arises from connecting economics to political competition) is invisible through an adjoint lens.

A political economist whose observation family $\mathcal{O}_B$ includes a cross-theoretic observation $\Phi^{\times}$ evaluating mixed-vocabulary consequences employs a non-adjoint observation. This agent sees a generativity gap $\Delta_{\mathcal{O}_B}(U_{\text{PE}};\, T_{\text{econ}}, T_{\text{pol}}) > 0$ (Definition 4.11), detecting that political business cycles, endogenous policy, and feasibility constraints arise specifically from the unification of economics and politics.

The disagreement between the two agents is structural, not irrational: both satisfy axioms (A1)–(A4). The dividing line is adjoint vs. non-adjoint observations, corresponding to single-domain vs. cross-domain research questions. This is Proposition 4.10 in action.[8]

6 Conclusion

This paper has argued that the value of unification is a generative effect, formalized the claim using the categorical framework for scientific theories developed by Halvorson [2019], Barrett and Halvorson [2016], and drawn out its decision-theoretic consequences. The main results are:

- The identification of genuine unification with non-conservative extension, and hence with generative effects in the sense of Fong and Spivak [2019] and Adam [2017] (Section 3).

[8] Full formalization would require specifying quantitative structure (real-valued functions, continuous logic). The semi-formal treatment captures the conceptual structure.

- The adjunction theorem's characterization of which observations can detect unification: those that are not left adjoints (Theorem 3.9, Corollary 3.10).
- The adjunction impossibility theorem: preferences representable by adjoint (single-domain) observation families are structurally blind to interaction value (Theorem 4.8).
- The observation-relativity of unification value: two agents can rationally disagree about whether unification is worth pursuing, with the dividing line being adjoint vs. non-adjoint observations (Proposition 4.10).

We close with limitations and directions for future work.

Limitations

The framework inherits the scope of the Halvorson–Barrett categorical approach: it applies to theories formulated in (multi-sorted) first-order logic. The examples of Section 5 show that the conceptual structure extends naturally to theories with richer mathematical content (differential equations, topological spaces), but full formalization of such cases would require extending the categorical framework to handle geometric or higher-order structure—a nontrivial undertaking.

The decision-theoretic results of Section 4 assume finite sets of candidate theories and finitely many observations. Extending to infinite observation families or continuous aggregators would require topological or measure-theoretic refinements. We also note that the axioms on preferences (Definition 4.3) do not incorporate probabilistic considerations; the relationship between our structural account of unification value and Myrvold's (2003) Bayesian account remains to be worked out.

A more pointed limitation: the framework has nothing to say about the *costs* of unification. Axiom (A3) makes more content weakly better—there is no parsimony axiom, no penalty for a larger signature, no accounting for the increased difficulty of verifying models or the risk of false bridge principles. This is deliberate. Our aim is to characterize when unification produces *epistemic* surplus—new consequences that no mere conjunction could yield—and which agents are positioned to see that surplus. Parsimony, tractability, and falsifiability are genuine virtues, but they are virtues of a different kind: they concern the pragmatics of working with a theory, not the logical structure

of what the theory says. A complete account of theory choice would need to balance generative value against these costs. We isolate the generative side and leave the balancing to future work.

Measuring unification: from scalars to cohomology

The generativity gap (Definition 4.11) is a scalar: it says *how much* better a unified theory is, given a particular observation family and aggregator, but not *in what way*. A natural refinement, flagged in Sections 3 and 4, is to replace the scalar gap with a structured algebraic object.

Adam [2017] develops exactly this kind of tool. His framework extracts vector spaces and modules from interacting systems via exact sequences, and higher cohomology groups measure the *type* and *degree* of generativity. Translating this machinery to the setting of theories would yield a vector-valued generativity gap: not just "U is better than $T_1 \sqcup T_2$ by this much," but "U is better in these specific dimensions—cross-theoretic consequences contributing here, novel single-domain consequences contributing there." This would enable finer comparisons between rival unified theories and connect the decision-theoretic framework to the three types of new consequences identified in Remark 3.7.

Reduction and unification

Nagelian reduction [Nagel, 1961, Schaffner, 1967] is a special case of the framework: bridge laws are non-conservative extensions producing type (3) generative effects (Remark 3.7), and the adjunction theorem explains why they are epistemically valuable. The key difference is that reduction is asymmetric (T_2 reduces *to* T_1), whereas the unification formalized here is symmetric. The relationship between the syntactic approach taken here and structuralist reconstructions of reduction [Dizadji-Bahmani et al., 2010] is a natural question for future work.

