STABLE LEARNING AND NO-ARBITRAGE PRICING BASED ON SENTIMENTS [1]

By Patrick H. O'Callaghan [2]

Australian Institute for Business and Economics

University of Queensland

A stable learner is a prediction model that generalizes to new samples without re-evaluation of the current kernel. In market settings, a learner is stable if, and only if, its pricing kernel is arbitrage-free. Stable pricing kernels preserve information and support the efficient markets hypothesis. We derive the current kernel on the basis of sentiments (a mapping from samples to rankings) for a broader class of stable learners. This extension of Gilboa and Schmeidler (2003) is justified by US Treasury bond market data for 1961-2023. When the data is insufficiently rich, stability requires out-of-sample validation: sentient agents may simulate or imagine the impact of novel data; artificial ones may engage in leave-one-type-out cross-validation; in market settings, one may bypass sentiments and directly rule out Dutch books. Our framework embeds current sentiments in a system of potential generalizations to enable within-sample testing of stability for any form of external validation.

From the past, the present acts prudently, lest it spoil future action.

Titian: Allegory of Prudence

[1]I thank Itzhak Gilboa for encouraging this project at an early stage. I thank Khoa Hoang for helpful discussions and pointers to the data.

This version has time stamp 21:55 (AEST), 4$^{\text{th}}$ April, 2023. Past versions can be found at `https://arxiv.org/abs/1904.02934` and `https://github.com/patrickocal/stable-learner`.

[2]Email address `p.ocallaghan@uq.edu.au` and ORCID iD `0000-0003-0606-5465`

1 Introduction

In general, prediction requires imagination: the capacity to go beyond the current training set of past cases. To paraphrase (**?**, pages 9 and 10): *the memory is tied down, by original impressions* [past cases]*, without any power of variation; in contrast, the imagination is free to transpose and change its ideas.* Well-established neuroscientific evidence points to an important role for the imagination and the degrees of freedom or flexibility that it provides in improving prediction and thus the chances of survival (**???**). In machine learning, where learners (prediction models) are artificial, the practice of splitting the data into a training set and a test set is a standard (**?**). Cases in the training set resemble memories whereas cases in the test set resemble instances of the imagination in the sense that they allow the learner to explore unseen and potentially novel data.

The goal of external validation is generalization: the ability to adapt appropriately to previously unseen data. The key to generalization is stability (**???**) Loosely speaking, a learner is stable if small changes in the data yield small changes in its predictions: continuity of the learning map from training sets to weighting functions (kernels). **?** provides famous examples of unstable learning where training on new cases causes current kernels to unravel in what is called catastrophic interference.

In this paper, we derive the learner's kernel on basis of sentiments. We formalise *sentiments* as a mapping from samples of past cases to rankings of underlying objects of interest. To account for all kinds of intelligence, we provide a framework that captures the behaviour of learners with the ability to externally validate their model. The axioms we provide are agnostic in relation to the method of external validation. The goal is instead to allow an observer to determine whether or not the learner is able to generalise to richer training sets of past cases in a consistent manner.

Stability is the key axiom in the present derivation of a learner's kernel on the basis of qualitative rankings, given any finite resampling of past cases.[3] The learner's kernel is a subjective assessment of how similar past cases are to the

[3] Here a *kernel* is simply a real-valued weighting function of two variables.

current prediction problem. Stable learners generalize to novel case types without needing to: re-evaluate past rankings; generate intransitive rankings; or suppress novel (transitive) rankings. In essence, a stable learner's kernel is not only stable in a literal sense, it also allows for meaningful, unbiased generalisations.

Stable learners stand in contrast to those that learn by doing. Children are encouraged to learn by making (noncatastrophic) mistakes. They are free to modify their beliefs and behaviour after the event: the intransitive rankings that novel case types may reveal (**?**) are of little consequence. But what of a market maker that exposes their pricing kernel either directly or indirectly (by buying and selling securities) on a financial market? We show that a market maker fails to be stable if, and only if, the returns on investment that they offer provide a free lunch to other agents today (i.e. before any novel case types arrive and before they have a chance to re-evaluate). This equivalence between stability (when novel data arrives tomorrow) and no-arbitrage (today) is at the heart of the present paper and it has a number of interesting implications.[4]

First, at the individual level, unstable learners are unlikely to survive.

Second, we can express stable pricing kernels as an empirical (geometric) mean. This means that, in stable markets (i.e. those with only stable learners), prices are efficient stores of information. A stable market price formation process is modular in cases and separable in securities. This implies that, if we are also given the order that past cases arrived, we can identify the stable market pricing kernel (i.e. the weighting function implied by market prices).

Third, what if, in advance, learners "massage" their kernels to ensure they are arbitrage-free (and thus stable)? This alternative method of external validation is reminscent of De Finetti's Dutch books argument for the formula of conditional probabilities: isn't this what bookies do? But then how then should an observer distinguish the stable learners that are prudent (i.e. ones that actually use their imagination or cross-validates their model)? This observational ambivolence carries over to the aggregate level. The market pricing kernel might be stable even if none of the individual agents in the market are prudent. This is a Deus-ex-machina

[4]In general, the learner is stable if, and only if, its kernel forms a groupoid over eventualities. This concept corresponds to the Jacobi equations in GS03.

or invisible-hand property: the market process provides benefits independently of the wisdom of its agents. All other things being equal, markets where arbitrage is easier to exploit should have more stable market pricing kernels.

The second implication speaks directly to the timeless topic of market efficiency **??**. Stable pricing kernels provide an explanation for why we might expect a passive (i.e. buy-and-hold) strategy to perform at least as well as any other strategy. (The semi-martingale hypothesis of **?**.) This explanation is as follows. If the market pricing kernel is stable, then the success of an active strategy depends solely on its ability to predict the impact of the next case type. Whilst this does simplify the task, it also means that the impact of the next case is unlikely to be large. This is the nature of the empirical mean: the most likely cases are those that have occurred most frequently in the past. The impact of a new case is likely to be crowded out by the frequency of those that it resembles. On the flip side, unstable pricing kernels come with arbitrage opportunities and learners modifying their kernels. Moreover, by virtue of the fact that unstable market pricing kernels cannot be expressed as an empirical mean: price movements are likely to be larger; passive strategies are less attractive; active strategies may indeed be less volatile.

The third implication motivates the present generalization of GS03. Similar to GS03, the learner is endowed with a mapping from finite resamplings of past cases to rankings of eventualities (sentiments). Via the basic axioms, all learners (stable or not) are characterised by a kernel that, for every pair of eventualities, gives rise to a linear function on samples of past cases. With sufficiently rich data, prudence (or the threat of arbitrage) is superfluous. Who needs an imagination when raw memory will do? To be more precise, a translation of the main result of GS03 to the present context tells us that a learner is guaranteed to be stable whenever resampling of past cases generates a *4*-diversity of rankings: for every subset of $k \leqslant 4$ eventualities, all $k!$ total orderings (of those k eventualities) feature in sentiments. The *4*-diversity axiom is sufficient but not necessary for a kernel that, for every single eventuality, gives rise to a linear function on samples of past cases. The stability axiom allows us to substantially weaken *4*-diversity to partial-*3*-diversity: for every subset of $k \leqslant 3$ eventualities, at least k total orderings feature in sentiments. Our main theorem then leads to the following hypothesis: we expect unstable learning to arise in settings where *4*-diversity fails to hold and where prudence is required to achieve stability: i.e. where the learners are inexperienced (short of relevant data) and where arbitrage opportunities are difficult to exploit.

Before turning to the model, let us consider some examples. These examples feature two kinds of learner: one who only looks back at past cases (within-sample, Inny), and another who also considers the potential impact of novel types of case (the prudent, Pru).

Example 1 (A setting where *4*-diversity holds)**.** *A tali (a primitive four-sided die) is rolled over and over again.*[5] *So far, it yields only ones, twos and threes, so that the empirical distribution, conditional upon the current data, assigns zero probability to fours. By resampling (with replacement), we obtain the empirical distribution conditional on potential samples of past cases. This conditional empirical distribution induces a single ranking of the set* $X = \{1, 2, 3\}$ *of eventualities (outcomes) for each sample. The resulting rankings map generates the arrangement of hyperplanes in fig. 1. To each region of this arrangement, sentiments assigns a distinct total ranking. In dice-like problems, the inherent symmetry between outcomes and past cases ensures all* $3! = 6$ *total rankings of* X *arise.*

Our learners, Inny and Pru, forecast the outcome of the next roll. Both reveal plausibility rankings that agree with fig. 1. By virtue of 4-diversity*, their ranking maps generalise to higher dimensions (four case types) without any need for external validation. The data is sufficiently rich for making predictions about* X*.*

Example 2 (A non-market setting where *4*-diversity fails to hold)**.** *Each month, the federal reserve announces its target overnight rate. Inny and Pru wish to forecast the next announcement. The consensus is that there will be a rate rise of* $0, 25$ *or* 50 *basis points. Inny and Pru have observed the same past announcements. Current circumstances are such that they both find that past cases generate just three distinct case types* t_1*,* t_2 *and* t_3 *where the indices capture the (increasing) degree to which the fed is "behind the curve". At (samples containing only cases of type)* t_3*, Inny and Pru agree with the plausibility ranking* $(0, 25, 50)$*, so that larger*

[5] Similar to a modern tetrahedral die, the winning number lands face down.

