Uninformative rungs: an order-theoretic stopping criterion for nested reference sets

García Hurtado, Alexis
ORCID 0009-0003-4636-8206

Abstract

When an object cannot be resampled and no generating process is available, a null hypothesis has to be built by declaring what may be permuted and what must be held fixed. The declaration is not unique, so the natural response is to report a family of nulls rather than one. A family with no principle behind it can be enumerated and cannot be terminated.

We give an order-theoretic stopping criterion. Rungs of such a family are partially ordered by containment of their reference sets, and for a fixed statistic the rungs on which that statistic is uninformative, meaning constant on the reference set, form a down set in that order. Descending past such a rung therefore produces arithmetic and not evidence, and the family can be truncated there along that path. The result assumes no group, no model, no alternative and no level, and it is proved in one line from set containment. We use *uninformative* throughout in this sense; it is not the Bayesian term of art, which describes a prior. A *rung* is a reference set with a position in that order, and a *null* is the hypothesis a rung carries; both words are in the title and both are demarcated in section 1.

We do not offer localisation as a contribution. Reading where a signal sits from the rung at which it stops surviving is stated and exercised in Theiler et al. (1992), and section 2 says where. What is new is the criterion for stopping and the separation that fixes its scope, not the reading.

We also give the reason a family is reported entire rather than reduced to one rung: the three grounds the literature offers for selecting a single conditioning each require something a unique object with no generating process does not supply, and section 2.8 sets them out.

We separate four reasons a rung can fail to reject and show that the two which are conditions on the reference set descend under containment while the two which are conditions on where the observed value falls do not. The scope of the criterion is therefore located by that separation rather than confessed. Where the rungs are orbits of subgroups the order is computable, which is a corollary and not the result; the exact-sampling property such rungs carry is a published result of others and is cited as theirs.

The framework is instantiated on a worked case: the King Wen sequence, the received ordering of the sixty four hexagrams of the I Ching, and a published family of six rungs over it. The six were presented as a sequence of increasing structure; four of their fifteen pairwise relations are incomparable, so they are a partial order and not a chain, established by exhibited witnesses. A preregistered experiment, committed before computation, refutes the chain reading at four ordered pairs. Two of the four are evidence about the object and rest on a statistic nobody built for this

purpose; the other two are demonstrations that the control behaves, and one of those two was not predicted.

Which rungs are orbits of subgroups is itself decided, by a second theorem with its own proof, and the case exercises one side of the boundary that theorem draws and cites the other. One of its two statistics could not have rejected at any rung whatever the data, for a reason computable before a draw, and we report that rather than omit it. Nothing here is a claim about the King Wen sequence: the statistics were chosen for their behaviour under the nulls, no reported percentile is extreme, and none is offered as a finding about the object.

1. Introduction

A rung of a ladder of nulls is *uninformative* for a statistic when the statistic takes the same value on every element of that rung's reference set. The word is used here in that sense and in no other. In particular it is not the Bayesian term of art: an *uninformative prior* is a prior distribution, this paper defines a property of a *reference set relative to a statistic*, and the two are unrelated. The collision is in the word only, and it is declared here because the word appears in the title of this paper before it appears in its text.

The other word of the title is demarcated in the same place and for the same reason. A **rung** is an element of the order this paper is about: a reference set with a position relative to the others. A **null** is the hypothesis a rung carries. Every rung is a null; not every null is a rung of anything, because a null on its own has no position. The two are used in that sense throughout and are not interchangeable.

1.1 The problem

Some objects cannot be resampled. There is one received arrangement of the sixty four hexagrams of the I Ching, one observed districting plan, one measured network. When the object is unique and no generating process is available, the usual apparatus has nothing to condition on: there is no population from which the observation is a draw, and no model whose parameters could be estimated and perturbed.

The standard response is to build a null by declaring what may be permuted and what must be held fixed. This works, and it immediately raises a harder question, because the declaration is not unique. A different choice of what to hold fixed gives a different null, a different reference set and a different answer, and nothing in the data selects among them.

A natural move is to stop choosing and report a family: a sequence of nulls, each conceding more structure than the last. This is done in practice and it is where the difficulty actually lives. A family with no principle behind it can be enumerated and cannot be terminated. Which rung's answer is the finding? What does it mean when a signal survives one rung and not the next? And when does adding a further rung stop buying anything at all?

1.2 What this paper does

This paper gives an order-theoretic answer to the third question and, through it, to the first two.

The rungs of such a family are partially ordered by containment of their reference sets. For a fixed statistic, the rungs on which that statistic is uninformative form a down set in that order: once a statistic is constant on a reference set it is constant on every reference set inside it, so descending past such a rung produces arithmetic rather than evidence. The complement is an up set, and its minimal elements are the most constrained rungs at which the statistic still varies. That set of minimal elements is what a ladder locates, and it is generally a set and not a single rung.

Section 2.8 answers the objection that one rung must be the right conditioning. Three grounds for choosing one are available in the published literature, and each requires something this kind of object does not supply, so the family is reported entire because choosing is not available rather than because it is hard.

The result requires nothing: no group, no model, no alternative hypothesis, no level, and no distributional assumption. It is a statement about set containment and it is proved in one line. What the paper adds beyond it is the accounting around it, which is where the work is: a separation of the reasons a rung can fail to reject, only one of which the result covers; a statement of what the order costs to use; and a characterisation of the case in which the order is computable, which is a second theorem and is second.

1.3 The case study, and what it is for

The framework is instantiated on a worked case: the King Wen sequence, the received ordering of the sixty four hexagrams of the I Ching, and a published family of six rungs over it. The case is not an illustration appended to a theory. It is a third of this paper and it does three things the theory alone cannot.

It shows the order is not obvious. The six rungs were published as a sequence of increasing structure, which describes a chain; they are not a chain, and four of their fifteen pairwise relations are incomparable. The relations were established by exhibiting a witness in one reference set and not the other, in both directions, for every pair.

It shows the partial order has an observable consequence. A preregistered experiment fixes, in writing and before computation, two statistics and their predicted behaviour at each rung. The chain reading requires a constancy to propagate at four ordered pairs and it does not propagate at any of them. Two of the four are evidence about the object and rest on a statistic nobody would have built for this purpose; the other two are checks that the instrument behaves, counted separately for that reason, and one of those was not among the predictions.

And it shows what the framework costs on a real object: every reference set is enormous, none is near the level bound, and one of the two statistics could not have rejected at any rung whatever the data, for a reason computable before a single draw.

1.4 What this paper does not do

The case study exercises one side of the boundary drawn by the second theorem and cites the other: all six of its rungs are generated by moves defined everywhere, and none by moves defined only where a local pattern occurs. The theorem is proved for both sides and demonstrated on one.

Two of the historical orderings the case study compares are taken from works that were not consulted in preparing this paper, and section 8 says which and what follows.

Nothing here is a claim about the King Wen sequence. The statistics used were chosen for their behaviour under the nulls and not for what they might say about the object, no percentile reported is extreme, and none is offered as a finding about the sequence.

1.5 Plan

Section 2 places the result among published work in eight subsections: the 1992 hierarchy that anticipates part of it and what it does not do, the three axes on which the antecedents sit, one demarcation for each axis, and the argument that a family is reported entire because the three published grounds for choosing are unavailable at once. Section 3 presents the object. Section 4 is the formal core. Sections 5 and 6 give the preregistered experiment and the methods, section 7 the results, and section 8 the discussion and the limitations. Section 9 lists the references, with the version read for each, what each entry rests on and what was not obtained, and section 10 says where the data, the code and the record of this work are deposited.

2. Related work and demarcation

2.1 The hierarchy of null hypotheses is not new, and it is from 1992

Ordering nulls rather than choosing one is published, and it is published under that description. Theiler, Eubank, Longtin, Galdrikian and Farmer (1992) devote their own section 2.2, titled *Toward a hierarchy of null hypotheses*, to five nulls presented in order, each preserving more of the observed series than the last: temporally independent data, an Ornstein-Uhlenbeck process, linearly autocorrelated Gaussian noise, a static nonlinear transform of that, and more general cases named without construction.

Three further things in that paper are ours only in the sense that we arrived at them independently.

Localisation is stated. Their section 2.2.4, page 82, observes that on a series which is a static nonlinear transform of a linear process, rejecting a lower null and failing to reject a higher one is what identifies where the structure sits. That is the inferential content of locating a signal along a family, written in 1992.

Localisation is exercised. Their section 4.3, with figure 10 on pages 91 and 92, reports two rungs on the same data set in two panels and reads the pair together rather than reporting one.

The multiplicity cost is declared. Their section 2.3, pages 82 and 83, warns that a battery of statistics will produce one significant result by chance, which is the cost a family of nulls incurs and which any ladder must pay.

And the problem this paper begins from is named in the same work. Their section 2.1, page 79, with its footnote 1, states that for present purposes only one experimental data set is available.