Geometric and higher-categorical extensions

De Haro [2026] proposes a complementary *geometric view* of theories as fiber bundles over a moduli space. A synthesis would ask whether generative effects correspond to changes in the topology of the model bundle—though Remark 1.8 suggests that the model functor alone cannot capture the

full syntactic generativity gap. Separately, lifting the framework to the 2-categorical structure of **Th** [Halvorson and Tsementzis, 2017]—where 2-cells relate translations—could yield a coarser but more intrinsic classification of which observations are sensitive to unification.

Implications for scientific practice

The observation-relativity result (Proposition 4.10) has a sociological prediction: the perceived value of interdisciplinary unification may correlate with the prevalence of cross-domain research questions in a community. A community of domain specialists, each asking only single-language questions, will rationally see no value in unification—not because they are narrow-minded, but because their observations are adjoint. A community that routinely asks cross-domain questions will see strict value. This suggests that institutional structures encouraging cross-domain inquiry (interdisciplinary programs, joint seminars, funding for boundary-spanning research) do not merely facilitate unification as a social process—they create the epistemic conditions under which unification becomes rationally valued.

References

Elie M. Adam. *Systems, Generativity and Interactional Effects*. PhD thesis, Massachusetts Institute of Technology, 2017.

Thomas William Barrett. Structure and equivalence. *Philosophy of Science*, 87(5):1184–1196, 2020.

Thomas William Barrett and Hans Halvorson. Morita equivalence. *The Review of Symbolic Logic*, 9(3):556–582, 2016.

Graciela Chichilnisky. Social choice and the topology of spaces of preferences. *Advances in Mathematics*, 37(2):165–176, 1980.

Graciela Chichilnisky. The topological equivalence of the Pareto condition and the existence of a dictator. *Journal of Mathematical Economics*, 9: 223–233, 1982.

Graciela Chichilnisky and Geoffrey Heal. Necessary and sufficient conditions for a resolution of the social choice paradox. *Journal of Economic Theory*, 31(1):68–87, 1983.

Sebastian De Haro. The geometric view of theories. *Synthese*, 207:37, 2026.

Foad Dizadji-Bahmani, Roman Frigg, and Stephan Hartmann. Who's afraid of Nagelian reduction? *Erkenntnis*, 73(3):393–412, 2010.

Anthony Downs. *An Economic Theory of Democracy*. Harper & Row, 1957.

Brendan Fong and David I. Spivak. *An Invitation to Applied Category Theory: Seven Sketches in Compositionality*. Cambridge University Press, 2019.

Erik Gartzke. The capitalist peace. *American Journal of Political Science*, 51(1):166–191, 2007.

Hans Halvorson. *The Logic in Philosophy of Science*. Cambridge University Press, 2019.

Hans Halvorson and Dimitris Tsementzis. Categories of scientific theories. In Elaine Landry, editor, *Categories for the Working Philosopher*, pages 402–429. Oxford University Press, 2017.

Philip Kitcher. Explanatory unification. *Philosophy of Science*, 48(4):507–531, 1981.

Finn E. Kydland and Edward C. Prescott. Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85(3): 473–491, 1977.

Margaret Morrison. *Unifying Scientific Theories: Physical Concepts and Mathematical Structures*. Cambridge University Press, 2000.

Wayne C. Myrvold. A Bayesian account of the virtue of unification. *Philosophy of Science*, 70(2):399–423, 2003.

Ernest Nagel. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World, 1961.

William D. Nordhaus. The political business cycle. *The Review of Economic Studies*, 42(2):169–190, 1975.

John R. Oneal and Bruce Russett. The Kantian peace: The pacific benefits of democracy, interdependence, and international organizations, 1885–1992. *World Politics*, 52(1):1–37, 1999.

Dani Rodrik. *The Globalization Paradox: Democracy and the Future of the World Economy*. Norton, 2011.

Kenneth Rogoff. The optimal degree of commitment to an intermediate monetary target. *The Quarterly Journal of Economics*, 100(4):1169–1189, 1985.

Leonard J. Savage. *The Foundations of Statistics*. Wiley, 1954.

Kenneth F. Schaffner. Approaches to reduction. *Philosophy of Science*, 34 (2):137–147, 1967.

James Owen Weatherall. Part 1: Theoretical equivalence in physics. *Philosophy Compass*, 14(5):e12592, 2019.