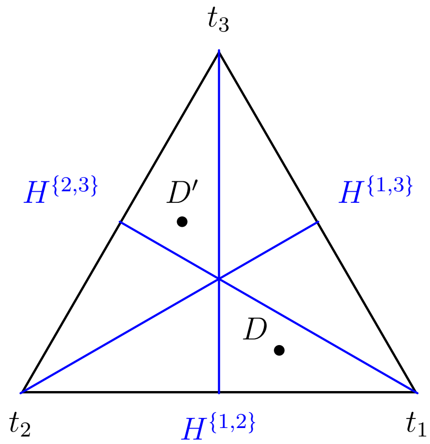


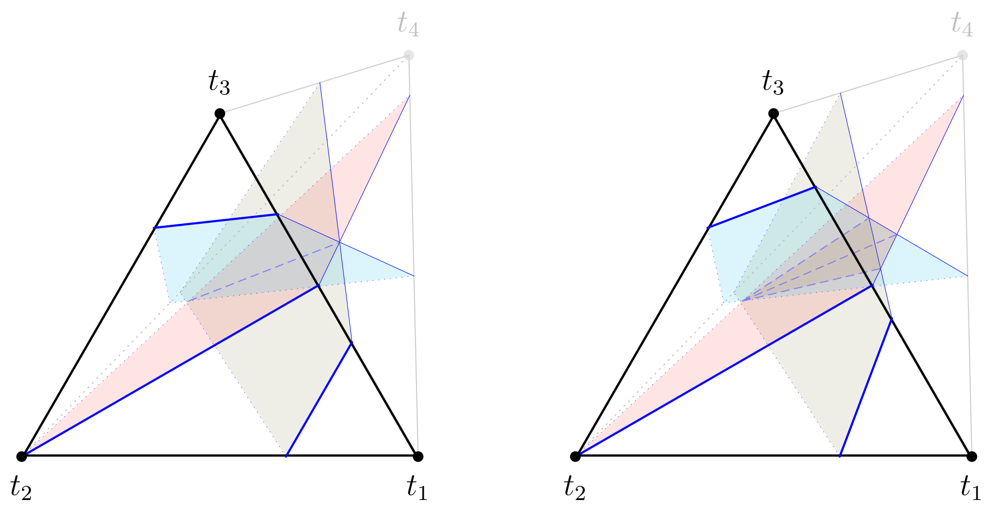
Figure 1: Each rational point in this simplex corresponds to a sample of past case types: t_1, t_2 and t_3. The total ranking $(1, 2, 3)$ arises at D (one is least likely and three is most likely). This reflects the relative frequencies of past cases in D. The inverse ranking $(3, 2, 1)$ arises at D'. Within each hyperplane $H^{\{x,y\}}$, faces x and y are equally plausible. The three hyperplanes form a centered arrangement.

rate rises are more likely. At t_1, the opposite ranking holds: both find that smaller rate rises are more likely. At t_2, they agree ranking is $(\{0, 50\}, 25)$, so that a 25 rise is most likely, and the others are equally plausible.

Figure 2 shows some subtle differences between their current (bold) rankings maps. These subtle differences turn out to be substantial once we extend each learner's rankings map to a hypothetical fourth case type. To test for stability, the observer assigns t_4 the novel ranking $(25, 0, 50)$ and then extends both learners' current rankings maps to samples that include cases of type t_4. Pru's extended rankings map can be chosen to form a congruent arrangement with six regions. In contrast, Inny's extended rankings map forms an incongruent arrangement with seven regions. Samples in the seventh region give rise to the intransitive ranking $(25, 0, 50, 25)$. We may not be able to observe Pru's imagination (cross-validation method), but since incongruence is a generic property for triples of hyperplanes, we can prove that he is almost certainly using one.

Example 3 (a bond-market setting that we develop in section 4)**.** *Inny and Pru are now fair market makers of* Zeros *(zero-coupon treasury bonds).*[6] *The*

[6]Similar to a fair insurer, a fair market maker sets the market spread to zero.



(a) Pru's rankings map (b) Inny's rankings map

Figure 2: In contrast with fig. 1, the current (bold) rankings maps features four regions and the current hyperplane arrangement is uncentered.

compound-interest formula for the accumulation process of such bonds is

$$a^{(x,y)} = \left(1 + r^{\{x,y\}}\right)^{-x+y},$$

where $r^{\{x,y\}}$ is the yield to maturity on a forward contract that accrues interest between dates x and y. If x is later than y, then the contract is to sell, and the market maker pays this yield, so that $r^{\{y,x\}} = r^{\{x,y\}}$. Let $X \subseteq \mathbb{R}_+$ index a suitable sequence of trading dates with $0 \in X$ being the spot date. The log-accumulation process has no arbitrages provided it satisfies

$$\text{for every distinct } x, y, z \in X, \quad \log a^{(x,y)} = \log a^{(x,z)} + \log a^{(z,y)}.^{7} \tag{1}$$

In eq. (6), reference to the data that drives the market-makers' prices is suppressed. When $\log a^{(x,y)}$ depends on past cases, it forms a vector that is normal (i.e. orthogonal) to a hyperplane such as those that separate regions in ranking maps such as fig. 1 or fig. 2. When the data is rich enough to generate the full diversity rankings (as in example 1), both Inny and Pru set prices that are currently arbitrage-free. Otherwise, only Pru satisfies this property. This is true even

[7] To see this, suppose that, for $x = 0 < y < z$, Inny sets $a^{(x,y)} < a^{(x,z)} \cdot a^{(z,y)}$. Pru would do well to buy the spot contract (x, y) and sell both the spot contract (x, z) and the forward contract (z, y). This yields a risk-free profit of $a^{(x,z)} \cdot a^{(z,y)} - a^{(x,y)} > 0$.

though they both observe the same collection of past cases. Equivalently, only Pru sets prices with the canonical form of an empirical (geometric) mean

$$B(x, D) = \prod_{c \in D} \left(1 + r_c^{\{0,x\}}\right)^{-x/n} \tag{2}$$

for every date x and finite sample D that we can generate by resampling (with replacement) from past cases.

Example 4 (A multi-sectoral investment setting)**.** *Suppose Inny and Pru are potential representative agents in a multi-sectoral macroeconomy (**??**). Instead of maturity dates, as in example 3, X is a variety of sectors with x_0 being the household (i.e. consumption sector). The following no-arbitrage equations may be derived from first-order conditions of the social planner's constrained optimisation problem provided we specify an explicit functional form for the instantaneous flow of utility. But here we take a nonparametric approach: we avoid the specification of a utility function and proceed from past cases to beliefs and decisions.*

For every pair of sectors x and y and nonempty set D of past cases, let $S_D^{(x,y)}$ denote a suitably normalised quantity of sector x inputs that the social planner invests in order to produce sector y outputs, conditional on D; then, for every sector z, the corresponding no-arbitrage equations are

$$\log S_D^{(x,y)} = \log S_D^{(x,z)} + \log S_D^{(z,y)} .^{8} \tag{3}$$

Just as in example 3, when the data is sufficiently rich to generate a sufficient diversity of rankings, Inny and Pru are indistinguishable. Otherwise, only the prudent social planner, Pru, has a rankings map that satisfies the above no-arbitrage conditions; only Pru sets the required rate of return on investment (with the gross return $R(x_0, \cdot)$ to consumption normalised to one) according to the canonical form of an empirical mean

$$R(x, D) = \prod_{c \in D} \left(1 + r_c^{\{x_0,x\}}\right)^{1/|D|} \tag{4}$$

for every sector x and finite sample D that we can generate by resampling (with replacement) from past cases; finally, only Pru satisfies the inter-sectoral (and

[8] We can confirm that, for rather general model specifications, these equations hold in our traditional simulations with explicit utility functions.

inter-temporal) consumption Euler equations

$$\log S_D^{(x,y)} = \log R(y, D) - \log R(x, D). \tag{5}$$

Example 5 (Supervised learning)**.** *We are given a training set* $D \subset X \times Y$ *consisting of* n *i.i.d cases (i.e. examples) drawn from some unknown probability distribution on* $(X \times Y)^n$*. That is* $D = \{(x_i, y_i) : i = 1, \ldots, n\}$*. The basic goal of supervised learning is to use the training set* D *to "learn" a function* $f_D : X \to Y$ *such that, for any given new value* x_{new}*,* f_D *provides an accurate prediction of the associated value* y_{new}*. That is, we hope that* $f_D(x_{\text{new}})$ *is suitably close to* y_{new}*.*

The rest of the paper is organised as follows. In section 2 we present the model and key definitions. In section 3 we present the axioms and main representation theorem. In section 4 we develop example example 3 by deriving a representation result for a bond-market setting. In section 4 we also discuss the implications for market efficiency identification of pricing kernels in more detail. In section 5 we provide a more detailed comparison of our main theorem with that of GS03 as well as matters such as second-order induction (**?**).

2 Model

The primitives of our model consist of the nonempty sets X and $\mathbb{C}^{\mathfrak{f}}$. For the current prediction problem, we interpret members of X as *eventualities* and members of $\mathbb{C}^{\mathfrak{f}}$ as *current* cases. Our first departure from GS03 is to allow for two forms of current case. That is we take $\mathbb{C}^{\mathfrak{f}}$ to be the union $\mathbb{C} \cup [\mathfrak{f}]$ of a set $\mathbb{C}$ of constant, *past* cases and a set $[\mathfrak{f}]$ of copies of the variable, *free* case $\mathfrak{f}$.

Remark. *By way of analogy with computer memory, a natural implementation is as follows. Take any case* $c \in \mathbb{C}^{\mathfrak{f}}$ *to consist of a pair* $p \times m$ *: a pointer* p *that references a memory location and the memory content* m*. For* $c \in \mathbb{C}$*, just as in the setting of* GS03*, content of* m *is meaningful. For* $c \in [\mathfrak{f}]$*, although* m *is empty (assigned a "garbage" value), the allocation is itself valuable. Pairs* $c, d \in [\mathfrak{f}]$ *are indistinguishable in terms of content and are in this sense copies of* $\mathfrak{f}$*.*

The learner is free to assign novel, imagined content to $\mathfrak{f}$ *and before going on to explore the prediction problem with this content in mind. As we will show,* *an observer (or more sophisticated learner) who observes past cases, but not the imagination, can nonetheless use* $\mathfrak{f}$ *to analyse the learners' behaviour.*

With case *resampling* from the literature on bootstrapping in mind, let

$$\mathbb{D} \overset{\text{def}}{=} \{D \subseteq \mathbb{C} : \#D < \infty\}$$

denote the set of (finite) *determinate or constant databases* or memories and let $\mathbb{D}^{\mathfrak{f}}$ denote the corresponding set of all finite subsets of $\mathbb{C}^{\mathfrak{f}}$. Finally, throughout the sequel, we take $\mathfrak{C}$ to denote a member of $\{\mathbb{C}, \mathbb{C}^{\mathfrak{f}}\}$ without reference to the latter set. Similarly,

$$\mathfrak{D} = \begin{cases} \mathbb{D} & \text{if, and only if, } \mathfrak{C} = \mathbb{C}, \text{ and} \\ \mathbb{D}^{\mathfrak{f}} & \text{otherwise.} \end{cases}$$

Sentiments. We now take a first step towards formalising the ranking maps of figs. 1 and 2. (For the translation from databases to rational vectors, see section A.) For each D in $\mathbb{D}$, the learner is endowed with a well-defined plausibility ranking $\preceq_D$ in the set $\mathrm{rel}(X)$ of binary relations on X. Denote the symmetric and asymmetric parts of $\preceq_D$ by $\simeq_D$ and $\prec_D$ respectively.