2.2 What that work does not do, and what this one adds

They do not formalise the hierarchy. The closing paragraph of their section 2.2, page 80, says the authors envision such a hierarchy; their section 2.2.5 says the more general cases would eventually be treated. **The order is described and not constructed.** No containment relation is proved between the rungs, and the nesting is not demonstrated at the level of the sets of surrogates the rungs generate, which is where a containment claim has to be made if it is to carry a theorem.

They do not report the whole ladder. Two rungs on one data set is a pair and not a family reported entire, and no criterion is offered for how far to descend or when to stop.

And there is no uninformative rung. Nothing in that work names the situation in which a statistic is constant on the reference set, distinguishes it from a failure to reject, or draws any consequence from it.

The distinction that carries the contribution is more specific than any of those. Their hierarchy nests CLASSES OF MODELS, and each class is instantiated by fitting a model to the observed data. That is exactly what an object with no generating process cannot supply: with one configuration and no process there is nothing to fit, and the hierarchy has no members. **This paper nests REFERENCE SETS by containment.** A reference set is a set of admissible configurations, and its containment in another is a fact about two sets, available whether or not any process exists. The result of section 4 is a statement about those sets, and it is what a hierarchy of fitted classes could not be without a process to instantiate it.

2.3 Three axes, and what each published work occupies

The antecedents to this work sit on three axes and not on one.

Ordering. Works that arrange nulls, models or conditionings by strength. Theiler et al. (1992), as above. Mahadevan, Krioukov, Fall and Vahdat (2006), who order network null models by the order of the degree correlations preserved and choose where to stop by measuring how much freedom remains at each. Gotelli and Ulrich (2012), on the trade-off between constraint and detection in ecological null models. Camargo (2026), on ordering conditionings for a single observed object.

Sampling. Works on how a reference set is drawn once it is specified. Besag and Clifford (1989), the generalized Monte Carlo significance test. Theiler et al. again, for surrogate generation. Stoepker and Castro (2024), on sequential computation of p-values and what a finite sample of a reference set can support.

Selection. Works that choose a reference set to maximise power against a stated alternative: Koning and Hemerik, and Koning, discussed in sections 2.6 and 2.7 and 4.7.

The third axis is not a refinement of the first two, and the difference is a presupposition. Ordering and sampling are alternative-free: a family of nulls can be arranged by strength, and a reference set can be drawn, with no alternative hypothesis anywhere. Selection cannot be. Choosing the subgroup that maximises power presupposes something to have power against. **That is why selection is a third column and not a region of the other two**, and it is why a work on that axis can be excellent and still not be an antecedent for a result that assumes no alternative.

2.4 The first boundary: ordering by containment, and what it is not

Chikina, Frieze, Mattingly and Pegden (2020) order a family of hypothesis events by containment and use that containment to license a step in their argument. The relation is stated as a set implication. It is the most formal ordering of hypotheses in the literature adjacent to this work, and the demarcation from the ladder of nulls rests on four grounds.

First, they nest events and not spaces. Their family is indexed by the threshold epsilon while the stationary distribution, the chain and the state space remain fixed. Containment of events under a single fixed null is not containment of nulls, and it is the null that a rung of the ladder changes.

Second, they do not read a profile. They fix a value of t, obtain an epsilon, and report one result. There is no sequence of decisions and consequently no shape along which a signal could be located.

Third, there is no uninformative rung. No statistic in their construction is invariant under either of their two nulls, and the notion of a conditioning under which a statistic becomes constant by theorem does not appear.

Fourth, and this is the difference neither framework can absorb into the other, their motive is validity and ours is localisation. Their nesting serves to permit a choice of epsilon after the data have been seen without paying a multiplicity correction. We order in order to locate, and we do pay multiplicity across the entire grid. The same formal relation is put to opposite uses.

2.5 The second boundary: a sampling failure is not an uninformative rung

A word of demarcation is owed at the first appearance of the term. A Markov chain used to generate a reference set can fail, and one way it fails is by being caught between two maps, so that the sampler returns a distorted reference set. The statistic in that situation still varies; what has broken is the sampling. That failure is diagnosed by inspecting the sampler and it is repaired by a better one.

An uninformative rung is the opposite situation. The statistic takes the same value on every element of the reference set, and it does so as a consequence of the conditioning imposed at that rung rather than of any accident of sampling. It is constant with data in abundance and with a sampler working exactly as intended, and no better sampler can recover a distribution from a point mass. The number reported in such a cell is arithmetic and not evidence, and the rung marks a boundary of what any test can establish about the object rather than a defect to be repaired.

2.6 Two monotonicities along a nested chain of subgroups

The nearest published relative of this work is not a result about uninformative rungs. It is a result about power, and it is asked along the same kind of chain.

Koning (2024) considers a nested sequence of subgroups and asks how a quantity behaves as one moves along it. That is the move this paper also makes, and it is published: ordering subgroups by containment and asking a monotonicity question along the order is not new here and is not claimed to be.

What is asked along the chain is where the two part.

	Koning (2024), Theorem 2	this paper, section 4
what is monotone	POWER	CONSTANCY of a statistic
under what	the p = 1 Gaussian case, a location-shift alternative, likelihood ratio tests, Neyman-Pearson	nothing. Any statistic, any configuration space, no model, no alternative, no level
what it recommends	use the ENTIRE group	where descending stops carrying information

The demarcation does not rest on our reading of their theorem. It rests on their own printed warning, which follows the theorem on the next page. They write, on page 7, two paragraphs after Theorem 2:

> However, Theorem 2 crucially relies on the Neyman-Pearson lemma, so as soon as we move beyond likelihood ratio tests there is little reason to believe that a the monotonicity property would still hold.

That sentence is the demarcation. A monotonicity that its authors expect to fail outside likelihood ratio tests and a monotonicity that follows from set containment for an arbitrary statistic are not the same result made twice; they are results of different kinds, and the difference is one their paper states about theirs rather than one this paper asserts about either.

The direction differs too, and it is not a detail. Their thesis is that a small subgroup can be dramatically more powerful than the whole group, so their interest runs toward the bottom of the chain for gain. This paper's interest runs toward the bottom for loss: the question is where descending stops buying anything, and the answer is a criterion rather than a recommendation.

2.7 The second instance: a chain and a partial order

The published statement quantifies over a nested sequence, that is over a chain. The six rungs of the case study are not a chain: four of their fifteen pairwise relations are incomparable, as section 7.2 establishes by exhibited witnesses.

This is not a refinement of the same picture. A monotonicity along a chain has a first member and a last and a single direction of descent; over a partial order there are several minimal informative rungs and no single descent, which is why the localisation of section 4 answers with a set rather than with a rung.

This demarcation carries weight in two places and is worth stating once for both. Here, it separates what is new in this paper from what is published. In section 7.4 it is what makes the partial order more than a bookkeeping preference: the chain reading is refuted there by measurement, at four ordered pairs, and a refutation is only interesting because a chain is what the published relative assumes.

2.8 Why a single object with no process is reported as a family

The natural objection to reporting a family is that one of its rungs must be the right conditioning and the paper should identify it. Three principles for identifying a right conditioning are available in the published literature, and each requires something this object does not supply.

principle	published in	what it requires
condition on a sufficient statistic	Besag and Clifford (1989)	nuisance parameters
instantiate a fitted model class	Theiler et al. (1992)	a generating process
maximise power against a stated alternative	Koning	an alternative

A single observed configuration with no generating process supplies none of the three. There are no nuisance parameters, because there is no parametric model; there is no process to fit; and there is no alternative, because the question is where structure sits rather than whether a particular structure is present.

The family is therefore reported entire not because choosing is hard, but because the three published grounds for choosing are unavailable at once. Section 4 says what such a family answers with: a set of minimal informative rungs whose size is a property of the object.

This is a statement about the principles that exist and not a proof that none can exist. A fourth principle, selecting a unique rung without nuisance parameters, a model class or an alternative, would leave the argument true of the three and no longer the whole answer.

3. The object

3.1 The sequence and its unit

The object is the King Wen sequence, the received ordering of the sixty four hexagrams of the I Ching. A hexagram is a figure of six stacked lines, each line either broken, called yin, or solid, called yang. The sequence assigns to each hexagram one of sixty four **positions**, and the assignment is the object under study: this paper analyses an ordering, not the figures it orders.

Everything below is a structure the received sequence carries. None of it is imposed by us, and where a structure is a convention of the tradition rather than a property of the arrangement it is said so.

3.2 The pairing

The sequence proceeds in thirty two **pairs**, each occupying two consecutive positions. Within a pair the two hexagrams are related by rotation through one hundred and eighty degrees, or, when the figure is unchanged by that rotation, by exchanging broken and solid lines throughout.

Eight hexagrams are unchanged by the rotation. They form four pairs whose members are related by the second rule rather than the first, and these are the **palindrome pairs**. The remaining twenty eight pairs are related by rotation. The distinction matters for this paper because several of the nulls in section 3.6 treat the two kinds differently, and one of them holds the four palindrome pairs fixed while randomising the other twenty eight; those four are the **anchor pairs** of that null.

