The Group Orbit Hypothesis: Inexpensive Variation and Transferable Geometric Learning

Abstract

Controlled variations of a geometric object can communicate which visible differences a recognition task should tolerate. When that relationship recurs across diverse objects, it may become a transferable abstraction: a learner can recognize unfamiliar forms through details it has not separately encountered on each one. The Group Orbit Hypothesis proposes that this learning opportunity overlaps with a computational one. If related realizations reuse substantial prior computation, a mixed allocation of independent bases, variants, and learning effort should improve recognition on unseen bases at fixed total time, or attain required quality sooner, relative to efficient alternatives emphasizing fresh geometry. The prediction concerns a practically meaningful range of coarse geometric tasks and budgets, rather than a universal exchange rate between related and independent examples.

The rationale connects distinct precedents. Exact symmetry supports nuisance averaging and reduced statistical complexity; nonlinear augmentation theory supplies a conditional feature-acquisition mechanism; empirical image studies demonstrate context-dependent exchanges between related views and additional data (S. Chen et al., 2020; Mei et al., 2021; Shen et al., 2022; Geiping et al., 2023). These results motivate the geometric conjecture without establishing its extent. Conditional families accommodate useful noninvertible edits while distinguishing preserved labels from observable information. A limited variance calculation separates coverage within families from coverage across bases, and resource accounting separates cheaper production of unchanged data from improved learning after reallocation. The proposed regime combines visible admissible variation, surviving structural information, recurring nuisance relationships, and consequential acquisition costs. Its practical breadth remains exposed to representative failures. An attained quality–time gain and evidence for transferred nuisance tolerance are related but distinct empirical commitments.

1 Introduction

A bracket can remain recognizable through several arrangements of mounting holes. Useful recognition requires both tolerance and discrimination: the predictor should tolerate the selected hole changes while retaining the relationship between major supporting regions that distinguishes the declared category. A small cut through a defining connection can change that answer, whereas extensive shallow detailing elsewhere may leave it intact. This distinction belongs to a task. The same hole can be incidental to coarse recognition and central to a machining-operation label or an assembly interface. Engineering defeaturing makes this dependence on purpose explicit through estimates of the effect of suppressed geometry on specified analyses (Gopalakrishnan and Suresh, 2007; Buffa et al., 2022).

Suppose training examples of one support category happen to be heavily perforated and those of another mostly smooth. A finite learner may acquire detailing density before the intended structural distinction. This is a geometric instance of shortcut learning: a rule can fit the available observations without supporting the desired transfer (Geirhos et al., 2020). A controlled relative changes the detailing of a known support while retaining evidence for its category. Its repeated label challenges the density-based rule at another input. Across several structurally different supports, that evidence may teach a reusable relationship between changing details and surviving form. It may also teach only how to recognize the familiar supports. Transfer beyond them is the substantive learning claim.

Fresh bases can teach the same nuisance relationship when their details vary broadly. They additionally reveal proportions and structural arrangements absent from existing families. Related realizations hold their source fixed while changing selected cues, which may make some useful constraints economical to repeat. These educational roles overlap; neither route has exclusive access to abstraction. Ordinary supervised examples with valid labels already supply the relevant possibility, as the longstanding use of transformation knowledge and vicinal learning illustrates (Simard et al., 1991; Chapelle et al., 2000). The question is how effectively a finite learner uses those constraints at their actual acquisition price.

Geometry offers a reason that the prices may differ substantially. An additional realization of a known base can preserve expensive common work while changing a cue that matters to recognition. Published experiments in self-adjusting computation demonstrate savings on suitable changes and workloads, together with initial overhead and crossover limits (Acar et al., 2009). This is a generic incremental-computation precedent; it supplies no CAD speedup ratio. The relevant premise is a material reduction in the marginal cost of usable related observations. Computational reuse and learning-relevant difference describe separate properties of those observations.

The Group Orbit Hypothesis predicts that these properties coincide across a practically useful range of coarse geometric learning tasks. A competitive mixture of base diversity, related variation, and learning effort should sometimes be a better purchase than allocating the same total time predominantly to fresh bases. The benefit must concern unfamiliar bases and the task's required quality, after acquisition, preparation, learning, and selection have been counted. More generated files alone do not satisfy the prediction. Existing augmentation–data comparisons already establish that repeated views and empirical data-equivalent benefits are not new general phenomena (Geiping et al., 2023). The proposed geometric contribution is the conjunction of controllable task-preserving variation, reduced acquisition price, and improved recognition under a complete time constraint.

The orbit name identifies a particularly clear mathematical reference: reversible transformations under which an answer remains constant. Many useful geometric changes are broader than this reference. We therefore develop the task and its observation families first, examine what the augmentation literature establishes, and then state the resource comparison and empirical commitment. The argument concerns a plausible learning regime whose occurrence and breadth remain to be determined.

2 Geometric families and transferable abstraction

2.1 The task, the base, and the observation

Let B denote a base geometry drawn from a population P_B, let U describe admissible detail and observation choices, and write

$$X = \Phi(B, U), \qquad Y = h(B), \qquad U \mid B \sim Q(\cdot \mid B). \tag{1}$$

A class contains bases sharing the target h; a family contains the permitted realizations of one base; and X is an observation actually supplied to the learner. Several brackets can have the same label while differing in proportions or arrangements that none of their selected detail changes connects. The model does not assert a unique simple ancestor for every detailed solid. Its conditional law permits base-dependent admissibility: a pocket appropriate to a broad face can destroy a narrow junction elsewhere. The allowed changes and their sampling frequencies are separate choices.