Sentiments $D \mapsto \preceq_D$ are thus a vector in $\mathrm{rel}(X)^{\mathbb{D}}$ of the form

$$\preceq_{\mathbb{D}} \overset{\text{def}}{=} \langle \preceq_D : D \in \mathbb{D} \rangle .$$

This subtle departure from the form $\{\preceq_D : D \in \mathbb{D}\}$ of GS03 is closer in structure to visual representations (figs. 1 and 2) and the graph $\{D \times \preceq_D : D \in \mathbb{D}\}$.[9]

Case types. As in GS03, two past cases $c, d \in \mathbb{C}$ are of the same *case type* if, and only if, the marginal information of c is everywhere equal to the marginal information of d. Formally, $c \sim^\star d$ if, and only if, for every $D \in \mathbb{D}$ such that $c, d \notin D$, $\preceq_{D \cup \{c\}} = \preceq_{D \cup \{d\}}$. By observation 1 of GS03, $\sim^\star$ is an equivalence relation on $\mathbb{C}$. The equivalence classes of $\sim^\star$ generate a partition $\mathbb{T}$ of case types.

We extend $\sim^\star$ to $\mathbb{C}^{\mathfrak{f}}$ by taking $[\mathfrak{f}]$ to be an equivalence class of its own, so that, for every $c \in \mathbb{C}$, $c \not\sim^\star \mathfrak{f}$. We let $\mathbb{T}^{\mathfrak{f}}$ denote the corresponding partition of $\mathbb{C}^{\mathfrak{f}}$. Like GS03, we also extend $\sim^\star$ to $\mathbb{D}^{\mathfrak{f}}$ by treating databases that contain the same number of each case type as equivalent. That is, $C \sim^\star D$ if, and only if, for every

[9] I thank Maxwell B. Stinchcombe for bringing this point to my attention.

$t \in \mathbb{T}^{\mathfrak{f}}$, the numbers $\#(C \cap t)$ and $\#(D \cap t)$ of that case type coincide. To enable a surjection $D \mapsto \langle \#(D \cap t) : t \in \mathbb{T}^{\mathfrak{f}} \rangle$ from databases to counting vectors, we impose

Richness Assumption. *For every $t \in \mathbb{T}^{\mathfrak{f}}$, there are infinitely many cases in t.*

Generalizations of sentiments $\preceq_{\mathbb{D}}$ to $\mathbb{D}^{\mathfrak{f}}$ require a suitable structure. The following definition provides that structure. It also simplifies the exposition by accommodating generalizations that restrict attention to subsets Y of X.

Definition 1. $\mathcal{R} \stackrel{\text{def}}{=} \langle \mathcal{R}_D : D \in \mathfrak{D} \rangle$ *is a* generalization *(or Y-generalization) of* $\preceq_{\mathbb{D}}$ *if, for some nonempty $Y \subseteq X$, the following all hold*:

1. *for every $D \in \mathbb{D}$ and every $x, y \in Y$, $x \mathcal{R}_D y$ if, and only if, $x \preceq_D y$,*
2. *for every $D \in \mathfrak{D}$, $\mathcal{R}_D$ belongs to* $\operatorname{rel}(Y)$*,*
3. *for every $D \in \mathfrak{D}$ and every $c, d \in \mathfrak{C} - D$, if $c \sim^{\star} d$ then $\mathcal{R}_{D \cup \{c\}} = \mathcal{R}_{D \cup \{d\}}$.*

A generalization $\mathcal{R}_{\mathfrak{D}}$ is proper *if $\mathfrak{D} = \mathbb{D}^{\mathfrak{f}}$ and improper otherwise.*

Part 1 of the definition implies that, for every Y-generalization $\mathcal{R}$ and every $D \in \mathbb{D}$, $\mathcal{R}$ is simply the restriction $\preceq_D \cap Y^2$ of $\preceq_D$ to Y. We therefore refer to part 1 of the definition of a generalization as the preservation or *nonrevision* condition.

Part 2 of the definition of a generalization ensures that, for every proper Y-generalization $\mathcal{R}$ and every $D \in \mathbb{D}^{\mathfrak{f}}$, $\mathcal{R}_D$ is a well-defined binary relation on Y. Let $\mathcal{I}_D$ and $\mathcal{P}_D$ respectively denote the symmetric and asymmetric parts of $\mathcal{R}_D$.

Part 3 of the definition ensures that, for proper generalizations $\mathcal{R}$, the partition $\mathbb{T}^{\mathfrak{f}}$ of case types generated by $\sim^{\star}$ is at least as fine the partition generated by the equivalence relation generated by $\mathcal{R}$. Two cases $c, d \in \mathbb{C}^{\mathfrak{f}}$ are *equivalent with respect to* $\mathcal{R}$, written $c \sim^{\mathcal{R}} d$, if, for every $D \in \mathbb{D}$ such that $c, d \notin D$, $\mathcal{R}_{D \cup \{c\}} = \mathcal{R}_{D \cup \{d\}}$. This notion allows us to partition the set of proper generalizations as follows.

Definition. *A proper generalization $\mathcal{R}$ is either* regular *or* novel. *It is novel whenever $[\mathfrak{f}]$ is a distinct equivalence class of $\sim^{\mathcal{R}}$, so that, for every $c \in \mathbb{C}$, $c \not\sim^{\mathcal{R}} \mathfrak{f}$.*

For any given novel generalization $\mathcal{R}$, $[\mathfrak{f}]$ is the unique novel case type that $\sim^{\mathcal{R}}$ generates. We impose this restriction, not because we think the imagination of learners is constrained in this way, or because the model would not work that way, but rather because one degree of freedom is sufficient for our purposes. From an evolutionary perspective: Pru represents a minimal deviation from Inny.

For every regular generalization $\mathcal{R}$, there exists $c \in \mathbb{C}$ such that $c \sim^{\mathcal{R}} \mathfrak{f}$. Thus, there are as many regular generalizations as there are past case types (i.e. $\#\mathbb{T}$). Yet every regular Y-generalization $\mathcal{R}$ is *equivalent* to the unique improper Y-generalization $\langle \preceq_D \cap Y^2 : D \in \mathbb{D} \rangle$ in the sense of

Observation 1. *For every regular Y-generalization $\mathcal{R}$ and improper Y-generalization $\acute{\mathcal{R}}$, for every $C \in \mathbb{D}^{\mathfrak{f}}$, there exists $D \in \mathbb{D}$ such that $C \sim^{\mathcal{R}} D$ and $\mathcal{R}_C = \acute{\mathcal{R}}_D$.*[10] *of*

See proof on page 45.

3 Axioms and main results

The basic axioms of GS03, which we rewrite in terms of generalizations, are the following. In each of these axioms, $\mathcal{R}$ is an arbitrary Y-generalization of $\preceq_{\mathbb{D}}$.

A0 (Transitivity axiom for $\mathcal{R}$). *For every $D \in \mathfrak{D}$, $\mathcal{R}_D$ is transitive.*

A1 (Completeness axiom for $\mathcal{R}$). *For every $D \in \mathfrak{D}$, $\mathcal{R}_D$ is complete.*

A2 (Combination axiom for $\mathcal{R}$). *For every disjoint $C, D \in \mathfrak{D}$ and every $x, y \in Y$, if $x \mathcal{R}_C y$ and $x \mathcal{R}_D y$, then $x \mathcal{R}_{C \cup D} y$; and if $x \mathcal{P}_C y$ and $x \mathcal{R}_D y$, then $x \mathcal{P}_{C \cup D} y$.*

A3 (Archimedean axiom for $\mathcal{R}$). *For every disjoint $C, D \in \mathfrak{D}$ and every $x, y \in Y$, if $x \mathcal{P}_D y$, then there exists $k \in \mathbb{Z}_{++}$ such that, for every pairwise disjoint collection $\{D_j : D_j \sim^{\mathcal{R}} D \text{ and } C \cap D_j = \varnothing\}_1^k$ in $\mathfrak{D}$, $x \mathcal{P}_{C \cup D_1 \cup \cdots \cup D_k} y$.*

The diversity axioms that now follow require that $\mathbb{D}$ is sufficiently rich to support Y-generalizations $\mathcal{R}$ with a variety of *total* orderings: i.e. complete, transitive and antisymmetric ($x \mathcal{R}_D y$ and $y \mathcal{R}_D x$ implies $x = y$). Let total($\mathcal{R}$) denote the set $\{R : \text{for some } C \in \mathfrak{D},\ R = \mathcal{R}_C \text{ is total}\}$ of of total orders that feature in $\mathcal{R}$. For $k = 4$, the following axiom is a restatement of the diversity axiom of GS03.

[10] This means that there is a canonical embedding of $\{C \times \mathcal{R}_C : C \in \mathbb{D}^{\mathfrak{f}}\}$ in $\{D \times \acute{\mathcal{R}}_D : D \in \mathbb{D}\}$. The converse embedding follows from the nonrevision condition of definition 1.

Diversity (k-Diversity axiom). *For every $Y \subseteq X$ of cardinality $n = 2, \ldots, k$, every regular Y-generalization $\mathcal{R}$ of $\preceq_{\mathbb{D}}$ is such that* $\#\mathrm{total}(\mathcal{R}) = n!$.

We say *k-diversity holds on Z* if the axiom holds with Z in the place of X. By lowering the bar for the required number of total orders, the following axiom substantively weakens *3*-diversity and *a fortiori 4*-diversity.

A4′ (Partial *3*-diversity). *For every $Y \subseteq X$ with cardinality $n = 2$ or 3, every regular X-generalization $\mathcal{R}$ of $\preceq_{\mathbb{D}}$ is such that $\#\mathrm{total}(\mathcal{R}) \geqslant n$.*

Example 2 (and example 6 of the appendix) provide examples of $\preceq_{\mathbb{D}}$ satisfying A0–A4′, but not *3*-diversity. For settings where the basic axioms hold, the next observation shows that partial-*3*-diversity is equivalent to

A4 (Conditional-*2*-diversity). *For every three distinct elements $x, y, z \in X$, one of the sets $\{D' : x \prec_{D'} y\}$ and $\{D' : y \prec_{D'} x\}$ contains both C and D such that $z \prec_C x$ and $x \prec_D z$. If $\#X = 2$, then 2*-diversity *holds on X.*

Observation 2. *For $\preceq_{\mathbb{D}}$ satisfying* A0–A3*,* conditional-*2*-diversity *and* partial-*3*-diversity *are equivalent.*

Proof. This follows by virtue of proposition 3 and the translation of section A. □

Partial-*3*-diversity is therefore stronger than *2*-diversity and, moreover, it plays the dual role of guaranteeing the existence and uniqueness of the representation. Via example 7, partial-*3*-diversity is weakest axiom with these properties.

The stability axiom is our final requirement. It is distinguished by the fact that it imposes structure on novel generalizations. First a definition that reduces the set of generalizations we test when establishing stability.