3.3 Orientation

A pair occupies two positions, and which of its two members occupies the earlier one is the pair's **orientation**. There are thirty two pairs and therefore thirty two orientations, each a binary choice, and they are logically independent of the order in which the pairs themselves are arranged. A null may randomise the arrangement of pairs, or the orientations within them, or both, and the nulls of section 3.6 differ from one another in exactly that way.

3.4 The two canons

The received text divides into two **canons**. The first occupies positions one through thirty and the second the remaining thirty four, so the first canon holds fifteen pairs and the second holds seventeen. The division is a feature of the transmitted text and not a construction of this analysis.

The **canon boundary** is the point between position thirty and position thirty one. Because the canons hold fifteen and seventeen pairs, the boundary falls between the fifteenth and sixteenth pairs.

The number of yang lines in the first canon is used as a statistic in section 5. It is an ordinary quantity about the received text and requires nothing from this framework.

3.5 The blocks

One of the nulls partitions the sixty four positions into sixteen **blocks** of four consecutive positions each, and permutes the blocks. A block therefore contains two pairs.

Blocks and canons are transversal, and this is a fact about the received sequence rather than about the partition. The canon boundary falls between positions thirty and thirty one, which lie inside the eighth block. That block is the only one straddling the boundary and it carries one pair from each canon. Permuting blocks can move it elsewhere, which is why a statistic counting something within a canon need not be invariant under a null that preserves blocks.

3.6 The six rungs

The case study defines a family of six rungs over this object, referred to here as P0 through P5 and specified in its Appendix A. Each is given by the moves that generate its reference set: which rearrangements of the sequence are permitted while the rest of the structure is held fixed. They are not a chain and have no least and no most constrained member: section 7.2 establishes that four of their fifteen pairwise relations are incomparable. One of them permutes positions freely and another fixes the arrangement of pairs entirely, permitting only orientation changes within them, and neither of those two lies above or below several of the others.

Their containment relations are not obvious from the descriptions and are established in section 7.2 rather than asserted here. One of the six is not defined at the point of use in the published appendix but named by cross-reference to an earlier section, and that is recorded in section 7.1.

3.7 Total and partial moves

The moves generating these reference sets are of two kinds, and section 4.8 turns the distinction into a theorem.

A move is **total** when it is defined at every configuration: permuting the members of a fixed partition, reversing an orientation, exchanging two blocks. Total moves compose and invert unconditionally.

A move is **partial** when it is defined only where a local pattern occurs, so that attempting it elsewhere is not a move at all. All six rungs of this case study are generated by total moves. No rung of this ladder is generated by partial moves, which bounds what the case study can demonstrate about the scope theorem of section 4.8.

4. The formal core

4.1 Reference sets

On two words. Rung and null are demarcated in the opening paragraph of section 1, beside the demarcation of *uninformative*, and that demarcation governs here.

Let X be the space of admissible configurations of the object and let $x_0 \in X$ be the observed configuration. A *rung* is a subset R of X containing x_0, called its *reference set*, together with a way of sampling from it. A statistic is a map $T : X \to \mathbb{R}$. The test at a rung compares $T(x_0)$ against the distribution of T over R.

Nothing so far assumes a group. This is deliberate: the central result does not need one.

4.2 Uninformative rungs

A rung R is *uninformative for* T when T is constant on R.

Being uninformative is relative to a statistic and to an observed configuration. A rung uninformative for one statistic may be informative for another, and this is the ordinary case rather than the exception.

4.3 The proposition

Let R *be uninformative for* T *and let* R' *be a subset of* R *containing* x_0*. Then* R' *is uninformative for* T*.*

Proof. T is constant on R and R' is contained in R. Therefore T is constant on R'. QED.

The proof is trivial, and that is a feature rather than something to apologise for. What is not trivial, and what the rest of this work supplies, is that the order under which the proposition is applied exists at all, that it can be computed rather than checked set by set, and that the reference sets it names can be sampled exactly. A result that holds for any nested family is worth having precisely because it constrains every ladder, not only the ones whose rungs happen to be well behaved.

Remark, graded form. The same argument gives slightly more than the statement above. If $K \subseteq H$ as reference sets, then $T(K) \subseteq T(H)$, so the number of distinct values T takes is monotone non-increasing as one descends the order, and a rung is uninformative exactly when that number is 1. The proposition is the case 1 of the graded statement.

We do not present this as a contribution. It follows in one line from the same set inclusion, it assumes nothing the proposition does not, and a statement this close to trivial is more likely to be somewhere in the literature than not. A search of the works held while this was written did not find it, over six phrasings, and that search covered ten works and not the field: it is reported as not-found, not as new.

Two further cautions, because the graded form invites a measurement it cannot support. The count of distinct values is a property of the reference set, and a sample of it can only undercount; the undercount is worst at the largest rungs, which is exactly where it would reverse the comparison. And on the object of section 7, at the sample sizes used there, the image of some statistics runs to five figures and cannot be counted at all. The graded form is true by set theory and is, on this object, mostly unmeasurable.

4.4 Three consequences

The uninformative region is an order ideal. For a fixed statistic T, the family of rungs uninformative for T is closed downward under containment. It is a down set in the containment order on reference sets.

Localisation acquires a definition, and its answer is a set. The complement of that ideal is an up set, and its minimal elements are the most constrained rungs at which T still varies. That set of minimal elements is what a ladder locates. The informal claim that a signal survives one conditioning but not another becomes a statement about position in an order, and the question of where a curve flattens is replaced by the question of which rungs are minimal in the informative region.

The answer to that question is a *set* of rungs and not a single rung, and its size is a property of the object rather than a choice of the analyst. In the instantiation of section 4.9 it has two members, and they are named there. Reporting a family rather than nominating one rung is therefore the report of the answer and not an evasion of the question. A reader who asks which rung is the right one is answered: these two, and here is why there are two.

Descent past an uninformative rung is arithmetic. Along any descending path, a rung once uninformative stays uninformative under further descent, so a cell reported below such a rung repeats a value that was already fixed. The ladder can be truncated at that point along that path without loss of information. This is what separates a reported ladder from a programme: a family of nulls with no criterion for when a rung ceases to carry information can be enumerated but cannot be terminated.

This is a criterion and not the boundary. Non-rejection has more than one cause, and only this one is settled by the proposition. Section 4.5 separates them, and the separation matters because truncating a path for the wrong reason would discard a rung that still had something to say.

4.5 Four causes of non-rejection

A rung can fail to reject for reasons of different kinds, and the differences matter more than the failure. Four are distinguishable here, and the list is not claimed to be complete.

The rung is uninformative. T is constant on R. The observed value is equalled by every element of the reference set, so the p-value is exactly 1. This is deterministic and it follows from the proposition of section 4.3: it depends on the object and on the statistic, and on no level, no sample size and no procedure.

The reference set is too small to reach the level. If $|R| < 1/\alpha$ then the smallest attainable p-value exceeds α and no configuration of the data permits rejection at that rung. This is deterministic by arithmetic rather than by any property of T, it depends on the rung and on the level, and **it is computable before any data are seen.**

The signal is exhausted. T varies on R, the reference set is large enough to resolve α, and this object does not cross. This one is inferential, and it is the only one of the three that a different object could answer differently.

The first two are down sets and they are incomparable with each other. A statistic constant on a set is constant on every subset, and a set with fewer than $1/\alpha$ elements has subsets with fewer still, so each is closed downward. Neither implies the other: a large orbit on which T is constant is uninformative and not too small, and a small orbit on which T varies is too small and not uninformative. Both exist.

Both are contained in non-rejection, and exhausted signal is non-rejection with neither of them present.

Only the first is what the proposition of section 4.3 is about. The down-set result concerns the constancy of a statistic on a reference set. The cardinality bound is also a down set and the proposition says nothing about it, because it is a fact about the size of the set and not about the statistic. A framework claiming both would be claiming a theorem about arithmetic that it has not proved and does not need.

The observed value is too common. T varies on R, the reference set is large enough to resolve α, and yet the set of elements of R at which T equals its observed value has measure greater than α, for reasons belonging to the structure of T rather than to the data. The smallest p-value attainable is that measure, so no configuration permits rejection. This is deterministic, and like the second cause it is computable before any data are seen; unlike the second it depends on T.

How the fourth differs from each of the other three, one at a time. From the first: there T is constant, the whole of R is one level set, the measure is 1 and the p-value is exactly 1; here T varies and the measure is strictly between α and 1. From the second: that one is a fact about the SIZE of R and holds whatever T is, while this one holds at reference sets of any size, including arbitrarily large ones. From the third: that one is inferential and a different object could answer it differently, while this one is settled by the definition of the statistic and no object crosses it.

And unlike the first two it is not closed downward. Passing to a smaller reference set can raise or lower the measure of the observed value, so a rung suffering this cause may sit above rungs that do not. Nothing in this paper depends on it being an order ideal, and it is not claimed to be one.