Geometric resources support these distinctions without prescribing one target. ModelNet provides CAD-derived recognition categories, and ABC supplies parametric curves and surfaces with geometric reference information (Wu et al., 2015; Koch et al., 2019). Fusion 360 Gallery connects shapes with human construction sequences, while DeepCAD treats constructive descriptions as learning data (Willis et al., 2021; Wu et al., 2021). Construction-operation annotations in the segmentation data introduced with BRepNet and machining-feature recognition in Hierarchical CADNet instead make local operations or features prediction targets (Lambourne et al., 2021; Colligan et al., 2022). Their relevance is the range of observable and semantic information that geometric learning can retain, rather than evidence for a particular coarse-archetype allocation.

Writing $Y = h(B)$ declares the intended label; it does not establish its recovery from X. For finite-label classification, a nondegenerate conditional label distribution on a set of positive observation probability produces an irreducible error contribution. A possible collision at a null boundary alone does not. The practical requirement is enough surviving information for the stipulated quality, not unique reconstruction of the hidden base. Sparse observations can miss a feature that is conspicuous on the solid, and partial observations can hide a defining connection. ScanObjectNN makes clutter and partial objects substantive recognition challenges; PointCloud-C and ModelNet40-C separately expose the difference between clean classification and robustness to altered observations (Uy et al., 2019; Ren et al., 2022; Sun et al., 2022).

2.2 Why the relation might recur

A coarse target can retain relations among extended regions while permitting many secondary configurations. Positions, extents, and combinations of holes or pockets may change local boundaries substantially without replacing the principal arrangement. Thus the visible variation a task permits can be much broader than the variation a modest training collection happens to contain. This is a geometric expectation about selected targets, not a consequence of small removed volume or confinement to an outer envelope. Thin connections can carry decisive information, and several individually acceptable changes can jointly erase it.

SolidLetters provides a useful neighboring example: character labels span controlled geometric and topological variation derived from fonts (Jayaraman et al., 2021). UVStyle-Net studies style on related solids, with held-out fonts and regenerated evaluation geometry that removes selected within-font randomness (Meltzer et al., 2021). A distinction incidental to character identity can be the object of a style judgment. Procedural data used for CAD-Recode further establishes generated geometry as a substantive reconstruction-learning resource (Rukhovich et al., 2025). These examples support task-relative controlled supervision; they do not measure the proposed coarse-recognition time advantage.

For transfer, a useful nuisance relationship must recur across sufficiently different bases. A unique editing signature can become another way to memorize a source. A recurring relationship offers a stronger opportunity: the learner may discover that selected detailing changes while a larger structural relation persists across unfamiliar proportions and arrangements. Admissibility can remain conditional on geometry. A tolerated pocket need not have identical coordinates or dimensions on every support for its relation to the coarse target to be shared. Ordinary repeated labels can communicate that constraint without supplying a separate correspondence to the learner.

Base diversity and family variation can therefore reinforce one another. More bases reveal contexts in which a nuisance relationship must work; more variants can challenge incidental cues that obscure what those bases share. Once the relationship is well learned, another structural instance may become more valuable than another relative. Independence alone is insufficient for coverage: independent draws from a restricted design population retain its omissions. The promising task also lies between two limits. A reliably observed statistic that already solves the categories leaves little abstraction to acquire; an observation that loses the target leaves nothing recoverable to learn. The proposed opportunity is recognizable but nontrivial structure surrounded by admissible variation that still induces errors.

2.3 Group orbits as a reference case

If a group G acts on the observed input space, its reversible maps respect identity and composition. The orbit $Gx = \{g \cdot x : g \in G\}$ collects observations reachable by the action. An invariant target is constant on each orbit, although several orbits can share a label. Correct symmetry removes distinctions within orbits while leaving the prediction problem across them. Functional invariance and symmetry of a probability law remain different requirements (Bloem-Reddy and Teh, 2020). Eliminating orientation does not determine which intrinsic shapes a learner has yet to encounter.

Four statements consequently need separate support: the observed action exists, the target respects it, the sampled law respects it, and the predictor respects it. Category labels can be rotation-invariant in an upright-biased population, and finite augmentation need not produce global predictor invariance (Lyle et al., 2020). A pose target would instead require an appropriately transformed output. Stability is different again: small changes can require small response changes rather than exact equality throughout a family. Scattering research makes this distinction explicit (Bruna and Mallat, 2013). The task and observation determine which relation is appropriate.

Subtraction supplies a short nongroup example. Removing a fixed region H gives $E_H(S) = S \setminus H$, with $E_H^2 = E_H$ on a domain closed under that operation. A bijective idempotent map is the identity, so any nontrivial such subtraction is noninvertible there. Recognizing the surviving form does not reverse the edit. Likewise, a reversible change of a richer description induces an action on observations only when indistinguishable descriptions remain indistinguishable after every transformation. Otherwise the transformed observation depends on hidden provenance. Conditional families remain meaningful without reversibility or an observed group action; the group theorems below illuminate special cases rather than certify destructive editing.