Definition. *A proper generalization $\mathcal{R}$ of $\preceq_{\mathbb{D}}$ is* testworthy *if it satisfies* A1–A3 *and, for some $D \in \mathbb{D}$ such that $\mathcal{R}_D$ is total, $\mathcal{R}_{\mathfrak{f}}$ is the inverse of $\mathcal{R}_D$.*[11]

Testworthy generalizations are distinguished by the ranking at $\mathfrak{f}$ and the fact that they need not satisfy A0. Stability requires that for any testworthy generalization we can find a suitable perturbation that also satisfies A0.

[11]Recall that the inverse $\mathcal{R}_D^{-1}$ of $\mathcal{R}_D$ satisfies $x \, \mathcal{R}_D^{-1} \, y$ if, and only if, $y \, \mathcal{R}_D \, x$.

Definition. *Let $\mathcal{R}$ and $\acute{\mathcal{R}}$ be generalizations of $\preceq_{\mathbb{D}}$. $\acute{\mathcal{R}}$ is a* perturbation *of $\mathcal{R}$ if $\acute{\mathcal{R}}_{\mathfrak{f}} = \mathcal{R}_{\mathfrak{f}}$ and, moreover, a* diverse *perturbation if* $\#\mathrm{total}(\acute{\mathcal{R}}) \geqslant \#\mathrm{total}(\mathcal{R})$.

A diverse perturbation of $\mathcal{R}$ does not suppress the novel, transitive rankings that $\mathcal{R}$ generates. If we allowed for nondiverse (i.e. dogmatic) perturbations, the learner could continue to "hide" intransitive rankings. Our main axiom is then

***4*-Stability.** *For every $Y \subseteq X$ with cardinality 3 or 4, every testworthy Y-generalization of $\preceq_{\mathbb{D}}$ that is novel has a diverse perturbation that satisfies* A0–A3.

Since every proper generalization is either regular or novel, the next observation confirms that *4*-diversity holds whenever *4*-stability holds vacuously; that is, whenever there are no novel generalizations for us to test.

Observation 3. *Let $\preceq_{\mathbb{D}}$ satisfy* A0–A4. *For every $Y \subseteq X$ of cardinality 3 or 4, the set of testworthy Y-generalizations is nonempty. If, for some Y, every testworthy Y-generalization is regular, then* *4*-diversity *holds on Y.*

Proof. Respectively, these two statements follow via lemma 3.2 and lemma 3.1. □

Our main theorem involves real-valued function $\mathbf{v}$ on the product $X \times \mathbb{C}$. We view $\mathbf{v}$ as a matrix and $\mathbf{v}(x, \cdot)$ as one of its rows. The matrix $\mathbf{v}$ is a *representation of* $\preceq_{\mathbb{D}}$ whenever it satisfies

$$\begin{cases} \text{for every } x, y \in X \text{ and every } D \in \mathbb{D}, \\ x \preceq_D y \quad \text{if, and only if,} \quad \sum_{c \in D} \mathbf{v}(x, c) \leqslant \sum_{c \in D} \mathbf{v}(y, c). \end{cases}$$

The matrix $\mathbf{v}$ *respects case equivalence* (with respect to $\preceq_{\mathbb{D}}$) if, for every $c, d \in \mathbb{C}$, $c \sim^{\star} d$ if, and only if, the columns $\mathbf{v}(\cdot, c)$ and $\mathbf{v}(\cdot, d)$ are equal.

Theorem 1 (Part I, Existence). *Let there be given X, $\mathbb{C}^{\mathfrak{f}}$, $\preceq_{\mathbb{D}}$ and associated generalizations, as above, such that the richness condition holds. Then* (1.i) *and* (1.ii) *are equivalent.*

(1.i) A0–A4 *and* *4*-stability *hold for* $\preceq_{\mathbb{D}}$.

(1.ii) *There exists a matrix $\mathbf{v} : X \times \mathbb{C} \to \mathbb{R}$ satisfying* (1.a) *and* (1.b) :

(1.a) $\mathbf{v}$ *is a representation of $\preceq_{\mathbb{D}}$ that respects case equivalence*;

(1.b) *no row of* $\mathbf{v}$ *is dominated by any other row, and for every three distinct elements* $x, y, z \in X$ *and* $\lambda \in \mathbb{R}$, $\mathbf{v}(x, \cdot) \neq \lambda \mathbf{v}(y, \cdot) + (1 - \lambda)\mathbf{v}(z, \cdot)$.[12]

Our uniqueness result is identical to that of GS03.

Theorem 1 (Part II, Uniqueness)**.** *If* (1.i) [*or* (1.ii)] *holds, then the matrix* $\mathbf{v}$ *is unique in the following sense: for every other matrix* $\mathbf{u} : X \times \mathbb{C} \to \mathbb{R}$ *that represents* $\preceq_{\mathbb{D}}$*, there is a scalar* $\lambda > 0$ *and a matrix* $\beta : X \times \mathbb{C} \to \mathbb{R}$ *with identical rows (i.e. with constant columns) such that* $\mathbf{u} = \lambda \mathbf{v} + \beta$*.*

A variety of statistical methods that are related to theorem 1 are discussed in GS03. In the next section, we provide a more detailed application to no-arbitrage asset pricing. This application also appeals to intermediate results in the proof of theorem 1, so let us unpack the key steps.

The first step in the proof of theorem 1 (see section A is to translate all the above concepts to the setting of rational vectors indexed by case types: as in figs. 1 and 2, i.e. vectors in $\mathbb{R}_+^{\mathbb{T}}$. Theorem 2 is the corresponding theorem for rational vectors. Definitions and intermediate results in the proof of theorem 2 are the subject of section B. For now, we maintain the discussion of these results in terms of samples (i.e. databases).

In section B, we show that A1–A4 alone yield a more general matrix representation than that of theorem 1. In particular, a matrix $v^{(\cdot,\cdot)} : X^2 \times \mathbb{C} \to \mathbb{R}$, $x \times y \times c \mapsto v_c^{(x,y)}$. For each pair $x, y \in X$, the row $v^{(x,y)}$ of this matrix is a pairwise representation provided that, for every $x, y \in X$ and $D \in \mathbb{D}$,

$$x \preceq_D y \text{ if, and only if, } \sum_{c \in D} v_c^{(x,y)} \geqslant 0.$$

(In the context of fig. 1, the projection of each such $v^{(x,y)}$ onto case-type space is orthogonal to the hyperplane $H_+^{\{x,y\}} \subset \mathbb{R}_+^{\mathbb{T}}$.)

A4 is characterised by the property that no row of $v^{(\cdot,\cdot)}$ dominates any another; nor is any pair collinear. The final key to the proof of theorem 1, stability, and the connection with no-arbitrage pricing is

[12]Observe that $\mathbf{v}(x, \cdot) - \mathbf{v}(z, \cdot)$ and $\mathbf{v}(y, \cdot) - \mathbf{v}(z, \cdot)$ are noncollinear if, and only if, the affine independence condition of (1.b) holds.

Definition. *For $Y \in 2^X$, $v^{(\cdot,\cdot)} : Y^2 \times \mathbb{C} \to \mathbb{R}$ satisfies the* groupoid property *(Jacobi identity) whenever, for every $x, y, z \in Y$, $v^{(x,z)} = v^{(x,y)} + v^{(y,z)}$.*

When the Jacobi identity holds on X, instead of the $\binom{\#X}{2}$ rows of $v^{(\cdot,\cdot)}$, we only need $\#X$ vectors in $\mathbb{R}^{\mathbb{C}}$ to summarise $\preceq_{\mathbb{D}}$. This is the crux of theorem 1.

Corollary 1 (a characterisation of stability)**.** *Let the number of case types be finite and let $\preceq_{\mathbb{D}}$ satisfy* A4*. Then $\preceq_{\mathbb{D}}$ satisfies 4*-stability *if, and only if, $\preceq_{\mathbb{D}}$ has a pairwise representation $v^{(\cdot,\cdot)}$ that satisfies the Jacobi identity. Moreover, for every other pairwise representation $u^{(\cdot,\cdot)}$, there exists $\lambda > 0$, such that $u^{(\cdot,\cdot)} = \lambda v^{(\cdot,\cdot)}$.*

See proof on page 45.

This suffices for us to formalise the log-accumulation process of example 3.

4 Application to no-arbitrage asset pricing

Fixed income securities (henceforth *bonds*) are commonly traded over the counter with certain parties (typically investment banks) acting as market makers. Similar to bookmakers in betting markets, market makers profit from the bid-ask spread. Unlike bookmakers, however, they do sieze on perceived opportunities to make money by trading with other market makers. To simplify the exposition, we assume the market makers are *fair*, so that the bid-ask spread is zero.

Recall that a (fixed income) *forward* is a contract between two parties to exchange a bond at a given date, price and amount in the future. Building on example 3, let X denote the set of settlement/maturity dates for forwards associated with given zero-coupon bond (a *Zero*) issued by the same entity. In particular, take $X = \{x_0, x_1, \ldots\}$ such that $x_0 = 0$ and $x_k < x_{k+1}$ for all feasible k.

Remark (on zero-coupon bonds)**.** *Although coupon-paying bonds are far more common, Zeros provide a natural asset for studying no-arbitrage pricing (**?**). Zeros are generated by "stripping" coupon-paying bonds of their coupons enroute to deriving benchmark zero-coupon yield curves (**?**). Since coupon stripping is reversible, restricting attention to Zeros is without loss of generality. Modulo surmountable complications relating to the uncertainty of the cashflows, the arguments that follow extend to dividend-paying stocks.*

Let $D^\star$ denote the (unique) current history of time-series data that are relevant to this market for Zeros. We generate the set $\mathbb{C}$ of past cases by taking each case $c \in \mathbb{C}$ to consist of market-relevant data from a given time interval (a block of time periods) in the past. Block resampling of time series data was originally developed by **??**. (For applications to finance problems see **??**.) For the present purposes, blocks need to be chosen so that, for any $k > 0$ and resampling $D = \{c_1, \ldots, c_k\}$, what matters is the number of repetitions of a given case type, not the order of the cases that form the pseudo-time series. In other words, $D \sim^\star D'$ for any permutation D' of the cases in D. It is straightforward to verify that this is indeed the case for the stationary bootstrap of **?**, section 2.

The free case $\mathfrak{f}$ has no additional structure beyond that of sections 2 and 3.