The list is open and is stated as open. Four causes are named and no total is asserted. What is claimed is that there are at least four, that three of them are structural and computable without data, that the first two are order ideals and the fourth is not, and that they are not ordered by one another.

4.5.1 What separates the two kinds, and it is not a list of four accidents

The four fall into two kinds, and the division is exact.

The first two are conditions on the reference set. That T is constant on R is a statement about the image of T over R; that $|R| < 1/\alpha$ is a statement about the size of R. **Both are preserved by passing to a smaller reference set, and for the same reason in each case: shrinking R can only collapse the image of T further and can only make R smaller.** Each is therefore an order ideal.

The last two are conditions on where the observed value falls in a distribution. The third asks whether $T(x_0)$ is extreme among the values T takes on R; the fourth asks how much of R shares that value. **Neither is preserved by shrinking R, and again for one reason: shrinking R changes the distribution against which $T(x_0)$ is compared, and it can move $T(x_0)$ towards the middle of that distribution or towards its edge.**

So the boundary between what descends and what does not is not a property of the four causes taken one at a time. It is the difference between a condition on a SET and a condition on a POSITION IN A DISTRIBUTION, and the taxonomy is what locates it.

This is why the scope of the proposition is a result and not an apology. The proposition proves the down-set property for the first cause; the second has it by arithmetic; the third and fourth do not have it at all. **The theorem does not fail to cover the third and fourth because it is weak. It cannot cover them, because they are not the kind of statement that descends**, and knowing which kind each cause belongs to was the work.

4.5.2 A region that contains an order ideal without being one

Write U for the family of rungs at which the set where T equals its observed value has measure greater than α, with no further condition. This is the fourth cause with its requirement that T vary removed.

U **contains the first cause's ideal.** If T is constant on R then every element of R carries the observed value, the measure is 1, and $1 > \alpha$ for any level worth using.

U **is not itself an order ideal.** Let T take the observed value on three fifths of H and let $K \subseteq H$ be a rung on which it takes that value at one element in thirty. Then H lies in U and K does not, at $\alpha = 0.05$.

So U contains an order ideal and is not one. The containment is not a near miss and it does not make U almost closed downward: descending from a rung in U can leave U at the first step, and the ideal inside it is reached only along the paths on which T becomes constant.

A caution on the reading. The fourth cause as stated above requires T to vary, so the fourth cause and the first are disjoint by construction and neither contains the other. The containment is a fact about U, the unqualified region, and it is stated here because U is what a reader computing measures will actually meet: nothing in a measure tells you whether the statistic varied.

Why this section exists. A single stopping criterion invites the question "which boundary is this rung at", and that question presupposes a unique answer. A rung can sit at more than one of these causes at once, and a rung can sit at none of the ones named. **The four are kept apart because collapsing the first into the third would report a theorem where an inference belongs, collapsing the second into either would report as a property of the statistic what is a property of the size of the set, and collapsing the fourth into the third would report as a fact about the object what is a fact about the arithmetic of the statistic.**

An instance of the fourth, in this paper. The control statistic of section 5 takes its observed value on exactly one sixteenth of any reference set that randomises the four orientations it counts, by independence, so its smallest attainable p-value is 0.0625. At $\alpha = 0.05$ it could not reject at any rung of section 7, whatever the data. The statistic varies, the reference sets are enormous, and the cause is neither of the first two.

4.6 What the order costs

The proposition is free. The order is not.

To apply it, one must know that one reference set is contained in another. For arbitrary conditionings this is a question about two subsets of a space whose size is the whole point of the exercise, and answering it by inspection is not available. One must also be able to sample each reference set, and for an arbitrary subset of X no sampler is given by the specification of the subset.

There is a third cost, and unlike the other two it is a number rather than a difficulty. Testing at level α requires $|R| \geqslant 1/\alpha$, since a reference set with fewer elements cannot produce a p-value at or below α whatever the data. This is the second cause of section 4.5 read as a design constraint rather than as a diagnosis, and it is checkable on any proposed rung before any data are seen: it needs the size of the reference set and nothing else.

That bound is the loose form of a sharper one, and the difference is worth a line because only the loose form is usually stated. $|R| \geqslant 1/\alpha$ is what is needed when the observed value is attained exactly once in the reference set; in general the smallest attainable p-value is the measure of the set on which T equals its observed value, and the requirement is that this measure be at most α. The size bound is the case of multiplicity one. A ladder can therefore satisfy the size bound at every rung and still carry a statistic that cannot reject anywhere, which is the fourth cause of section 4.5 and is instantiated by the control of section 5.

We report it for the case study because it is cheap and because a ladder that never states it is asking a reader to assume it. At $\alpha = 0.05$ the floor is twenty elements. The six rungs of section 4.9 have reference sets between 4.30×10^9 and 1.27×10^{89}, so the smallest exceeds the floor by a factor of about 2×10^8. The sizes are totally ordered and the rungs are not, so the smallest reference set does not belong to a rung below all the others. No rung of this ladder is anywhere near the bound, and none is at risk of the second cause.

What that check does not establish. It concerns the size of the reference sets and says nothing about how many draws were taken from them, which is a separate quantity and a separate risk. The reference sets are large enough to carry the reported resolution; whether enough was drawn from them is a question about the sample size and is treated where the sample size is reported.

4.7 Subgroup rungs, a corollary and not the foundation

Nothing in sections 4.1 to 4.6 assumes a group, and this section adds one. **What follows is a corollary of the proposition of section 4.3 in the case where the rungs happen to be orbits, not a second foundation and not the general result restated.** The distinction is worth a sentence because the group case is the one every worked example here belongs to, and a reader who met it first would reasonably take it for the theory.

Let G be a group acting on X. A rung is a *subgroup rung* when its reference set is the orbit Hx_0 of some subgroup $H \leqslant G$. For such rungs, three things follow at once, **and none of the three is a contribution of this work.** Two are standard and the third is somebody else's published theorem, named where it is used.

Containment becomes computable. If $H \leqslant K$ then $Hx_0 \subseteq Kx_0$, and containment of subgroups is decided on generators. An order that would otherwise require comparing sets is decided by comparing finite presentations.

Sampling becomes exact by construction, and this is not ours. Uniform sampling from an orbit requires no argument about reachability: every element of the orbit is gx_0 for some $g \in H$, and drawing g uniformly from H draws uniformly from the orbit up to the stabiliser, which is accounted for in the obvious way. The exact test a subgroup buys is the published result of Koning and Hemerik, who give it with $|S|$ transformations and make it the subject of their paper. We claim no part of it. It is cited here because a published theorem stands where this passage would otherwise carry an unattributed assertion, and the other two bullets of this section are standard group and lattice theory rather than anyone's contribution.

On the version cited. Koning and Hemerik is cited as arXiv:2202.00967v3, which is the version read. No printed version is held and none is claimed; if a journal version is cited later, its numbering is declared at the point of citation.

Meets are always available. The subgroups of G form a lattice, so any two rungs have a greatest lower bound. One can always name a rung lying below two given rungs and compute it, which is what allows a branching family to be treated as a single object rather than as a set of unrelated chains.

Remark, stated so that it is not discovered later. The map from subgroups to reference sets is order preserving but is not in general a lattice homomorphism: the orbit of an intersection can be a proper subset of the intersection of the orbits. The lattice structure is a property of the subgroups and is inherited by the reference sets only as an order, not as a lattice. Nothing above depends on more than the order. **In the instantiation of section 4.9 the action is free, and there the two coincide.**

4.8 Scope: which rungs are subgroup rungs. A second theorem, and it is second

The general hypothesis that fixing an invariant always amounts to choosing a subgroup is false, and the boundary can be stated exactly.

This result is proved here and it is not the spine of the paper, for a reason that is a limitation of the evidence and not a preference. Its whole content is a dividing line between rungs reachable by total moves and rungs reachable only by partial moves. **The case study of section 4.9 exercises one side of that line only:** all six of its rungs are generated by total moves and not one is a partial-move rung. A result whose content is a boundary cannot carry a paper that demonstrates one side of it and cites the other. It is a theorem, it is proved, and it is second.

Criterion. Let $R \ni x_0$ and let $\mathrm{Stab}(R) = \{g \in G : gR = R\}$ be the setwise stabiliser, which is a subgroup for any R. Then R is the orbit of a subgroup of G if and only if $\mathrm{Stab}(R)$ acts transitively on R, and in that case $\mathrm{Stab}(R)$ is the largest subgroup whose orbit is R.

Proof. If $Hx_0 = R$ for a subgroup H, then $HR = (HH)x_0 = Hx_0 = R$, so $H \leqslant \mathrm{Stab}(R)$, whence $R = Hx_0 \subseteq \mathrm{Stab}(R)x_0 \subseteq R$. The converse is immediate. QED.

The criterion is exact but expensive. The following distinction covers the cases that arise in practice and can be read off a specification.