3 What augmentation and invariant learning establish

3.1 Constraints, averaging, and feature acquisition

The statistical case for related observations extends beyond adding independent information. Tangent Prop expresses selected local insensitivity, and vicinal risk minimization replaces isolated training records with chosen local distributions (Simard et al., 1991; Chapelle et al., 2000). Small-noise squared-error analysis connects training to regularization at the stated approximation order (Bishop, 1995). These constructions encode different knowledge: a target may be constant along selected directions, locally smooth, or predictable under a specified corruption. Their common relevance is that another valid observation can constrain a predictor's behavior where the original records did not.

One constructive linear result makes that opportunity precise. In Section 3.1 of Wu et al. (2020), the favorable ridge-regression construction uses the component of a transformed covariate orthogonal to the original training span, together with an adjusted response and signal, size, and noise conditions. Appending the raw transformed observation need not improve estimation. Their separate mixing analysis concerns regularization. The positive lesson is that selected transformation information can supply a useful new constraint without a new independent source label. Geometric relatives need not satisfy that construction, so it supports a possible mechanism rather than a general raw-augmentation guarantee.

Orbit averaging provides a different explanation. Under exact distributional invariance and square integrability, normalized Haar averaging over a compact group has a conditional-expectation interpretation and reduces variance (S. Chen et al., 2020). Further learning consequences require the associated estimator, complexity, loss, or optimization conditions. Approximate invariance introduces distributional discrepancy and a bias–variance tradeoff. Thus averaging can remove nuisance variability without changing the relevant expectation in the exact case, while a preserved label alone does not establish that probability statement.

Other augmentation models identify alternatives to realizing every view. Dao et al. (2019) connect stochastic augmentation processes to kernels, feature averaging, and variance regularization; their exact process correspondence requires reversibility and a common stationary law. Marginalized corrupted features optimize expectations for specified corruption models (van der Maaten et al., 2013). Analytical augmentation moments give exact treatments of specified input-space and linear quantities, with approximations for nonlinear outputs or losses (Balestriero et al., 2022b). A large family can therefore describe an integration problem whose useful effect does not require exhaustive enumeration. Analytical availability is not itself a measured cost advantage, but such alternatives belong in the resource comparison when they are economical.

Nonlinear feature acquisition supplies a more direct precedent for changing what finite learning discovers. In a stylized high-dimensional patch model, Shen et al. (2022) show that suitable feature-balancing augmentation allows rare informative features to be learned before fitting through dominant noise. The result retains assumptions on feature frequencies, initialization, activation, noise, sample size, and training. Its comparison is at the first constant-margin training fit; it does not rule out later acquisition of those features without augmentation. This is stronger than evaluating a fixed feature map or predictor, but remains a restricted learning result. The geometric inference is that varied secondary cues may help a learner acquire persistent structural evidence soon enough to matter under a budget.

Related-view learning offers adjacent support for this inference. SimCLR emphasizes augmentation composition, and view-selection analysis studies retained task information alongside irrelevant information shared between observations (T. Chen et al., 2020; Tian et al., 2020). Content-identification results preserve a smooth invertible observation map and further generative, style-variation, and objective assumptions (von Kügelgen et al., 2021); they do not establish recovery from arbitrary lossy geometry. A causal interpretation similarly requires the observed augmentation to correspond to a justified latent nuisance intervention (Ilse et al., 2021). These accounts make the relationship between views consequential. They supply motivation for supervised geometric transfer under their own boundaries, rather than a required replacement for ordinary labels.

3.2 Correct symmetry can reduce statistical complexity

Task knowledge can also reduce the functions that remain plausible. Invariant-classifier bounds relate generalization to underlying base-space complexity under specified geometric and robustness assumptions (Sokolić et al., 2017). Linear equivariant and kernel feature-averaging analyses establish conditional expected-risk benefits (Elesedy and Zaidi, 2021; Elesedy, 2021). Their target, action, sampling law, estimator, and noise assumptions are essential. The kernel result, for example, uses an invariant law, a compatible bounded kernel, an invariant bounded target, and independent finite-variance noise; strict improvement requires a nonzero relevant gap. A generalization bound, an actual risk comparison, and a sample-complexity rate answer different questions.

For uniform covariates on high-dimensional spheres or hypercubes, Mei et al. (2021) quantify invariant random-feature and kernel-learning gains under specified action degeneracy and polynomial sample/feature scaling. A separate identity, their Proposition 8, credits Li et al. (2019): full finite-group augmentation in kernel ridge regression agrees with the compatible invariant-kernel estimator on the *same original examples*, under the corresponding regularization convention. This changes how one sampled dataset is used. It does not make augmented observations independent or equate their estimator with one trained on a larger fresh sample.

Geometric rate theory supplies another precise positive case. Tahmasebi and Jegelka (2023) analyze invariant Sobolev regression on smooth connected compact boundaryless manifolds, with smooth isometric compact Lie-group actions, uniform covariates, and stated noise and regularization conditions. Quotient dimension and volume govern the gains. Effective finite actions and positive-dimensional actions differ: a finite factor is distinct from a possible change in the rate exponent. Spherical kernel results likewise depend on spectral and target-regularity assumptions, with the approach to favorable asymptotics depending on the action (Bietti et al., 2021). These results demonstrate substantial statistical value for suitable known structure without supplying rates for arbitrary editing families.