Before turning to the interpretation of $\preceq_{\mathbb{D}}$ in the present setting, we introduce current market prices. From the market maker's perspective, current prices reflect the rest of the market's view on yields to maturity conditional on $D^\star$. Since we take current market prices (and rates) to be fixed, we suppress reference to $D^\star$. The market maker observes past cases, and current prices and forms a view about how prices might change when the information changes. The compound-interest formula for the *market accumulation process* (at current market rates) is then

$$a^{(x,y)} = \left(1 + r^{\{x,y\}}\right)^{-x+y} ,$$

where $r^{\{x,y\}}$ is the market's implied forward yield for the contract that accrues interest between dates x and y. If x is later than y, then the contract is to sell, and the market maker pays this yield, so that $r^{\{y,x\}} = r^{\{x,y\}}$. For each $x \in X$, the *market price* of a Zero that pays out one-dollar at date x is defined as $b(x) \overset{\text{def}}{=} a^{(x,0)}$.

Let $X \subseteq \mathbb{R}_+$ index a suitable sequence of trading dates with $0 \in X$ being the spot date. It is well known that a Zero that pays out one dollar at time $x > 0$ is arbitrage-free if, and only if, the log-accumulation process satisfies

$$\text{for every } x, y, z \in X, \quad \log a^{(x,y)} = \log a^{(x,z)} + \log a^{(z,y)} .^{13} \tag{6}$$

This no-arbitrage condition is a special case of the Jacobi identity.

[13]To see this, w.l.o.g. suppose that, for some $x < y < z$, $a^{(x,z)} < a^{(x,y)}a^{(y,z)}$. The market maker would do well to sell the contract (x, z) and buy the contracts (x, y) and (y, z). In the absence of counterparty risk, the difference between interest paid and received is risk-free.

We now explain how a market maker might infer her own subjective accumulation process from past cases. Our market maker thinks in terms of economic profits (relative to the market price). For every finite resampling D and every date x and y, let $x \preceq_D y$ if, and only if, at D, the market maker finds y (weakly) more plausible than x in answer to the question

> "Hold current market prices fixed and consider a one-dollar investment today. Given D, which maturity will yield a higher return?"

Remark (alternative interpretation)**.** *It is also possible to interpret $\preceq_{\mathbb{D}}$ in terms of statements of the market maker's intention to buy or sell forwards. Let (x, y) be shorthand for the forward contract where the buyer accumulates interest between dates x and y. For $x < y$, accumulating interest over (y, x) simply means selling (x, y). Holding current prices fixed, for each D in $\mathbb{D}$, we have $x \preceq_D y$ if, and only if, given D, the market maker would buy (x, y). Under this interpretation, if at $D^\star$ the market maker agrees with the market, then $x \simeq_{D^\star} y$ for every $x, y \in X$. Thus, if the market maker agrees with the market, then her sentiments are centered as in fig. 1. Our point is that the market maker may not agree with the market and thus the rankings map may be uncentered as in fig. 2. To operationalise this interpretation, simply suppress reference to $a^{(\cdot,\cdot)}$ in what follows.*

Next, we introduce the market maker's (possibly negative and subjective) markup function. This is a markup relative to current rates $r^{\{\cdot,\cdot\}}$. A *markup function* $\mu : X^2 \times \mathbb{C} \to \mathbb{R}$, is characterised by three conditions: for time intervals of length zero, the yield is zero; fair pricing; and case equivalence. These are, respectively, formalised as follows: for every $x, y \in X$ and every $c, d \in \mathbb{C}$, $\mu_c^{\{x,x\}} = 0$; $\mu_c^{\{y,x\}} = \mu_c^{\{x,y\}}$; and $c \sim^\star d$ if, and only if, $\mu_c^{\{x,y\}} = \mu_d^{\{x,y\}}$. We extend to $\mathbb{D}$ by taking $\mu_D^{\{x,y\}}$ to be the (geometric) *mean markup conditional on* D

$$1 + \mu_D^{\{x,y\}} = \prod_{c \in D} \left(1 + \mu_c^{\{x,y\}}\right)^{\frac{1}{|D|}} .$$

The market maker's subjective forward yield conditional on D is then the following modification of the market yield $r^{\{\cdot,\cdot\}}$: for every $x, y \in X$ and $D \in \mathbb{D}$

$$1 + \rho_D^{\{x,y\}} = \left(1 + r^{\{x,y\}}\right) \cdot \left(1 + \mu_D^{\{x,y\}}\right) .$$

In turn, the market maker's subjective accumulation process $A^{(\cdot,\cdot)} : X^2 \times \mathbb{D} \to \mathbb{R}$ modifies the market accumulation process $a^{(\cdot,\cdot)}$. For every $x, y \in X$ and $D \in \mathbb{D}$,

$$A_D^{(x,y)} := \left(1 + \rho_D^{\{x,y\}}\right)^{-x+y} = a^{(x,y)} \cdot \left(1 + \mu_D^{\{x,y\}}\right)^{-x+y}.$$

Note that the market maker agrees with the current market accumulation process whenever the cases in $D^\star$ are such that the positive markups countervail those that are negative. That is, for every $x, y \in X$, $\mu_{D^\star}^{(x,y)} = 0$.

The market maker's (subjective) empirical bond price function $B : X \times \mathbb{D} \to \mathbb{R}$ modifies the market price $b : X \to \mathbb{R}$. Thus, for a Zero with a one-dollar face value the price at time x, conditional on D is

$$B(x, D) := A_D^{(x,0)} = b(x) \cdot \left(1 + \mu_D^{\{0,x\}}\right)^{-x}.$$

This reflects the inverse relationship between bond prices and yields.

When $D^\star$ belongs to $\mathbb{D}$, the number of case types is finite, the following corollary of theorem 1 characterises $\measuredangle$-stability in the Zero-market setting.

Corollary 2. *Let $\preceq_{\mathbb{D}}$ satisfy* A4. *Then $\preceq_{\mathbb{D}}$ satisfies $\measuredangle$-*stability *if, and only if, there exists empirical implied yield and empirical bond price functions, such that*

$$\begin{cases} \textit{for every } x, y \in X \textit{ and every } D \in \mathbb{D}, \\ x \preceq_D y \quad \textit{if, and only if,} \quad B(x, D) \geqslant B(y, D). \end{cases} \tag{$*$}$$

Moreover, for every other empirical bond price function $\acute{B}$ that satisfies $(*)$, there exists a scalar $\lambda > 0$ such that $\log \acute{B} = \lambda \log B$; and, for every $D \in \mathbb{D}$, the associated accumulation process $A_D^{(\cdot,\cdot)}$ is arbitrage-free.

Proof of corollary 2. Via corollary 1, $\preceq_{\mathbb{D}}$ satisfies $\measuredangle$-stability if, and only if, there exists a pairwise Jacobi representation $v^{(\cdot,\cdot)} : X^2 \times \mathbb{C} \to \mathbb{R}$. Recalling that $0 \in X$, for every $x \in X$ and $D \in \mathbb{D}$, let

$$B(x, D) = \exp\left(-\tfrac{1}{|D|} \textstyle\sum_{c \in D} v_c^{(0,x)}\right). \tag{7}$$

For the proof of $(*)$, suppose that w.l.o.g., $x \preceq_D y$, so that $\sum_{c \in D} v_c^{(y,x)} \leqslant 0$. Via the Jacobi identity we have $\sum_{c \in D} \left\{v^{(y,0)} + v^{(0,x)}\right\} \leqslant 0$. Via the representation property $v^{(y,y)} = 0$. Another application of the Jacobi identity yields $v^{(y,0)} = -v^{(0,y)}$. Thus,

$$-\log B(x, D) + \log B(y, D) = \tfrac{1}{|D|} \textstyle\sum_{c \in D} \left\{v_c^{(0,x)} - v_c^{(0,y)}\right\} \leqslant 0.$$

We now show that the bond price is a suitable function of the empirical yield function. For every $x, y \in X$ and $D \in \mathbb{D}$, take $\log A_D^{(x,y)} = \frac{1}{|D|}\sum_{c\in D} v_c^{(x,y)}$. That is,

$$\log a^{(x,y)} + \tfrac{y-x}{|D|}\sum_{c\in D}\log\left(1+\mu_c^{\{x,y\}}\right) = \tfrac{1}{|D|}\sum_{c\in D} v_c^{(x,y)}.$$

Take $D = \{c\}$ and note that the Jacobi identity implies $v_c^{(x,y)} = -v_c^{(y,x)}$. Thus

$$\log a^{(x,y)} + (y-x)\log\left(1+\mu_c^{\{x,y\}}\right) = -\log a^{(y,x)} - (x-y)\log\left(1+\mu_c^{\{y,x\}}\right).$$

Moreover, note that $\log a^{(x,y)} = -\log a^{(y,x)}$, so that, by cancelling terms, we obtain

$$\log\left(1+\mu_c^{\{x,y\}}\right) = \log\left(1+\mu_c^{\{y,x\}}\right), \tag{8}$$

so that $\mu_c^{\{x,y\}} = \mu_c^{\{y,x\}}$. Then $v_c^{(x,x)} = 0 = \log a_c^{(x,x)}$ implies that $\mu_c^{\{x,x\}} = 0$. Finally, note that for $c \sim^\star d$, the property $\mu_c^{\{x,y\}} = \mu_d^{\{x,y\}}$ is inherited from $v_c^{(x,y)} = v_d^{(x,y)}$, so that we have an empirical implied yield function. All of the above arguments are reversible, so that, given the existence of such empirical yield and bond price functions satisfying $(*)$, it follows that $\leq_{\mathbb{D}}$ satisfies *4*-stability.

If $\acute{B}$ is another empirical bond price function that satisfies $(*)$, then via eq. (7) and corollary 1, for some $\lambda > 0$, $\log\acute{B} = \lambda\log B$. The fact that, for every D, $A_D^{(\cdot,\cdot)}$ is arbitrage-free follows by virtue of the fact that the Jacobi identity holds element-wise for $v^{(\cdot,\cdot)}$. In particular, since, for every $x, y, z \in X$ and $c \in D$, $\log A_c^{(x,y)} = v_c^{(x,y)}$,

$$\log A_D^{(x,y)} = \tfrac{1}{|D|}\sum_{c\in D}\log A_c^{(x,y)} = \tfrac{1}{|D|}\sum_{c\in D}\left\{\log A_c^{(x,z)} + \log A_c^{(z,y)}\right\}.$$

Taking exponents, we obtain the no-arbitrage condition $A_D^{(x,y)} = A_D^{(x,z)}\cdot A_D^{(z,y)}$. $\square$

***prop-fourdiv-empty here ***

The uniqueness result in corollary 2 is stronger than in the general setting of part II of theorem 1. This is a consequence of the fact that our empirical bond yield satisfies the property $r^{(x,x)} = 0 = \mu^{(x,x)}$, for every $x \in X$.