Total and partial moves. A rung may be specified by the moves that generate its reference set. A move is *total* when it is defined at every configuration: permuting the elements of a fixed partition, reversing an orientation, exchanging two blocks. Total moves compose and invert unconditionally, so the moves of a rung generate a group and the reference set is an orbit. A move is *partial* when it is defined only where a local pattern occurs. The canonical example is the two by two swap used to sample binary matrices with fixed row and column margins, which requires ones in opposite corners of a rectangle and is undefined where they are absent. Partial moves compose only sometimes. They generate a groupoid rather than a group, and the reference set they reach is not an orbit.

The consequence for the literature. Where a group acts, reachability of the reference set is a construction and there is nothing to prove. Where only partial moves are available, reachability is a theorem, and it must be proved for each family. That theorem is exactly what Besag and Clifford (1989) supplied for the fixed margin case, showing that any admissible table is reachable from any other by a succession of swaps, and it is the reason their machinery is required precisely where the subgroup description fails. The null models that dominate the ecological literature on this problem are of the partial kind, and the ladder framework applies to them through the containment order alone, without the computational advantages of section 4.7.

The scope of this work is therefore not the scope of its central result. The proposition of section 4.3 holds for any nested family. What section 4.7 supplies is a class of rungs for which the order is computable, the sampling is exact and the family has a lattice structure. Where that class does not apply, the proposition still does, and the cost is that containment must be established directly and the sampler must be proved to connect.

4.9 Instantiation in the case study

The six rungs of Appendix A of the case study are orbits of subgroups of the symmetric group on the sixty four positions. Writing them in the order in which they are printed: the full symmetric group; the group generated by permutations of the thirty two pairs together with the thirty two orientation reversals; the stabiliser within that group of the partition into the two received canons; the subgroup fixing the four palindrome pairs in position and in orientation; the subgroup carrying blocks of four consecutive positions to blocks; and the group of the thirty two orientation reversals alone.

Every one of these is generated by total moves, so each is a subgroup rung and section 4.7 applies to all six.

The family is a partial order and not a chain. The fifteen pairwise containments were computed from predicates derived from the printed definitions. Eleven are strict containments, four pairs are incomparable, and no two rungs coincide. Each refutation is accompanied by an exhibited witness, that is a permutation lying in one rung and not in the other, checked in both directions. The order has **two minimal elements**, the anchor rung and the orientation rung, and the sharpest of the four incomparabilities is between them: the anchor rung fixes four pairs in orientation while the orientation rung reverses all thirty two, so neither contains the other.

The canon division and the block partition are transversal, and this is a property of the object rather than of the formalisation: fifteen and seventeen pairs place the boundary between the two canons inside the eighth block of four positions.

The printed order is a valid linear extension of the partial order, since each of the eleven strict containments prints the larger group before the smaller. No figure and no conclusion of the case study is affected. What the present framework corrects is the description of the family as one of increasing structure, which describes a chain: a later rung concedes a different part of the architecture, not a larger part of it.

On the ambiguity at the most constrained rung. The printed prose of the case study admits two readings of one rung's specification, and the two differ by a factor that is not an accident of the example: at n pairs the index of one group in the other is $n!/((n/2)!\,2^{n/2}) = (n-1)!!$, the number of ways to partition n labelled objects into $n/2$ unordered blocks of two. That count is exactly what the two readings disagree about, since one leaves the block partition free and the other holds it fixed. At $n = 32$ it is 191898783962510625. The identity is stated here because it converts a discrepancy observed at five scales into a fact about what the ambiguity is, and a reader checking the rung will otherwise rediscover the factor without its explanation.

4.10 Retrodiction, and what it does not establish

The proposition predicts the shape of the published table before the table is consulted. Seven cells of that table are reported as constant. **Four of those seven lie above rungs contained in them and generate, between them, seven predicted constancies; the remaining three sit at minimal rungs and predict nothing.** All seven predictions hold and none is refuted.

What this does not show. If the family were the chain that the case study's wording suggests, the proposition would predict four further constancies, and those four also hold. No cell of the table distinguishes the chain reading from the partial order reading. The retrodiction is therefore evidence that being uninformative is inherited downward, and it is silent on whether the order is a chain or a lattice. The lattice is established by the fifteen containments and their witnesses, and independently by a preregistered experiment whose predictions were fixed in writing before any of its statistics was computed. That experiment tests the four pairs that the chain reading orders and the partial order leaves incomparable, and its strongest instance uses a statistic nobody would have built for the purpose: the total yang in the first received canon.

The two establishments are independent and neither strengthens the other. The containments are set inclusions with exhibited witnesses; the experiment is a measurement whose predictions were fixed before it ran. **The limitation this section states about Table A1 is unaffected: no cell of the published table distinguishes the chain reading from the partial order, and that is why a second route had to be built rather than found.**

This limitation is stated here rather than in a note because a reader who discovers it should find it already written. It is not removed by editing. Removing it would require an object, or a statistic on this object, for which the two readings disagree in some cell.

4.11 Status

Established. The proposition, in the general form of section 4.3. The separation of the four causes of non-rejection in section 4.5, of which the first is what the proposition covers, the second and the fourth are arithmetic, and only the third is inferential. The criterion of section 4.8, proved and second. The six identifications and the fifteen containments of section 4.9, verified against the printed definitions and, for the block rung, against the replication code. The seven retrodictions of section 4.10, and, independently of them, the preregistered experiment named there.

Established and not ours. The exact test a subgroup buys, section 4.7, which is the published result of Koning and Hemerik. The two other consequences of that section are standard group and lattice theory. The graded remark of section 4.3 is stated and declined: it is one line from the proposition and is reported as not-found in the works held rather than as new.

Measured, and reported because it is cheap. No rung of section 4.9 is near the cardinality bound of section 4.6; the smallest reference set exceeds it by about eight orders of magnitude. That concerns the size of the reference sets and not the number of draws taken from them.

Established, and it is the scope statement in positive form. The down-set property holds for exactly the first two causes of section 4.5 and fails for the other two, and the division is the division between a condition on a reference set and a condition on where the observed value falls in a distribution. The proposition of section 4.3 proves the first; the second is arithmetic. **The theorem's scope is therefore located by the taxonomy rather than confessed by it.**

Also established. The unqualified region of section 4.5, the rungs where the observed value has measure above the level, contains the first cause's order ideal and is not itself an order ideal.

Not established. That the four causes of section 4.5 exhaust non-rejection: the list is open and no total is asserted. That the fourth cause is closed downward; it is not, and nothing here depends on it being so. That the graded remark is new. That the partial-move side of the boundary in section 4.8 is exercised by anything in this work; it is cited and not demonstrated.

A defect in the printed prose of the case study. The description of the block rung admits a literal reading under which a theorem printed two paragraphs below it would be false. **The code and the theorem agree with each other against that sentence, and both are in the published replication package, so the discrepancy is checkable today from the published**

artifacts alone. The two readings differ by the index computed in section 4.9. How the published record is to be corrected is not settled here.

5. The preregistered experiment

5.1 What it was for

The six rungs form a partial order and not a chain, and the fifteen containments of section 7.2 establish that by exhibited witnesses. The published table of section 7.3 does not: every constancy in it is consistent with both readings, because the chain reading predicts a superset of the partial order's constancies and all of them hold. A reader entitled to ask whether the partial order is required by any measured quantity would, on that evidence alone, be entitled to say no.

This experiment exists to remove that gap. It asks for a statistic on this object for which the two readings disagree in some cell, and it fixes what would count as an answer before looking.

5.2 The ordering is proved by the record and is not asserted

The predictions were committed in one act, alone, with nothing computed: that commit contains the document and no statistic, no percentile and no result. The instrument that evaluates them was written and run in the next commit. **The order in which those two things happened is a property of the deposited commit graph and can be checked by anyone**, which is why this section makes no appeal to the good faith of the authors and why the deposit of section 10 is of the repository and its graph rather than of a snapshot of its files.

The document is not edited. It was not edited when the results came in, and it is not edited now: a preregistration that improves after the fact is not a preregistration, and the rule holds even in the cases where an edit would plainly make it better. Two such cases arose and both are recorded below rather than corrected in the document.

5.3 Two tiers, and the first is a control

The design has two tiers and only the second carries evidence.

Tier one, statistic C1: A POSITIVE CONTROL, and it is labelled as one wherever it appears. C1 counts how many of the four palindrome pairs keep their received orientation. It was constructed for this experiment and declared so in advance. Its role is to fail loudly if the instrument is wrong: a demonstration whose guaranteed case does not come out has not been run correctly, and the preregistration commits to believing nothing else in the document if C1 misbehaves.