All these gains leave a prediction problem across structural instances. Knowing that a target is constant along one family does not determine its value on an otherwise unrelated base. For example, let $X = (z, u)$, with a group acting only on u and an otherwise unrestricted target $h(z)$. Two invariant targets can agree on every view of each observed z and disagree at an unseen z'. Smoothness or shared features can connect those values, which is why the geometric recurrence argument matters. Repeated observations can help a learner use its base coverage effectively; they cannot make absent task-relevant distinctions appear solely through the name or size of an orbit.

3.3 Finite coverage and the learning rule

Useful augmentation need not exhaust a family. Tahmasebi et al. (2026) analyze finite-dimensional projection estimators under measure-preserving actions. For invariant targets, action-stable approximation spaces, independent original samples and uniform group draws, and the stated moment conditions, their squared-error analysis separates invariant-subspace estimation from finite augmentation. Projection approximation bias remains separate. The result explains how additional views can reduce a particular source of error before exhaustive averaging. It supplies no general training optimum or rate for dependent edit sequences.

The quantity being learned and the way it is used remain consequential. Shao et al. (2022) study PAC learnability with an invariant correct labeling function without requiring an invariant covariate distribution. Their realizability regimes concern its relation to a hypothesis class. Augmented empirical risk minimization can improve sample complexity yet remain suboptimal, and failing to distinguish natural from transformed examples can be costly under relaxed realizability. Yang et al. (2023) compare augmented fitting and consistency regularization under linear rank and subspace assumptions, with separate nonlinear bounds. Such comparisons identify alternatives to ordinary augmented training without implying that ordinary supervision cannot exploit useful families.

Augmenting data, averaging losses, and averaging predictions are therefore different operations. Finite training can leave encouraged invariance unrealized away from the observed inputs (Lyle et al., 2020). Even a perfectly specified nuisance relationship has a price in learning effort. An already effective predictor may benefit from a few additional constraints, while another learner may fail to acquire the relation within the available time. The number of realizable variants alone determines neither situation.

3.4 Empirical exchanges and adverse distributional effects

The closest empirical predecessor to the allocation question is Geiping et al. (2023). Their image-learning comparisons infer augmentation–data exchanges from fitted learning curves. The exchange depends on policy, original data scale, learner, evaluation distribution, and fit range; some inverse comparisons are undefined. Limited fixed views recover substantial benefits in some settings, and their analysis identifies optimization effects beyond enforced invariance. These are contextual comparisons between procedures and risks, not a portable numerical value for a transformed example.

Repeated-view training supplies complementary evidence. Hoffer et al. (2020) report convergence and computational-utilization benefits in their studied large-batch settings, while Fort et al. (2021) report improved image test performance in comparisons matching updates and gradient evaluations. These controls establish meaningful learning-resource comparisons. They do not measure the complete time to obtain and prepare geometric bases. When acquisition is material, a reduction in the price of related geometry can alter the available allocation, but its benefit still depends on the cost of using the resulting observations.

Neighboring domains make that extension plausible. PointAugment and PointWOLF report gains from learned or locally varying point-cloud transformations (Li et al., 2020; Kim et al., 2021). Domain randomization supports synthetic-to-real transfer in specified localization and detection tasks (Tobin et al., 2017; Tremblay et al., 2018). Controlled variability can encourage reliance on persistent evidence. Real engineering details nevertheless obey constraints and co-occurrences that a synthetic law may miss. More variation does not guarantee coverage of the observations or structural sources that define a useful application.

There is also a statistical reason that valid variation can lose. Lin et al. (2024) explain stochastic augmentation through implicit spectral regularization in linear models; changes to bias and variance can be favorable or adverse depending on parameterization, covariance structure, and prediction criterion. Lawrence et al. (2026) isolate a further distinction with invariant linear targets and potentially noninvariant Gaussian covariates. In the full-rank ridgeless regime with n observations, d input dimensions, and $n > d+1$, full training augmentation has no higher expected risk than ordinary fitting, while prediction symmetrization can hurt. Specified overparameterized strong-coupling regimes allow augmentation itself to harm. Distributional asymmetry is consequently neither a universal prohibition nor an irrelevant detail whenever labels remain correct.

For geometry, a detail excluded from the target can still be an efficient predictive cue under a particular evaluation law. Reducing reliance on it may aid transfer when its association changes across intended conditions, yet burden finite learning when that association remains reliable. The application must determine which conditions matter. Choosing a broader nuisance law and lowering the price of sampling an unchanged law have different effects; a favorable outcome from one should not be attributed automatically to the other.

Selecting useful variation is already a substantive learning problem. Domain transformations can have learned compositions, invariance extent can be selected through marginal likelihood, and augmentation distributions can be learned from data (Ratner et al., 2017; van der Wilk et al., 2018; Benton et al., 2020). Affinity and diversity jointly describe empirical augmentation tradeoffs (Gontijo-Lopes et al., 2021). Such evidence supports deliberate policy choice without making a diversity measure a certificate of semantic validity or transfer. Selection effort belongs in a competitive resource account.