Our diversity axiom, A4**,** has a straightforward interpretation in the present setting. Given *4*-stability, *2*-diversity implies that, for every distinct x and y, there exist c and d such that $v_c^{(x,y)} < 0 < v_d^{(x,y)}$. This, via the arguments that lead to eq. (8), is equivalent to both $\mu_d^{\{x,y\}} < 0 < \mu_c^{\{x,y\}}$ and $\rho_c^{\{x,y\}} < r^{\{x,y\}} < \rho_d^{\{x,y\}}$. In words, *2*-diversity requires that the market maker's data is rich enough to contain at least one case where her markup between date x and y is positive, and at least one where it is negative. Conditional-*2*-diversity generalizes this notion to require that for every distinct x, y, z, $\mu_C^{\{x,y\}} < 0 < \mu_D^{\{x,y\}}$ for some C and D such that $\mu_C^{\{x,z\}} \cdot \mu_D^{\{x,z\}} > 0$. An equivalent characterisation is also available in terms of the affine independence condition (1.b) of theorem 1 for the matrix $\beta : X \times \mathbb{C} \to \mathbb{R}$, $x \times c \mapsto \beta(x, c) = B(x, \{c\})$.

The considerably stronger conditional-*3*-diversity arises (implicitly) in the final step of our proof of theorem 3. When it holds for $\preceq_{\mathbb{D}}$, stability is unnecessary (provided A0–A3 hold). In terms of mark ups, we may characterise conditional-*3*-diversity as, for every distinct x, y, z and w, one of the half spaces $\{D : \mu_D^{\{x,w\}} > 0\}$ or $\{D : \mu_D^{\{x,w\}} < 0\}$ contains $D_1, \ldots, D_6$ and permutations $\pi_1, \ldots, \pi_6$ of x, y and z such that for $i = 1, \ldots, 6$,

$$\mu_{D_i}^{\{\pi_i(x),0\}} < \mu_{D_i}^{\{\pi_i(y),0\}} < \mu_{D_i}^{\{\pi_i(z),0\}}.$$

Yet *4*-diversity is stronger still, requiring the above to hold on both half spaces.

This disparity between conditional-*2*-diversity and *4*-diversity reflects the value of experience or of rich data. It also reflects the value of a prudent imagination when the data fails to be rich. Outside of market settings, this may be as far as we can go, but in the present context we can say more.

Market efficiency We begin with a proposition that justifies our claim that when *4*-stability holds, so will the efficient markets hypothesis in the usual (**?**) sense: passive (buy-and-hold) strategies outperform active ones. To this end, let $\beta : X \times \mathbb{D}^{\mathfrak{f}} \to \mathbb{R}_+$ be an empirical bond price function such that its restriction to $X \times \mathbb{D}$ coincides with B of corollary 2 and let

$$g_c(x, D) := \frac{\beta(x, D \cup \{c\}) - \beta(x, D)}{\beta(x, D)}$$

denote the *proportional price increment* for the maturity date x given the data D, following the arrival of a new case c. By new case, we mean that c can be regular or novel. Under stable pricing, the proportional price increment converges to zero exponentially.

Proposition 1. *For every $x \in X$, $D \in \mathbb{D}$ of cardinality n and $c \in \mathbb{C}^{\mathfrak{f}} - D$,*

$$1 + g_c(x, D) = \left(\left(1 + \rho_c^{\{0,x\}}\right)^{-x} / B(x, D)\right)^{\frac{1}{n+1}} \leqslant (1 + \rho)^{\frac{2x}{n+1}}$$

where $\rho = \operatorname{argmax}\left\{\left(1 + \rho_d^{\{0,x\}}\right)^{-x} : d \in D \cup \{c\}\right\}$.

Proof. Let $1, \ldots, n$ be an enumeration of D and identify c with $n+1$. Moreover, let us suppress reference to $\{0, x\}$, so that $\rho_{n+1} = \rho_c^{\{0,x\}}$. We first derive the equality. By corollary 2, and the fact that β agrees with B on $X \times \mathbb{D}$, we obtain

$$\beta(x, D) = B(x, D) = \prod_i^n (1 + \rho_i)^{\frac{-x}{n}}.$$

and, by manipulating exponents, we obtain

$$\beta(x, D \cup \{c\}) = \left(\prod_i^{n+1} (1 + \rho_i)^{\frac{-x}{n}}\right)^{\frac{n}{n+1}}.$$

Substituting for $B(x, D)$ and noting that $\frac{n}{n+1} = 1 - \frac{1}{n+1}$, we arrive at the expression

$$1 + g_c(x, D) = \frac{B(x, D)^{1 - \frac{1}{n+1}} \cdot (1 + \rho_{n+1})^{\frac{-x}{n+1}}}{B(x, D)}.$$

For the inequality, note that $B(x, D)^{-1} = \exp(\frac{1}{n} \sum_1^n \nu_i)$ where $\nu_i = x \log(1 + \rho_i)$. Let $\nu_{n+1} = x \log(1 + \rho_{n+1})$. Then, for $\nu := \max\{|\nu_i| : i = 1, \ldots, n+1\}$, we have

$$\frac{1}{n} \sum_1^n \nu_i \leqslant \left|\frac{1}{n} \sum_1^n \nu_i\right| \leqslant \frac{1}{n} \sum_1^n |\nu_i| \leqslant \nu = x \log(1 + \rho).$$

Then, since exp is an increasing function,

$$B(x, D)^{-1} \leqslant \exp(\nu) = (1 + \rho)^x$$

Mutatis mutandis, the same argument yields $(1 + \rho_{n+1})^{-x} \leqslant (1 + \rho)^x$. Extending this pair of bounds extend to the product brings us to the desired inequality. □

Key to the proof of proposition 1 is that the new pricing kernel β coincides with the old one B on $X \times \mathbb{D}$. Proposition 1 thus demonstrates that $\mathcal{4}$-stability implies the usual notion of stability of statistical learning (**???**). That is to say it implies continuity of the learning map (here induced by the ranking map) $L : \bigsqcup_{i \geqslant 1} Z^n \to \mathcal{H}$ from the sample space to the hypothesis space (**?**).

The following corollary of proposition 1 confirms that, for stable pricing kernels, the impact of the arrival of a given case type is decreasing in its frequency. This is a decreasing marginal impact of information property: the information carried by a common case type is already baked into the price. We discuss the impact of this result on market efficiency following the statement and proof.

Corollary 3. *For $D \in \mathbb{D}$ of cardinality n, let $\mathbb{T}_D$ be the set of case types that feature in D and let $c, d \in \mathbb{C} - D$ be cases of type $s \in \mathbb{T}_D$. Then, for every $x \in X$,*

$$|g_d(x, D \cup \{c\})| < |g_c(x, D)|$$

and, moreover, if n_s is the frequency of s in D, then $g_c(x, D)$ tends to 0 as $n_s \to n$.

Proof. Take $\gamma_c = g_c(x, D)$ and $\gamma_d = g_d(x, D \cup \{c\})$. For any c' of type t, let $\rho_t = \rho_{c'}^{\{0,x\}}$. We claim that

$$1 + \gamma_c = \left((1 + \rho_s)^{n - n_s} / \prod_{t \neq s} (1 + \rho_t)^{n_t} \right)^{\frac{-x}{n(n+1)}}. \tag{9}$$

First note that, since, for each $t \in \mathbb{T}_D$, D contains $n_t > 0$ cases of type t, we have

$$B(x, D) = \left(\prod_{t \in \mathbb{T}_D} (1 + \rho_t)^{n_t} \right)^{\frac{-x}{n}} = \left(\prod_{t \neq s} (1 + \rho_t)^{n_t} \cdot (1 + \rho_s)^{n_s} \right)^{\frac{-x}{n}}$$

Then a substitution for $B(x, D)$ in the expression for $1 + \gamma_c$ of proposition 1 followed by a straightforward manipulation of exponents yields eq. (9).

Let θ denote the main ratio (inside the brackets) of eq. (9). For convergence of γ_c to 0 note that $n_s \to n$ implies that, for every $t \neq s$, $n_t \to 0$. Thus both the numerator and the denominator of θ tend to one and $\gamma_c \to 0$.

For monotonicity, since d also belongs to s, we extend eq. (9) to obtain

$$1 + \gamma_d = \left((1 + \rho_s)^{(n+1) - (n_s + 1)} / \prod_{t \neq s} (1 + \rho_t)^{n_t} \right)^{\frac{-x}{(n+1)(n+2)}}. \tag{10}$$

Note that the exponent of the main numerator in eq. (10) simplifies to $n - n_s$. Thus, the main ratio in eq. (10) is also equal to θ. In both eq. (9) and eq. (10), the exponent involving x is negative, so that, via the connection with θ,

$$\gamma_c < 0 \quad \text{iff} \quad \theta > 1 \quad \text{iff} \quad \gamma_d < 0. \tag{11}$$

Combining eq. (9) and eq. (10), substituting θ and simplifying the exponent:

$$\frac{1+\gamma_d}{1+\gamma_c} = \theta^{2x\cdot\frac{n+1}{k}}, \tag{12}$$

where $k = n(n+1)^2(n+2)$. Since the exponent here is positive, we arrive at

$$\frac{1+\gamma_d}{1+\gamma_c} < 1 \quad \text{iff} \quad \theta < 1. \tag{13}$$

Via eq. (11), there are two cases to consider. The most obvious is $\gamma_d, \gamma_c > 0$, for then $\theta < 1$ and $\gamma_d < \gamma_c$ follows from eq. (13). The remaining case is $\gamma_c, \gamma_d < 0$ so that, via eq. (11), $\theta > 1$. Then, since $1 + \gamma_d = 1 - |\gamma_d|$ (and, mutatis mutandis, the same is true of γ_c), via eq. (13), we arrive at

$$\frac{1-|\gamma_d|}{1-|\gamma_c|} > 1 \quad \text{iff} \quad \theta > 1.$$

A simple rearrangement then yields the desired inequality. □

In terms of market efficiency, implies that, if we assume that, more often than not, the future resembles the past in the sense that past cases with a higher frequency are more likely to continue to arrive more frequently, corollary 3 means that the most likely cases are crowded out by the volume other cases of the same type. When the pricing kernel is stable, the only way an active strategy can exploit the information in past cases is to more accurately predict the type of the case. But, since the marginal return to the most frequent cases is diminishing, this will be an uphill struggle.

Yet corollary 3 goes further. It tells us that there is value in diversity, value to the information discovery process, for novel case types are more rewarding. This explains why financial institutions spend vast resources on research. When the price kernel is stable, the largest price movements come out of "left field". This frequentist idea of market efficiency allows research and active strategies to co-exist with passive buy-and-hold. Most of the time buy-and-hold is likely to be better, but if one can also discover the nature of less frequent, novel cases prior to their arrival, one can improve by actively researching the novel case generation process. In this way, an active imagination and hard work can help an entrepreneurial market maker to go beyond prudence. (And also beyond the formal scope of this paper.)