A property of C1 that was computable before any draw and was not computed. The observed value is four out of four. At any rung that randomises those four orientations each flips independently with probability one half, so the observed value has mass exactly 1/16, that is 0.0625. **The smallest p-value C1 can attain is therefore 0.0625, and it exceeds 0.05. C1 could not have rejected at the five percent level at any rung of this ladder, and that follows from the structure of the statistic with no data at all.**

This does not damage C1 in the role it was given, because a positive control is asked whether it varies where it should and is constant where it should, not whether it rejects. **It does mean that C1 was, by construction, incapable of the thing a reader may assume a statistic in a table is doing**, and the fact belongs beside it rather than in a footnote.

It is also an instance of the fourth cause named as open in section 4.5. That section lists a statistic which varies on the reference set but attains its observed value on a fraction bounded below by α for structural reasons, and declines to analyse it. The control of this experiment is exactly such a statistic. The list of causes was declared open, and one of the things it was open about was sitting inside this paper's own design.

Tier two, statistic C2: the statistic that was not constructed for this. C2 is the total number of yang lines in the first received canon. It is a quantity someone would compute without any of this framework in view, and the balance of yang between the two received canons is an ordinary question about the received text. Its predictions are given rung by rung in the preregistration, each with its reason, and the interesting case is stated there in advance: the canon boundary falls inside a block, so permuting blocks can move a straddling block elsewhere and change canon membership.

5.4 What came back

The results are in sections 7.4 and 7.5 and are not repeated here. In summary: every preregistered prediction held, none failed, the control behaved, and the chain reading is refuted at four ordered pairs of which two are evidential and two are the control.

Two evidential refutations rather than four, because the two carried by the control are demonstrations that the instrument works and are not evidence about the object. Section 7.5 counts them separately for that reason.

5.5 Two things the preregistration did not do, recorded and not repaired in it

First, the floor above. The document does not state that C1 cannot reject at five percent, and it should have: the property is derivable from the definition of the statistic in one line and would have been available on the day it was written. Nobody was misled, because C1 is nowhere offered as evidence, but the omission is a design defect of the document and is attributed rather than diffused. **The preregistration was drafted by the assistant working on this review, at the authors' direction, and this omission is that drafting's.**

Second, an unpredicted cell. The document predicts C1 at P3 and at P5 and makes no statement about C1 at P4. That cell exists because the design commits to reporting both statistics at all six rungs, and a fourth refutation of the chain reading falls out of it. It is real and it is not preregistered, and section 7.5 keeps it separate for that reason.

Neither is fixed by editing the document, and the reason is the same in both cases. The value of a preregistration is that it says what was thought before, and a document that improves whenever its authors learn something is a record of nothing. Both defects are reported here, in the paper, where a reader meets them at the same time as the design.

5.6 What the experiment cannot settle

It does not show that the published table discriminates the two readings; it does not, and this is a different measurement on the same object with statistics that did not exist when that table was computed. It does not show that the partial order is the only correct reading: a coarser or finer third reading is untouched, as the preregistration says in its own final section. And it says nothing about the sequence itself. C2 was chosen for its behaviour under the nulls and not for what it might say about the object, and no percentile in section 7.4 is extreme or is offered as a finding.

6. Methods

6.1 Sampling the reference sets

Each of the six rungs of section 3.6 is sampled by drawing uniformly from the group that generates its reference set and applying the result to the observed configuration. Because every one of the six is generated by total moves, the moves compose and invert without condition, the reference set is an orbit, and a uniform draw from the group gives a uniform draw from the orbit up to the stabiliser.

This is exact sampling by construction and not a Markov chain. There is no burn-in, no mixing argument, and no convergence to check, and none of the diagnostics that a chain would require applies here. The property is the published result cited in section 4.7 of the formal core and it is not a contribution of this work; what belongs to this work is only the observation that all six rungs of this ladder qualify for it.

Each sampler was checked against the published replication code of the case study rather than against the description in prose: draws from each published sampler were tested against a predicate derived from that rung's printed definition, and the outcome is reported in section 7.1.

6.2 Draws and the seed

Twenty thousand draws are taken per cell, at every rung and for every statistic, and the same count is used throughout so that no cell has a resolution another lacks. The pseudo-random seed is fixed at 20260722 and is stated here so that any figure in section 7 can be reproduced exactly rather than approximately.

The design commits to reporting every statistic at every rung, including rungs about which nothing was predicted. That commitment is why section 7.4 contains cells with no prediction beside them, and it is also why one refutation of the chain reading falls out of a cell the preregistration did not anticipate, as section 5.5 records.

6.3 How percentiles are printed

A percentile or a p-value estimated from a finite sample is printed at the resolution the sample supports and no finer. The rule follows Stoepker and Castro (2024), whose section 1.4 separates two information goals that are usually conflated: estimating the accept-or-reject decision, and estimating the p-value itself. Sample-size guidance is mostly written for the first while analysts report the second, and the rule below is what that distinction demands of a reported figure. A printed last digit claims a resolution of half a unit in that digit, and it is printed only when the standard error of the estimate is at most that half unit. Where the resolution is not supported, fewer digits are printed or an interval is given. The rule and the sample size appear where the figures do rather than in an appendix, because a reader meeting a figure is the one who needs them.

A cell in which the statistic does not vary is not printed as a percentile of zero. It is printed as constant, with the argument for its constancy attached. A percentile of zero beside a constant cell reads as an extreme measurement, and the two are not the same kind of statement: one is arithmetic that no sampling could change, the other an estimate.

6.4 Constancy is argued and variation is exhibited

A finite sample can refute constancy and cannot establish it. Twenty thousand draws that happen to agree are consistent with a statistic that varies rarely.

Accordingly, no cell is reported as constant on the strength of the draws. A cell is reported as varying when a drawn element with a different value is exhibited; it is reported as constant when the sample shows no variation and a structural argument for the constancy is available and is printed with it. A third category is reserved for a cell with no variation and no argument, so that such a cell can never be reported as a constancy result. Section 7.4 records that no cell fell into it.

6.5 Reference set sizes

The size of each reference set is computed as the order of the group generating it. That step is licensed by the action being free on the observed configuration, which makes the orbit map a bijection, and it would not be available otherwise. The sizes are reported in section 7.6 against the bound of section 4.6.

7. Results

Every figure in this section was computed by an instrument in the deposited record, and each is traceable to the run that produced it. Percentiles are printed at the resolution the sample size supports: a printed last digit claims a resolution of half a unit, and it is printed only when the standard error of the estimate is at most that half unit. The sample size is twenty thousand draws per cell throughout.

7.1 The six rungs and their identification

The ladder has six rungs. Each is specified in the case study by the moves that generate its reference set, and each of the six was identified with a subgroup by deriving a predicate from the printed definition and checking that the published sampler satisfies it. Five of the six definitions are given directly; the sixth, the pair-preserving null, is named by cross-reference rather than defined at the point of use, and that is a property of the printed text rather than of the identification.

The cross-check ran five thousand draws from each rung's published sampler against that rung's predicate. There were no mismatches at any rung.

7.2 The containment order

There are fifteen unordered pairs among six rungs. Each was decided by exhibiting either a containment or a witness lying in one reference set and not the other, and every witness was checked in both directions.

P0 / P1	P0 strictly contains P1
P0 / P2	P0 strictly contains P2
P0 / P3	P0 strictly contains P3
P0 / P4	P0 strictly contains P4
P0 / P5	P0 strictly contains P5
P1 / P2	P1 strictly contains P2
P1 / P3	P1 strictly contains P3

P1 / P4	P1 strictly contains P4
P1 / P5	P1 strictly contains P5
P2 / P3	INCOMPARABLE
P2 / P4	INCOMPARABLE
P2 / P5	P2 strictly contains P5
P3 / P4	INCOMPARABLE
P3 / P5	INCOMPARABLE
P4 / P5	P4 strictly contains P5

strict containments	11
incomparable pairs	4
equalities	0

Four of the fifteen pairs are incomparable, so the six rungs form a partial order and not a chain. The sharpest of the four is P3 against P5: the most constrained rung of the printed sequence is not contained in P3, because P3 holds the four palindrome pairs fixed in orientation while P5 randomises all thirty two orientations, so neither contains the other.

Two of the four incomparabilities are facts about the object rather than about the formalism. P4 is not inside P2 because the boundary between the two received canons falls inside a block: the canons are fifteen and seventeen pairs, so the boundary sits between pair slots fourteen and fifteen, and those two slots lie in the same block of four consecutive positions. The received division of the text and the block partition are transversal. P3 is not inside P2 because P3 permits any of its twenty eight free pairs to cross the canon boundary, which P2 forbids.

The printed order of the six is nevertheless a valid linear extension of the partial order: all eleven strict containments print the larger group before the smaller. No figure of the case study is affected. What is imprecise is the description of the sequence as one of increasing structure, which describes a chain; a later rung concedes a different part of the architecture and not a larger part of it.