Quality also needs a stable meaning. AugMix improves common-corruption robustness in its tested image settings, whereas spatial-robustness studies and augmentation-margin lower bounds limit stronger extrapolations (Hendrycks et al., 2020; Engstrom et al., 2019; Rajput et al., 2019). Aggregate gains can coexist with class-specific harms (Balestriero et al., 2022a). Better average

Table 1: Distinct support for the hypothesis. These comparisons motivate different parts of the geometric claim; their favorable effects cannot be multiplied into a speedup.

Evidence	Established comparison	Open geometric question
Orbit averaging (S. Chen et al., 2020)	Variability under stated action and probability assumptions.	Which useful nuisance expectations do the families convey?
Invariant complexity (Mei et al., 2021; Tahmasebi and Jegelka, 2023)	Statistical gains in specified invariant learning regimes.	How much shared structure can be used across new bases?
Feature acquisition (Shen et al., 2022)	A conditional change in what finite nonlinear learning acquires.	Do varied details help reveal transferable structural evidence?
Finite coverage (Tahmasebi et al., 2026)	Partial averaging for specified projection estimators.	How much variation is useful to the actual learner?
Adverse distributional effects (Lin et al., 2024; Lawrence et al., 2026)	Helpful and harmful fitting under different laws and regimes.	Which nuisance tolerance benefits the intended population?
Views and additional data (Geiping et al., 2023; Hoffer et al., 2020; Fort et al., 2021)	Empirical quality and learning-resource exchanges.	Does cheaper geometric acquisition improve total time to quality?

recognition under a declared law is worthwhile; it does not certify every admissible edit. Difficult classes and nuisance conditions must remain part of quality when they determine the task's usefulness.

4 Dependence and the cost of learning

An elementary variance identity clarifies the two statistical resources. Let $W_1, \ldots, W_n$ be iid family states containing the base and all shared randomness, including shared label or persistent observation noise. Given each W_i, draw m iid views from the specified conditional law, independently across families. Fix a predictor independently of these evaluation draws and suppose its losses L_{ij} have finite second moments. Define $V_b = \mathrm{Var}(\mathbb{E}[L \mid W])$ and $V_v = \mathbb{E}[\mathrm{Var}(L \mid W)]$. Then

$$\widehat{R}_{n,m} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} L_{ij}, \qquad \mathrm{Var}(\widehat{R}_{n,m}) = \frac{V_b}{n} + \frac{V_v}{nm}. \tag{2}$$

More views reduce the within-family contribution; more bases reduce both terms. At the same record count, iid fresh family–view pairs with the same single-example marginal law instead give variance $(V_b + V_v)/(nm)$. Equal labels do not imply equal mean losses across bases, and shared noisy labels remain shared measurements.

Equation (2) evaluates a fixed predictor. A separately trained random predictor can be conditioned upon; fitting these same records introduces additional dependencies, so the identity is not a learning curve or a generalization bound for an unspecified learner. Conditional view dependence introduces covariance terms, and a different nuisance law can change the expectation. Unequal family sizes under record-wise averaging also change base weights. Identical one-example marginals can coexist with different joint training laws, which is why nominal sample count alone does not settle the value of repeated supervision.

For cost, let c_b represent obtaining the first usable realization of a fresh base, c_v an additional relative, and e the learning effort. A schematic account is

$$C(n, m, e) = n\{c_b + (m-1)c_v\} + C_{\text{learn}}(n, m, e) + C_{\text{other}}(n, m, e). \tag{3}$$

The opportunity is a consequential regime with $c_v < c_b$. Preparation, handling, validity checks, selection, and other necessary work remain in the effective acquisition prices or the remaining terms. Costs can vary across geometry, and concurrent work can overlap or compete for resources. This is explanatory accounting rather than an elapsed-time law. Cheap observations still require learning effort, and savings behind another bottleneck may barely change completion time.

Two comparisons must be distinguished. The first preserves requested bases, relatives, targets, and the learning procedure. Producing exactly those observations, or preserving their full joint sampling law, more cheaply changes their acquisition price without adding a new dependence penalty. Both sides use the same statistical information. This can establish a computational improvement even when the collection itself was not the best possible learning purchase.

The second fixes total time and changes the allocation. Fresh bases, more relatives, and further learning now compete. Efficient alternatives should receive comparably broad nuisance opportunities, ordinary augmentation, and economical uses of known structure, together with reasonable learning and selection effort. Point-cloud classification comparisons show why augmentation, losses, and evaluation choices can materially alter apparent progress (Goyal et al., 2021). A weak comparator denied ordinary efficiencies would not establish the proposed resource advantage.

Reducing a price weakly expands affordable choices when previous choices remain available. It does not imply a strict improvement in the best choice or that a practical procedure finds one. Newly affordable allocations may be redundant, hard to learn from, or inferior to more base coverage. The substantive prediction is that useful new allocations occur often enough to matter in the intended geometric domain. It concerns an attained quality–time relation, including the cost of selecting the procedure, rather than an allocation supplied by an uncharged optimum.