With unstable pricing kernels, the story is very different, whilst the yield accumulation process $A = \{A_D^{(x,y)} : x, y \in X, D \in \mathbb{D}\}$ is well-defined, since the groupoid property fails to hold, the pricing kernel is not. Whilst it is mathematically possible to derive equivalent statements to proposition 1 and **??** for A, it is somewhat meaningless in the presence of the resulting arbitrage opportunities that are present and likely to cause instability to A. Instead, the value of A is that we can study out-of-equilibrium market activity in the absence of a well-defined pricing kernel (in the sense of **??**). To our knowledge, this feature is absent in the literature.

?, p.60 points out that "Markets can be efficient even if many market participants are quite irrational." Here, we can show that markets can be efficient, in the sense that the pricing kernel is stable (and thus acting as if it were guided by a prudent market maker), even if no agent is prudent. Given the above discussion, this possibility should be clear when the data is rich. But what if the data fails to be *4*-diverse?

Throughout the following, we suppose that conditional-*2*-diversity holds. For in its absence, uniqueness and existence are not guaranteed.

Recall that in our derivation above, we only assumed the market's accumulation and bond price were available conditional on the current set $D^\star$. Suppose the analyst is able to extend the market accumulation process and condition on any finite sample and, moreover, is able to identify an arbitrage-free bond price $b : X \times \mathbb{D} \to \mathbb{R}$. (Below we provide a rudimentary algorithm for this generalization.) Then our results say that the market bond price induces a rankings map $\leq_{\mathbb{D}}$ that satisfies *4*-stability. Prudent pricing emerges from market activity even in the absence of prudent agents.

This inductive notion of market efficiency differs from the more Bayesian and forward-looking definitions of **?** and **?**. It is not a form of weak efficiency, for we may define cases to include more than historical data on prices. We may include features such as firm size or price-earnings data or indeed any of the many other factors that are extensively discussed in the literature (**???**).

In our view, the fact that, in market settings, stable pricing can arise even in the absence of both rich data (*4*-diversity) and prudent agents is a tribute to the power of markets: it is an instance of Adam Smith's invisible hand at its best. While the threat of arbitrage may supplant the imagination and provide an alternative form of external validation, stability also provides a way to identify or "memorize" past cases. This is a more operational, non-behavioural form of efficiency.

Structural breaks and second-order induction. Of course some new types of cases might for good reason require a re-evaluation of past ones. Such cases lead to the formation of a new rankings map $\grave{\precsim}_{\mathbb{D}}$. We would expect such cases to be less common than other forms of novel case since they represent structural breaks or regime changes. Agents that are sufficiently imaginative or privately informed about such cases may be able to profit from the associated upheaval once they arrive. But novel cases that are "independent" of past cases (in the sense that they do not cause a re-evaluation) should not generate arbitrage opportunities. This is the essence of stability.

Algorithm for identifying the pricing kernel from past cases Under stable pricing, cases combine in a modular way, markets assimilate the information in new types of cases without re-pricing the information of old case. This is the role of the nonrevision property of generalizations (see definition 1). Price movements reflect the marginal information of new cases.

We now describe a classification method for deciding the nature of a new case and the implications for market maker(s) who themselves act as analysts. Consider the arrival of j new cases $c_1, \ldots, c_j$. Under stable pricing, the analyst takes new price movements to reflect the market view on the information value of new cases. That is, for each $i = 1, \ldots, j$, the analyst estimates the markup associated with case c_i to be $\hat{\mu}^{\{x,y\}} = b(x, D^\star \cup \{c_i\}) - b(x, D^\star)$. This is reasonable provided she already has used the history of prices to estimate $b : X \times \mathbb{D} \to \mathbb{R}$. Moreover suppose that, for each $i = 1, \ldots, k-1$, there exists $d \in D^\star$ such that for every $x, y \in X$, $\hat{\mu}^{\{x,y\}}_{c_i} = \mu^{\{x,y\}}_d$. She concludes that each of the cases $c_1, \ldots, c_{k-1}$ is a "copy" of some past case type. Then $D^\star \cup \{c_1, \ldots, c_{k-1}\}$ belongs to $\mathbb{D}$ and the analyst's market model is unchanged: all that has changed is the location of the current sample $D^\star$ in $\mathbb{D}$.

Case c_k is different. The analyst finds that price movements are such that, for every $d \in D^\star$ (and hence every case in $\mathbb{C}$), $\hat{\mu}_{c_k}^{\{\cdot,\cdot\}} \neq \mu_d^{\{\cdot,\cdot\}}$. This reveals c_k to be a new type of case. The analyst then generalizes her model of the market to let $\grave{\mathbb{C}}$ and $\grave{\mathbb{D}}$ to respectively be the new sets of all cases and finite resamplings that she can generate from $\grave{D} = D^\star \cup \{c_k\}$. Assuming no-arbitrage/prudent pricing, there no need for her to update $\mu_d^{\{\cdot,\cdot\}}$ for $d \in D^\star$. That is, the new market pricing function $\grave{b} : X \times \grave{\mathbb{D}} \to \mathbb{R}$ satisfies $\grave{b}(\cdot, D) = b(\cdot, D)$ for every $D \in \mathbb{D}$. It is in this sense that markets deliver efficiency via prudent pricing.

Now consider two possibilities for $c_{k+1}, \ldots, c_j$. The first is that the price movements associated with all of these cases are similar to cases in $\grave{\mathbb{D}}$. That is, for each $i = k+1, \ldots, j$, there exists $d \in \grave{D}$ such that $\hat{\mu}_{c_i}^{\{\cdot,\cdot\}} = \grave{\mu}_d^{\{\cdot,\cdot\}}$. This is what we would expect if new case types arrive infrequently. It would represent a consolidation of her updated model $\grave{b}$ and $\grave{\preceq}_{\grave{\mathbb{D}}}$.

The second possibility is where many of the cases $c_k, \ldots, c_j$ turn out to be novel. If new case types do indeed arrive infrequently, then it is likely to reflect a re-evaluation of cases in $D^\star$ and a structural break from the past. In this scenario, it may well be worth checking to see if a re-evaluation or even re-specification of past cases reduces the number of case types: thus generating a more parsimonious model.

5 Discussion of the axioms and theorem 1

We begin by restating the existence part of the main theorem of GS03.

Theorem (GS03, existence)**.** *Let there be given X, $\mathbb{C}$ and $\preceq_{\mathbb{D}}$, as above, such that the richness condition holds. Then* (i) *and* (ii) *are equivalent.*

(i) A0–A3 *and 4*-diversity *hold for $\preceq_{\mathbb{D}}$.*

(ii) *There exists a matrix $\mathbf{v} : X \times \mathbb{C} \to \mathbb{R}$ satisfying* (a) *and* (b)*:*

  (a) $\mathbf{v}$ *is a representation of $\preceq_{\mathbb{D}}$ that respects case equivalence*;

  (b) *For every four distinct elements $x, y, z, w \in X$ and every $\lambda, \mu, \theta \in \mathbb{R}$ such that $\lambda + \mu + \theta = 1$, $\mathbf{v}(x, \cdot) \nleq \lambda\mathbf{v}(y, \cdot) + \mu\mathbf{v}(z, \cdot) + \theta\mathbf{v}(w, \cdot)$. For $\#X < 4$, no row is dominated by an affine combination of other rows.*

Although diversity axioms play an important technical role, they are not obviously behavioral. Instead, diversity axioms impose restrictions on what is beyond the learner's control and on what is central to inductive inference: experience. We contend that $D^\star$ may not be so rich as to support $\preceq_{\mathbb{D}}$ satisfying *4*-diversity. That is to say, there may exist $Y \subseteq X$ such that $\#Y = 4$, and such that the data is insufficiently rich to support all $4! = 24$ strict rankings.

It is natural to ask whether *4*-stability is simply requiring that, on Y such that $\preceq_{\mathbb{J}}$ fails to satisfy *4*-diversity, there exists a testworthy Y-generalization that is novel and satisfies *4*-diversity. This is not the case. When the basic axioms hold, *4*-diversity implies that the current ranking map features a centered arrangement of hyperplanes (for every $Y \subseteq X$ of cardinality four). In contrast, the axioms of theorem 1 do not even imply the existence of a centered generalization. (See for instance fig. 3 of section E.) It is well-known in the literature on hyperplane arrangements that numerous complexities arise when the arrangement is uncentered.

A casual comparison of condition (1.b) and (b) confirms that the present framework accommodates the less experienced or equivalently those settings where the data is not rich. By doing so, we have identified an important role for the imagination and in particular, a prudent imagination. Our results show how inexperienced learners that are prudent can survive the initial phase before going on to become experienced learners in their own right. We have established that there is more than one kind of stable learner. That is to say, whereas experienced (*4*-diversity) learners are stable, so are those that prudently appeal to external validation: either via their imagination or via leave-one-type-out cross validation.

We have shown that, even in the absence of prudent learners, provided the market structure allows agents to exploit arbitrage opportunities, market pricing is prudent. By this we mean that it is as if a prudent market maker were guiding the price formation process. This novel form of efficiency is grounded in inductive inference. Moreover, it implies a modular nature to the way information is built into market prices. It implies stability in the value of information in past cases.

Appendix A. The proof of theorem 1

Similar to GS03, we translate the model into one where databases are represented by vectors, the dimensions of which are case types. To allow us to focus on aspects of the present model, proceed directly to rational vectors and present the axioms and a corresponding theorem (theorem 2) which, as we confirm, holds if, and only if, theorem 1 does. The proof of theorem 2 can be found in section B.

Case types as dimensions. From our definition of case types in section 2, $\mathbb{T} = \mathbb{C}_{/\sim^\star}$ and $\mathbb{T}^{\mathfrak{f}} \stackrel{\text{def}}{=} \mathbb{T} \cup [\mathfrak{f}]$. Let $\mathfrak{T}$ be a free variable in $\{\mathbb{T}, \mathbb{T}^{\mathfrak{f}}\}$. When no possible confusion should arise, we use $\mathfrak{f}$ as shorthand for $[\mathfrak{f}]$. It is straightforward to show that the following construction would work if instead we were to work with any partition T of $\mathbb{C}$ that is at least as fine as $\mathbb{T}$. The present construction is the one with the lowest feasible number $\#\mathbb{T}$ of dimensions.