7.3 The retrodiction over the published table

The published table reports twenty four cells. It was computed and published before the order-theoretic property was formulated, so it can be used as a retrodiction: the property requires that a statistic constant at a rung be constant at every rung below it, and a single cell carrying a percentile where the property demands a constant would refute it.

constant cells in the published table	7 of 24
of those, cells lying above a rung contained in them	4
cells at a minimal rung, which predict nothing	3
constancies the four generate between them, under the containment relation recomputed for this check	7
predictions that hold	7
predictions refuted	0

non-constant entries, recomputed	17
non-constant entries the published appendix asserts	17

The retrodiction holds and no cell contradicts the property. Four of the seven constant cells lie above a rung contained in them and generate the seven predictions between them; the other three sit at rungs with nothing below them, so they are consistent with the property without testing it. The retrodiction is evidence that the property is inherited downward, and it is silent on whether the order is a chain or a lattice: no cell of this table distinguishes the two readings, which is why the experiment of section 5 had to be built rather than found.

7.4 The preregistered experiment

Two statistics were fixed in writing, with their predictions, before either was computed. Percentiles are two-sided positions of the observed value in the sampled reference distribution.

STAT C1, observed value 4

P0	varies	94	[0, 4]	no prediction made	
P1	varies	94	[0, 4]	no prediction made	
P2	varies	94	[0, 4]	no prediction made	
P3	constant, by argument			predicted constant	agrees
P4	varies	94	[0, 4]	no prediction made	
P5	varies	94	[0, 4]	predicted varies	agrees

STAT C2, observed value 86

P0	varies	18	[70, 108]	predicted varies	agrees
P1	varies	20	[68, 114]	predicted varies	agrees
P2	constant, by argument			predicted constant	agrees
P3	varies	21	[68, 112]	predicted varies	agrees
P4	varies	15	[74, 106]	predicted varies	agrees
P5	constant, by argument			predicted constant	agrees

On the STAT C1 row. **93.75 is exact, and the five varying cells are one arithmetic fact measured five times.** STAT C1 counts how many of the four palindrome pairs keep their received orientation. Each flips independently with probability one half at every rung that randomises them, so the observed value of four has mass exactly one sixteenth and the percentile is exactly 93.75. The five figures printed above lie within 2.04 standard errors of that number. A reader meeting 94.0 at one rung and 93.5 at another would see two measurements of two rungs, and there is one number there.

On the printed resolution. Not one of the nine varying cells supports a decimal. The smallest standard error among them is 0.167 points and a decimal would claim 0.05, so every percentile above is printed to the unit. This is the same limitation the case study's own published table carries, applied here to our figures.

On the three constant cells. Each rests on a proof and not on the draws, and each is printed as constant with that fact attached rather than as a percentile of zero, which a reader would take for an extreme measurement. At P3 the four palindrome pairs are fixed in position and orientation, so STAT C1 takes the value four on every element. At P2 the pairs permute only within their canon, so the multiset over the first thirty positions is invariant and any sum over it is invariant. At P5 the pair order is fixed and a flip exchanges two positions inside one pair, so nothing crosses the boundary at position thirty.

A run of this kind can refute constancy and cannot establish it. Every cell reported as varying carries an exhibited draw whose value differs from the observed; every cell reported as constant carries a structural argument. A third category was reserved in advance for a cell with no variation and no argument, and no cell fell into it.

7.5 What the experiment refutes

The chain reading requires that constancy propagate from the first member of each of four ordered pairs to the second. In all four cases it does not.

EVIDENTIAL, on the load-bearing statistic

(P2, P3)	STAT C2 constant at P2 and varying at P3	predicted
(P2, P4)	STAT C2 constant at P2 and varying at P4	predicted

DEMONSTRATIONS OF THE CONTROL, counted separately

(P3, P5)	STAT C1 constant at P3 and varying at P5	predicted
(P3, P4)	STAT C1 constant at P3 and varying at P4	NOT PREDICTED

The two evidential refutations are the two carried by STAT C2, which is the total yang in the first received canon: a quantity someone would compute without any of this framework in view. The refutation therefore does not rest on a statistic built to produce it.

The other two are demonstrations that the control behaves, and they are not counted as evidence. STAT C1 was declared in advance to be a constructed control, included because the preregistration required an instrument check before anything else was to be believed. **And one of the two was not predicted.** The preregistration made no statement about STAT C1 at P4; that cell exists because the design commits to reporting both statistics at all six rungs, and the refutation falls out of it. It is a real refutation and it is not a preregistered one, and merging it with the other three would claim more foresight than the document had.

7.6 Reference set sizes against the level bound

Testing at level α requires a reference set of at least $1/\alpha$ elements. At $\alpha = 0.05$ the bound is twenty.

rung	reference set size
P0	1.2689e+89
P1	1.1301e+45

rung	reference set size
P2	1.9977e+36
P3	8.1843e+37
P4	5.8892e+27
P5	4.2950e+09

The smallest exceeds the bound by a factor of about 2×10^8 and exceeds the twenty thousand draws by a factor of about 2×10^5. No rung of this ladder is near the bound. These are sizes of reference sets and not counts of draws: the reference sets are large enough to carry the reported resolution, and whether enough was drawn from them is the separate question treated above.

8. Discussion and limitations

8.1 What the criterion buys

The result is small and it is not weak. A statistic constant on a reference set is constant on every reference set inside it, so a family of nulls has a place where descending stops buying anything, and that place can be named without a model, an alternative or a level. What a reported ladder gains is a reason to stop, and what a reader of one gains is the ability to tell a cell that is arithmetic from a cell that is evidence.

The case study shows the order is not read off the presentation. Six nulls published as a sequence of increasing structure are a partial order with four incomparable pairs, and the difference has an observable consequence: a preregistered experiment refutes the chain reading at four ordered pairs, two of them on a statistic that was not built for the purpose.

8.2 On the terminology, and the risk it carries

We use *uninformative* for a rung on which a statistic is constant. A term search over the index reachable by a live web search, run on the phrasings this paper uses, returned no settled sense of the word attached to a null or to a test.

That search bounds the risk and does not remove it. **The phrase may be in use somewhere the index did not reach, or in use with a meaning a term query cannot see.** The search was a live query against a search index at an instant, it was not a reading of any literature, and a negative result over an index is sound only over what the index covered. A reader who knows the term in another sense should read it here as defined in section 4 and nowhere else.

The one collision found is stated in the abstract and again here: *uninformative prior* is a settled term of art and it describes a prior distribution. This paper describes a property of a reference set relative to a statistic. The words coincide and the objects do not.

8.3 Two orderings of the case study rest on works we did not obtain

The case study compares five orderings. This paper uses one of them, the received sequence, whose provenance the case study states without a source because it is the received text.

Its other two historical orderings are reconstructions that rest on philological works. **We did not obtain any of those works.** Nothing in this paper depends on those two orderings, no figure reported here is computed from them, and no claim here would change if their provenance were different.

We state it because a reader relying on section 7 is relying on the case study, and the provenance of the case study's data is part of what they are relying on. A dependency we do not use is still a dependency a reader inherits, and the honest report is that we did not check it rather than that it did not arise.

8.4 The fourth cause, and what does not descend

Section 4.5 names four reasons a rung can fail to reject and section 4.6 shows that the two which are conditions on the reference set descend under containment while the two which are conditions on where the observed value falls do not.

The fourth of them deserves a limitation of its own, because it is the one a reader is most likely to meet without recognising. A statistic can vary on the reference set, the reference set can be far above the level bound, and the observed value can still be shared by so large a fraction of the set that no configuration permits rejection. This is deterministic and computable before any data are seen, and this paper contains an instance: the control statistic of section 5 could not have rejected at any rung, for a reason available from its definition.

That cause is not closed downward. Descending can raise or lower the fraction carrying the observed value, so a rung suffering it may sit above rungs that do not, and no analogue of the stopping criterion is available for it. Nothing here depends on it being an order ideal and it is not claimed to be one.

And a caution on how it is met in practice. The unqualified region, the rungs where the observed value has measure above the level, contains the uninformative rungs and is not itself an order ideal. That region is what a reader computing measures will actually meet, because **nothing in a measure tells you whether the statistic varied.** The two must be separated by looking at the statistic and not at the number.

8.5 The list of causes is open

Four causes are named and no total is asserted. We do not claim they exhaust non-rejection, and we have not tried to prove that they do. What is claimed is that there are at least four, that three of them are structural and computable without data, that the first two are order ideals and the fourth is not, and that they are not ordered by one another.

A fifth would not disturb any result here. It would extend the accounting, which is what the accounting is for.

8.6 What the case study cannot show

All six of its nulls are generated by moves defined at every configuration. None is generated by a move defined only where a local pattern occurs. The scope theorem of section 4.8 is proved for both sides of that line and demonstrated on one, and a family built on partial moves would exercise a part of the result that no measurement in this paper touches.

Nothing here is a claim about the King Wen sequence.

9. References

Each entry states the version that was read. Where the version read is not the version of record, both are given and the reading is attributed to the one that was read. Works that were not obtained are marked as such and are not cited as though they had been consulted.