5 The Group Orbit Hypothesis as an empirical commitment

Group Orbit Hypothesis. *For representative coarse geometric recognition tasks with diverse base structure and substantial visible secondary variation that retains sufficient target information, inexpensive related generation will create a practically useful regime in which a competitive mixed allocation of independent bases, related realizations, and learning effort improves recognition on unseen bases at fixed total time, or attains prespecified quality sooner, compared with efficient alternatives emphasizing fresh bases. The advantage is expected across a meaningful range of budgets or quality requirements. A principal proposed mechanism is that recurring within-family changes teach nuisance relationships that transfer while preserving discrimination.*

This forecasts the practical occurrence of a favorable regime, not merely its logical possibility. Coarse target granularity, visible admissibility, surviving information, structural diversity, recurring nuisance relationships, and meaningful acquisition costs are properties that can be investigated before a winning result is observed. Their conjunction motivates the prediction; it does not define success. The intended task lies between a readily solved distinction and an observation that has lost its target. A convenient taxonomy or isolated favorable budget cannot by itself establish the claim's breadth.

Let π denote a specified feasible learning and allocation policy and $R_\pi(t)$ its expected risk after total elapsed time t, including training randomness and evaluation on independent test bases under the declared observation law. For a risk threshold r, define

$$\tau_\pi(r) = \inf\{t : R_\pi(t) \le r\}, \qquad \inf \varnothing = +\infty. \tag{4}$$

This is a threshold of expected risk, not the expected first successful time of an individual random run. The claim concerns lower risk at agreed times or lower $\tau_\pi(r)$ over a useful range. Required class-specific performance, nuisance tolerance, or reliability can accompany the scalar criterion. Neither monotone learning curves nor attainability of every threshold is assumed. The target population, observation law, quality requirements, comparison strategies, and range of interest need substantive meanings before outcomes select their interpretation.

Evidence for the resource outcome and evidence for its proposed explanation have different demands. A quality–time improvement can arise through regularization, better optimization, or changed weighting without establishing transferred nuisance tolerance. Conversely, improved tolerance can fail to earn its total cost. Evidence for the proposed mechanism concerns reduced reliance on irrelevant detail on unfamiliar bases while retaining response to target-changing structure. Family consistency alone is inadequate: a constant predictor is invariant and fails at recognition. The main outcome is useful discrimination on new base information under the complete resource account.

Transfer must match the independence level required by the application. New descendants of training bases measure within-family generalization. New bases from the declared population address the central claim; independently designed shapes or another acquisition process test further transfer. Nominally distinct bases may share design templates, requiring separation at that higher level. Independence also does not ensure representativeness. Many relatives of compact brackets cannot supply an absent population of thin-sheet supports or unfamiliar structural combinations merely by increasing the record count.

The anticipated return from further relatives should depend on unresolved learning needs. Once a nuisance relationship is acquired, additional bases or learning effort can become preferable. New bases can also introduce contexts in which familiar nuisance relations are harder to use. These interactions make a useful mixed allocation plausible without fixing a ratio, requiring smooth returns, or promising superiority under unlimited data. Existing invariant predictors, conventional augmentation, or economical integration may already communicate most of the useful knowledge; credible comparisons must allow them to do so.

Persistent absence of mixed-allocation gains across representative intended tasks would weaken the practical claim, especially when generation is verified cheap, variation is substantial and admissible, target information survives, and nuisance-related errors remain. Gains that disappear after separating source families or counting complete time would fail to support the stated outcome. Benefits confined to familiar families would particularly weaken the transfer explanation. Missing structure or observation ambiguity can explain specific failures, but losing cases cannot all be excluded afterward by redefining the task domain or calling every unsuccessful variation uninformative. The hypothesis risks a substantive prediction: controlled geometric relationships will often be cheap and learnable enough to improve recognition beyond their sources.

References

Umut A. Acar, Guy E. Blelloch, Matthias Blume, Robert Harper, and Kanat Tangwongsan (2009). An experimental analysis of self-adjusting computation. *ACM Transactions on Programming Languages and Systems* 32(1):3:1–3:53.

Randall Balestriero, Léon Bottou, and Yann LeCun (2022a). The Effects of Regularization and Data Augmentation are Class Dependent. *Advances in Neural Information Processing Systems* 35.

Randall Balestriero, Ishan Misra, and Yann LeCun (2022b). A Data-Augmentation Is Worth A Thousand Samples: Analytical Moments And Sampling-Free Training. *Advances in Neural Information Processing Systems* 35.

Gregory Benton, Marc Finzi, Pavel Izmailov, and Andrew G. Wilson (2020). Learning Invariances in Neural Networks from Training Data. *Advances in Neural Information Processing Systems* 33.

Alberto Bietti, Luca Venturi, and Joan Bruna (2021). On the Sample Complexity of Learning under Geometric Stability. *Advances in Neural Information Processing Systems* 34:18673–18684.

Christopher M. Bishop (1995). Training with Noise is Equivalent to Tikhonov Regularization. *Neural Computation* 7(1):108–116.

Benjamin Bloem-Reddy and Yee Whye Teh (2020). Probabilistic Symmetries and Invariant Neural Networks. *Journal of Machine Learning Research* 21(90):1–61.

Joan Bruna and Stéphane Mallat (2013). Invariant Scattering Convolution Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8):1872–1886.

Annalisa Buffa, Ondine Chanon, and Rafael Vázquez (2022). Analysis-aware defeaturing: Problem setting and a posteriori estimation. *Mathematical Models and Methods in Applied Sciences* 32(2):359–402.

Olivier Chapelle, Jason Weston, Léon Bottou, and Vladimir Vapnik (2000). Vicinal Risk Minimization. *Advances in Neural Information Processing Systems* 13:416–422.

Shuxiao Chen, Edgar Dobriban, and Jane H. Lee (2020). A Group-Theoretic Framework for Data Augmentation. *Journal of Machine Learning Research* 21(245):1–71.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton (2020). A Simple Framework for Contrastive Learning of Visual Representations. *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research 119:1597–1607.

Andrew R. Colligan, Trevor T. Robinson, Declan C. Nolan, Yang Hua, and Weijuan Cao (2022). Hierarchical CADNet: Learning from B-Reps for Machining Feature Recognition. *Computer-Aided Design* 147:103226.

Tri Dao, Albert Gu, Alexander Ratner, Virginia Smith, Chris De Sa, and Christopher Ré (2019). A Kernel Theory of Modern Data Augmentation. *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research 97:1528–1537.

Bryn Elesedy (2021). Provably Strict Generalisation Benefit for Invariance in Kernel Methods. *Advances in Neural Information Processing Systems* 34.

Bryn Elesedy and Sheheryar Zaidi (2021). Provably Strict Generalisation Benefit for Equivariant Models. *Proceedings of the 38th International Conference on Machine Learning*, Proceedings of Machine Learning Research 139:2959–2969.

Logan Engstrom, Brandon Tran, Dimitris Tsipras, Ludwig Schmidt, and Aleksander Madry (2019). Exploring the Landscape of Spatial Robustness. *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research 97:1802–1811.

Stanislav Fort, Andrew Brock, Razvan Pascanu, Soham De, and Samuel L. Smith (2021). Drawing Multiple Augmentation Samples Per Image During Training Efficiently Decreases Test Error. *arXiv preprint arXiv:2105.13343*. Revised February 2022.

Jonas Geiping, Micah Goldblum, Gowthami Somepalli, Ravid Shwartz-Ziv, Tom Goldstein, and Andrew Gordon Wilson (2023). How Much Data Are Augmentations Worth? An Investigation into Scaling Laws, Invariance, and Implicit Regularization. *International Conference on Learning Representations*.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence* 2:665–673.

Raphael Gontijo-Lopes, Sylvia J. Smullin, Ekin D. Cubuk, and Ethan Dyer (2021). Tradeoffs in Data Augmentation: An Empirical Study. *International Conference on Learning Representations*.

Sankara Hari Gopalakrishnan and Krishnan Suresh (2007). A formal theory for estimating defeaturing-induced engineering analysis errors. *Computer-Aided Design* 39(1):60–68.

Ankit Goyal, Hei Law, Bowei Liu, Alejandro Newell, and Jia Deng (2021). Revisiting Point Cloud Shape Classification with a Simple and Effective Baseline. *Proceedings of the 38th International Conference on Machine Learning*, Proceedings of Machine Learning Research 139:3809–3820.

Dan Hendrycks, Norman Mu, Ekin D. Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan (2020). AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty. *International Conference on Learning Representations*.

Elad Hoffer, Tal Ben-Nun, Itay Hubara, Niv Giladi, Torsten Hoefler, and Daniel Soudry (2020). Augment Your Batch: Improving Generalization Through Instance Repetition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8129–8138.

Maximilian Ilse, Jakub M. Tomczak, and Patrick Forré (2021). Selecting Data Augmentation for Simulating Interventions. *Proceedings of the 38th International Conference on Machine Learning*, Proceedings of Machine Learning Research 139:4555–4562.

Pradeep Kumar Jayaraman, Aditya Sanghi, Joseph G. Lambourne, Karl D. D. Willis, Thomas Davies, Hooman Shayani, and Nigel Morris (2021). UV-Net: Learning From Boundary Representations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11703–11712.

Sihyeon Kim, Sanghyeok Lee, Dasol Hwang, Jaewon Lee, Seong Jae Hwang, and Hyunwoo J. Kim (2021). Point Cloud Augmentation With Weighted Local Transformations. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 548–557.

Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo (2019). ABC: A Big CAD Model Dataset for Geometric Deep Learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9601–9611.

Joseph G. Lambourne, Karl D. D. Willis, Pradeep Kumar Jayaraman, Aditya Sanghi, Peter Meltzer, and Hooman Shayani (2021). BRepNet: A Topological Message Passing System for Solid Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12773–12782.

Hannah Lawrence, Elyssa Hofgard, Vasco Portilheiro, Yuxuan Chen, Tess Smidt, and Robin Walters (2026). To Augment or Not to Augment? Diagnosing Distributional Symmetry Breaking. *International Conference on Learning Representations*.

Zhiyuan Li, Ruosong Wang, Dingli Yu, Simon S. Du, Wei Hu, Ruslan Salakhutdinov, and Sanjeev Arora (2019). Enhanced Convolutional Neural Tangent Kernels. *arXiv preprint arXiv:1911.00809*.

Ruihui Li, Xianzhi Li, Pheng-Ann Heng, and Chi-Wing Fu (2020). PointAugment: An Auto-Augmentation Framework for Point Cloud Classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6378–6387.

Chi-Heng Lin, Chiraag Kaushik, Eva L. Dyer, and Vidya Muthukumar (2024). The good, the bad and the ugly sides of data augmentation: An implicit spectral regularization perspective. *Journal of Machine Learning Research* 25(91):1–85.

Clare Lyle, Mark van der Wilk, Marta Kwiatkowska, Yarin Gal, and Benjamin Bloem-Reddy (2020). On the Benefits of Invariance in Neural Networks. *arXiv preprint arXiv:2005.00178*.