Each entry also rests on something, and the entries do not all rest on the same thing. Two facts decide what an entry rests on: whether the entry was checked against the artifact in the version this list names, and whether the identity was checked against a registry. They are recorded here for all ten entries, and **every count in this section is read off this table and not written beside it.**

The column records the CHECK and not a holding, and the difference is not book-keeping. A retained copy is a snapshot of one version and it ages: the work it came from can be revised, and the case study analysed here will have a further version, because this paper corrects it. **The version identifier does not age.** So what the entry rests on is a check made against a named, immutable version, which any reader can repeat from the identifier, and not the presence of a file anywhere.

entry	checked against the named version	identity registry-verified
Besag and Clifford (1989)	no	yes
Camargo (2026)	no	no
Chikina, Frieze, Mattingly and Pegden (2020)	no	no
García Hurtado (2026)	yes	yes
Gotelli and Ulrich (2012)	no	yes
Koning (2024)	yes	no
Koning and Hemerik	yes	no
Mahadevan, Krioukov, Fall and Vahdat (2006)	no	yes
Stoepker and Castro (2024)	yes	no
Theiler, Eubank, Longtin, Galdrikian and Farmer (1992)	no	yes

Four of the ten were checked, in preparing this list, against the artifact in the version this list names. For the other six the check was against the reading record written when the work was read. **The two are not the same warrant and this paper does not present them as one.**

We say it because the difference has already cost something. One entry in this list gave a title that is not the work's title, and the reading record had the right one: the record was correct and the list was not. **A check against a record catches a list that drifted from the record. Only a check against the artifact catches a record and a list that agree and are both wrong**, and for six of these ten that check was not available while this list was being prepared.

Five of the ten are registry-verified on their identity and five are not. The two columns are independent and the table shows all four combinations: the case study is in both, Camargo is in neither, Theiler is registry-verified and was checked against the reading record, and Stoepker was checked against the artifact and has no registry check. **The five that are not registry-verified**

carry the sentence *Identity not registry-verified.* in their own entry below; the five that are carry no such sentence, and its absence is what marks them.

Besag, J., and Clifford, P. (1989). Generalized Monte Carlo significance tests. *Biometrika* 76(4), 633 to 642. Read in the printed article, in full, and cited in section 4.8 for the reachability theorem it supplies in the fixed margin case.

Camargo, S. (2026). Network-normative belief updating in high-dimensional ideological space. Read as arXiv:2605.10726v1. There is no DOI, no journal version and no other arXiv version, and the identity was taken from the artifact's own first page. **Identity not registry-verified.**

Chikina, M., Frieze, A., Mattingly, J. C., and Pegden, W. (2020). Separating effect from significance in Markov chain tests. *Statistics and Public Policy* 7(1), 101 to 114, `10.1080/2330443X.2020.1806763`, taken from the artifact. **Identity not registry-verified.** Read in the published version. A preprint of an earlier paper by three of the authors, arXiv:1608.02014v2, was also read, and its theorem numbering differs from the published one; every theorem reference in this paper is to the published numbering.

García Hurtado, A. (2026). Statistical structure of the historical orderings of the I Ching hexagrams: pair rule, family gradient, and the limits of demonstrability. Zenodo, version DOI 10.5281/zenodo.21609654. The case study of sections 3, 5 and 7. The archived version behind that DOI is the version against which every statement in this paper about that work was checked, including its source text and its replication code, both of which were obtained.

Gotelli, N. J., and Ulrich, W. (2012). Statistical challenges in null model analysis. *Oikos* 121(2), 171 to 180, `10.1111/j.1600-0706.2011.20301.x`. Read in the published version, in full.

Koning, N. W. (2024). More power by using fewer permutations. *Biometrika* 111(4), 1405 to 1412, `10.1093/biomet/asae031`, recorded from the preprint's own front matter. **Identity not registry-verified. Read as arXiv:2307.12832v3, not as the printed article**, which was not obtained. The sentence quoted in section 2.6 is transcribed from the preprint, page 7, the paragraph beginning *Moreover, the result proves*, and no concordance between preprint and print is claimed.

Koning, N. W., and Hemerik, J. More efficient exact group-invariance testing: using a representative subgroup. *Biometrika* 111(2), 441 to 458, `10.1093/biomet/asad050`, recorded from the preprint's own front matter. **Identity not registry-verified. No year is given for this entry on purpose: the work appeared in advance access in 2023 and in print in 2024, this repository recorded both, and neither was chosen.** Read as arXiv:2202.00967v3; the printed article was not obtained and no concordance between the two is claimed. The result cited in section 4.7 is quoted from the preprint.

Mahadevan, P., Krioukov, D., Fall, K., and Vahdat, A. (2006). Systematic topology analysis and generation using degree correlations. *ACM SIGCOMM Computer Communication Review* 36(4), 135 to 146, `10.1145/1151659.1159930`. Read in full, in two versions; the published version is the one cited. A second DOI exists for the same work and is recorded in this paper's repository rather than resolved here.

Stoepker, I. V., and Castro, R. M. (2024). Inference with sequential Monte-Carlo computation of p-values: fast and valid approaches. Read as arXiv:2409.18908v1. There is no DOI and the identity is taken from the artifact. **Identity not registry-verified.** The printing rule of section 6.3 follows the distinction drawn in their own section 1.4 between estimating a decision and estimating a p-value.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D* 58, 77 to 94. Read in the printed article, in full. Every section and page reference in section 2 is to that printing.

9.1 What this paper did not obtain, stated rather than omitted

The case study compares five orderings and this paper uses one of them. Its other two historical orderings rest on philological works that were not obtained in preparing this paper: two on the reconstruction of a manuscript sequence, and one on the palace arrangement that generates a third ordering.

Those three works are not listed above, because this paper does not use the orderings they source, and listing them would claim a consultation that did not happen. They are named here because the case study's data provenance is part of what a reader is relying on when they rely on section 7, and because none of the three was obtained. Section 8 states what follows.

10. Data and code availability

Two deposits are named in this paper and they are different objects. **The CASE STUDY** analysed here is published with its own replication package, and the version-of-record deposit behind DOI `10.5281/zenodo.21609654` is the version against which every statement in this paper about that work was checked. **THE PRESENT PAPER** is deposited under DOI `10.5281/zenodo.21750029`. The two identifiers share their first seventeen characters and differ only from the eighteenth, so they are given here with their roles attached rather than left to be told apart by their digits.

The instruments, records and drafting history of the present paper are deposited in full on Zenodo under `10.5281/zenodo.21750029`, which was reserved before publication so that the deposit can name its own location from inside itself. **The deposit is not a selection made for how it reads: nothing is withheld because of what it shows about this work**, and the list that follows is the demonstration rather than the assurance. It contains the checkers, the numerical instruments, the dated log of every run, the preregistration, the superseded drafts, and the register of defects found in this work's own apparatus.

One class of file is excluded, on a ground that has nothing to do with what it shows, and it is named rather than omitted: copies of works by others. Six third-party articles were retained while this review was made and they are not deposited. **A retained copy is a snapshot of one version and the warrant that matters is the version identifier, which is immutable and public**, so the six are listed in a manifest in the deposit with their size, their sha256 and their identifier, and each is obtainable from that identifier. **Nothing in this paper is computed from any of them**, and the checks that read them are stated as not reached when the file is absent rather than passed over.

Three claims made in this paper are verifiable only against the commit graph of that deposit, and not against any snapshot of its files:

- the preregistration of section 5 was committed before the instrument that ran it;
- the order-theoretic result of section 4 was adopted before the figures of section 7 were computed;
- the retrodiction of section 7.3 was measured before the discriminating statistics of section 7.4 existed.

Each is a statement about the order in which things were done. This is also why the DOI is reserved rather than assigned after the fact: a deposit that cannot state its own address until after it is sealed has to be sealed twice, and the second sealing is the one nobody can check against the first. A deposit of files alone would leave all three as assertions of the authors; the graph makes them checkable by anyone.

Colophon

This document is generated. `PAPER.md` is assembled from marked spans by `assemble.py` and converted to LaTeX by `make_tex.py`. The PDF is compiled from `PAPER.tex` with a standard LaTeX chain and not by any instrument of this repository.

This PDF is NOT reproducible byte for byte. A PDF carries timestamps and embeds fonts, and any two compilations differ, including two by the same person on different dates. **The reproducible objects are `PAPER.md` and `PAPER.tex`, by sha256.** Without the record below, a reader who recompiles and obtains different bytes cannot tell the expected difference from a tampering.

The deposited copy was compiled with pdfTeX 3.141592653-2.6-1.40.27 (TeX Live 2025), at Overleaf, on 2026-08-02.

```
sha256 of PAPER.md  b58e6f8633a247c9c40dea999acd98eae0f677af24656b4e52736039bd8f9d97
```

Rebuild and verify, from a clean clone:

```
git clone null-ladder-history.bundle null-ladder
cd null-ladder
python assemble.py && python make_tex.py
python verify_paper2.py && python mutate_verify.py
```