Song Mei, Theodor Misiakiewicz, and Andrea Montanari (2021). Learning with invariances in random features and kernel models. *Proceedings of the 34th Conference on Learning Theory*, Proceedings of Machine Learning Research 134:3351–3418.

Peter Meltzer, Hooman Shayani, Amir Khasahmadi, Pradeep Kumar Jayaraman, Aditya Sanghi, and Joseph Lambourne (2021). UVStyle-Net: Unsupervised Few-Shot Learning of 3D Style Similarity Measure for B-Reps. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9690–9699.

Shashank Rajput, Zhili Feng, Zachary Charles, Po-Ling Loh, and Dimitris Papailiopoulos (2019). Does Data Augmentation Lead to Positive Margin? *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research 97:5321–5330.

Alexander J. Ratner, Henry R. Ehrenberg, Zeshan Hussain, Jared Dunnmon, and Christopher Ré (2017). Learning to Compose Domain-Specific Transformations for Data Augmentation. *Advances in Neural Information Processing Systems* 30.

Jiawei Ren, Liang Pan, and Ziwei Liu (2022). Benchmarking and Analyzing Point Cloud Classification under Corruptions. *Proceedings of the 39th International Conference on Machine Learning*, Proceedings of Machine Learning Research 162:18559–18575.

Danila Rukhovich, Elona Dupont, Dimitrios Mallis, Kseniya Cherenkova, Anis Kacem, and Djamila Aouada (2025). CAD-Recode: Reverse Engineering CAD Code from Point Clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9801–9811.

Han Shao, Omar Montasser, and Avrim Blum (2022). A Theory of PAC Learnability under Transformation Invariances. *Advances in Neural Information Processing Systems* 35.

Ruoqi Shen, Sébastien Bubeck, and Suriya Gunasekar (2022). Data Augmentation as Feature Manipulation. *Proceedings of the 39th International Conference on Machine Learning*, PMLR 162, pp. 19773–19808.

Patrice Simard, Bernard Victorri, Yann LeCun, and John Denker (1991). Tangent Prop - A formalism for specifying selected invariances in an adaptive network. *Advances in Neural Information Processing Systems* 4:895–903.

Jure Sokolić, Raja Giryes, Guillermo Sapiro, and Miguel Rodrigues (2017). Generalization Error of Invariant Classifiers. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research 54:1094–1103.

Jiachen Sun, Qingzhao Zhang, Bhavya Kailkhura, Zhiding Yu, Chaowei Xiao, and Z. Morley Mao (2022). Benchmarking Robustness of 3D Point Cloud Recognition Against Common Corruptions. *arXiv preprint arXiv:2201.12296.*

Behrooz Tahmasebi and Stefanie Jegelka (2023). The Exact Sample Complexity Gain from Invariances for Kernel Regression. *Advances in Neural Information Processing Systems* 36.

Behrooz Tahmasebi, Melanie Weber, and Stefanie Jegelka (2026). Data Augmentation: A Fourier Analysis Perspective. *Proceedings of the 39th Conference on Learning Theory*, Proceedings of Machine Learning Research 336:6114–6155.

Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola (2020). What Makes for Good Views for Contrastive Learning? *Advances in Neural Information Processing Systems* 33.

Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel (2017). Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield (2018). Training Deep Networks With Synthetic Data: Bridging the Reality Gap by Domain Randomization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 969–977.

Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Duc Thanh Nguyen, and Sai-Kit Yeung (2019). Revisiting Point Cloud Classification: A New Benchmark Dataset and Classification Model on Real-World Data. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1588–1597.

Laurens van der Maaten, Minmin Chen, Stephen Tyree, and Kilian Weinberger (2013). Learning with Marginalized Corrupted Features. *Proceedings of the 30th International Conference on Machine Learning*, Proceedings of Machine Learning Research 28(1):410–418.

Mark van der Wilk, Matthias Bauer, ST John, and James Hensman (2018). Learning Invariances using the Marginal Likelihood. *Advances in Neural Information Processing Systems* 31:9938–9948.

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello (2021). Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style. *Advances in Neural Information Processing Systems* 34:16451–16467.

Karl D. D. Willis, Yewen Pu, Jieliang Luo, Hang Chu, Tao Du, Joseph G. Lambourne, Armando Solar-Lezama, and Wojciech Matusik (2021). Fusion 360 Gallery: A Dataset and Environment for Programmatic CAD Construction from Human Design Sequences. *ACM Transactions on Graphics* 40(4):54:1–54:24.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao (2015). 3D ShapeNets: A Deep Representation for Volumetric Shapes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1912–1920.

Sen Wu, Hongyang Zhang, Gregory Valiant, and Christopher Ré (2020). On the Generalization Effects of Linear Transformations in Data Augmentation. *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research 119:10410–10420.

Rundi Wu, Chang Xiao, and Changxi Zheng (2021). DeepCAD: A Deep Generative Network for Computer-Aided Design Models. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6772–6782.

Shuo Yang, Yijun Dong, Rachel Ward, Inderjit S. Dhillon, Sujay Sanghavi, and Qi Lei (2023). Sample Efficiency of Data Augmentation Consistency Regularization. *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research 206:3825–3853.