Translation to counting vectors. Let $\mathbb{Z}_+$ denote the set of nonnegative integers and $\mathbb{Z}_{++}$ those that are (strictly) positive. Let $\mathbb{L} \subseteq \mathbb{Z}_+^{\mathbb{T}}$ denote the set of counting vectors $L : \mathbb{T} \to \mathbb{Z}_+$ such that $\{t : L(t) \neq 0\}$ is finite and let $\mathbb{L}^{\mathfrak{f}}$ denote the corresponding subset of $\mathbb{Z}_+^{\mathbb{T}^{\mathfrak{f}}}$. Then let

$$\mathfrak{L} = \begin{cases} \mathbb{L} & \text{if, and only if, } \mathfrak{T} = \mathbb{T}, \text{ and} \\ \mathbb{L}^{\mathfrak{f}} & \text{otherwise.} \end{cases}$$

Modulo notation, the following construction is identical to GS03. For every $D \in \mathbb{D}$, let $L_D : \mathbb{T} \to \mathbb{Z}_+$ denote the function $t \mapsto L_D(t) = \#(D \cap t)$. For each $D \in \mathbb{D}$, let $\preceq_{L_D} \stackrel{\text{def}}{=} \preceq_D$. We need to establish that $\preceq_{\mathbb{L}} \stackrel{\text{def}}{=} \langle \preceq_L : L \in \mathbb{L} \rangle$ is well-defined. For every $L \in \mathbb{L}$, the richness assumption (on $\mathbb{T}^{\mathfrak{f}}$) guarantees the existence of $D \in \mathbb{D}$ such that $L_D = L$. By definition, $\sim^\star$ is such that, for every $C, D \in \mathbb{D}$, $C \sim^\star D$ if, and only if, $L_C = L_D$. Straightforward mathematical induction on the cardinality of C shows that $C \sim^\star D$ implies $\preceq_C = \preceq_D$. This construction of $\preceq_{\mathbb{L}}$ ensures that the same notion of equivalence that we introduced in observation 1 also applies here. Thus, $\preceq_{\mathbb{L}} \equiv \preceq_{\mathbb{D}}$.

Translation to rational vectors. Similarly, let $\mathbb{Q}_+$ denote the nonnegative rationals and $\mathbb{Q}_{++}$ those that are (strictly) positive. Take $\mathbb{J} \subseteq \mathbb{Q}_+^{\mathbb{T}}$ to be the set of rational vectors with $\{t \in \mathbb{T} : J(t) \neq 0\}$ finite and take $\mathbb{J}^{\mathfrak{f}}$ to denote the corresponding subset of $\mathbb{Q}_{+}^{\mathbb{T}^{\mathfrak{f}}}$. For each $J \in \mathbb{J}$, by virtue of the fact that $\mathbb{Z}_{++}$ is well-ordered and J has finite support, there exists (unique) minimal $k_J \in \mathbb{Z}_{++}$ such that $L_J \stackrel{\text{def}}{=} k_J J$ belongs to $\mathbb{L}$. Let $\leq_J \stackrel{\text{def}}{=} \leq_{L_J}$. (This definition acquires meaning below once we translate and apply the combination axiom.) In this way, $\leq_{\mathbb{J}} = \langle \leq_J : J \in \mathbb{J} \rangle$ is well-defined, and we may introduce axioms for $\leq_{\mathbb{J}}$ directly: i.e. without first introducing axioms for $\leq_{\mathbb{L}}$. We first demonstrate that $\leq_{\mathbb{J}}$ and $\leq_{\mathbb{D}}$ are equivalent. First note that, for every $I, J \in \mathbb{J}$ such that $L_I = L_J$, $\leq_I = \leq_J$. Then, let $L' = L_J$ and take any D such that $L_D = L'$. Then $\leq_J = \leq_D$. The reverse embedding follows by virtue of the fact that $\mathbb{L} \subset \mathbb{J}$. Thus, $\leq_{\mathbb{J}} \equiv \leq_{\mathbb{D}}$.

Construction of generalizations of $\leq_{\mathbb{J}}$. We follow common practice by letting 2^X denote the collection of nonempty subsets $Y \subseteq X$. For each $Y \in 2^X$, we will denote the set of regular, novel and testworthy Y-generalizations (of $\leq_{\mathbb{D}}$ or $\leq_{\mathbb{J}}$) by $\text{reg}(Y, \cdot)$, $\text{nov}(Y, \cdot)$ and $\text{test}(Y, \cdot)$ respectively. Recalling that every Y-generalization is either regular or novel, let $\text{ext}(Y, \cdot)$ denote the set of all Y-generalizations. We now clarify what it means to be a generalization of $\leq_{\mathbb{J}}$.

For each $t \in \mathfrak{T}$, we take $\delta_t : \mathfrak{T} \to \mathbb{R}$ to be the function satisfying $\delta_t(s) = 1$ if $s = t$ and $\delta_t(s) = 0$ otherwise. (When $\mathfrak{T}$ is finite, these are simply the basis vectors for $\mathbb{R}^{\mathfrak{T}}$.) When we wish to emphasise that the vectors belong to in $\mathbb{R}^{\mathbb{T}^{\mathfrak{f}}}$, then, for each $\mathbb{T}^{\mathfrak{f}}$, we will write $\delta_t^{\mathfrak{f}}$. Let

$$\mathfrak{J} = \begin{cases} \mathbb{J} & \text{if, and only if, } \mathfrak{T} = \mathbb{T}, \text{ and} \\ \mathbb{J}^{\mathfrak{f}} & \text{otherwise.} \end{cases}$$

For every $I \in \mathbb{J}$ and $J \in \mathfrak{J}$, we write $I \equiv J$ whenever $I = J$ or $J = I \times 0$. (In the latter case, $J(t) = I(t)$ for every $t \in \mathbb{T}$ and $J(\mathfrak{f}) = 0$.) This notion reflects the fact that, for the purposes of the present model, such I and J are equivalent.

Definition 2. *$\mathcal{R} = \langle \mathcal{R}_J : J \in \mathfrak{J} \rangle$ is a* generalization *or a* Y-generalization *of* $\leq_{\mathbb{J}}$ *if, and only if, for some nonempty $Y \subseteq X$ both the following hold*

1. *for every $J \in \mathfrak{J}$, $\mathcal{R}_J \in \text{rel}(Y)$, $\mathcal{I}_J \stackrel{\text{def}}{=} \mathcal{R}_J \cap \mathcal{R}_J^{-1}$ and $\mathcal{P}_J \stackrel{\text{def}}{=} \mathcal{R}_J - \mathcal{R}_J^{-1}$;*

2. *for every $J \in \mathbb{J}$ and $L \in \mathfrak{J}$ such that $J \equiv L$, $\mathcal{R}_L = \leq_J \cap (Y^2)$.*

A generalization $\mathcal{R}_{\mathfrak{I}}$ (of $\preceq_{\mathbb{J}}$) is proper if $\mathfrak{I} = \mathbb{J}^{\mathfrak{f}}$ and improper otherwise. A proper generalization is either regular or novel. $\mathcal{R}$ is novel if, for every $s \in \mathbb{T}$, there exists I in $\mathbb{J}$ such that, for $J = I \times 0$ (in $\mathbb{J}^{\mathfrak{f}}$), we have $\mathcal{R}_{J+\delta^{\mathfrak{f}}_s} \neq \mathcal{R}_{J+\delta^{\mathfrak{f}}_{\mathfrak{f}}}$.

For every regular Y-generalization $\mathcal{R}$ of $\preceq_{\mathbb{D}}$ such that $Y = X$, observation 1 implies $\mathcal{R} \equiv \preceq_{\mathbb{D}}$. And, via $\preceq_{\mathbb{J}} \equiv \preceq_{\mathbb{D}}$ and transitivity of equivalence, we conclude that $\mathcal{R}$ is equivalent to $\preceq_{\mathbb{J}}$. Two sets of generalizations are isomorphic if there exists a canonical isomorphism between equivalent generalizations.

Lemma 1.1 (proof on page 45)**.** *For every $Y \in 2^X$, $\mathrm{reg}(Y, \preceq_{\mathbb{J}})$ is isomorphic to $\mathrm{reg}(Y, \preceq_{\mathbb{D}})$ and $\mathrm{nov}(Y, \preceq_{\mathbb{J}})$ is isomorphic to $\mathrm{nov}(Y, \preceq_{\mathbb{D}})$.*

Axioms and theorem. We restate the axioms for Y-generalizations $\mathcal{R}$ of $\preceq_{\mathbb{J}}$.

A0$^\flat$ For every $J \in \mathfrak{I}$, $\mathcal{R}_J$ is transitive on Y.

A1$^\flat$ For every $J \in \mathfrak{I}$, $\mathcal{R}_J$ complete on Y.

A2$^\flat$ For every $I, J \in \mathfrak{I}$, every $x, y \in Y$ and every $\lambda, \mu \in \mathbb{Q}_{++}$, if $x \mathrel{\mathcal{R}_I} y$ and $x \mathrel{\mathcal{R}_J} y$, then $x \mathrel{\mathcal{R}_{\lambda I+\mu J}} y$; moreover, if $x \mathrel{\mathcal{P}_I} y$ and $x \mathrel{\mathcal{R}_J} y$, then $x \mathrel{\mathcal{P}_{\lambda I+\mu J}}$.

A3$^\flat$ For every $I, J \in \mathfrak{I}$ and every $x, y \in Y$ if $x \mathrel{\mathcal{P}_J} y$, then there exists $0 < \lambda < 1$ such that, for every $\mu \in \mathbb{Q} \cap (\lambda, 1)$, $x \mathrel{\mathcal{P}_{(1-\mu)I+\mu J}} y$.

For $k = 2, 3, 4$, k-diversity is defined for generalizations of $\preceq_{\mathbb{J}}$ in exactly the same way. We continue to use the term k-diversity in this setting. The following are conditional-*2*-diversity and partial-*3*-diversity respectively.

A4$^\flat$ For every three distinct elements $x, y, z \in Y$, one of the two subsets $\{J' : x \prec_{J'} y\}$ and $\{J' : y \prec_{J'} x\}$ of $\mathbb{J}$ contains both I and J such that $z \prec_I x$ and $x \prec_J z$. If $\#Y = 2$, then *2*-diversity holds on Y.

A4$'^\flat$ For every $Y' \subseteq Y$ with cardinality $n = 2$ or 3, every Y'-generalization $\mathcal{R}$ of $\preceq_{\mathbb{J}}$ is such that $\#\mathrm{total}(\mathcal{R}) \geqslant n$.

A proper generalization $\mathcal{R}$ of $\preceq_{\mathbb{J}}$ is *testworthy* if it satisfies A1$^\flat$–A3$^\flat$ and, for some $J \in \mathbb{J}$ such that $\mathcal{R}_{J\times 0}$ is total, $\mathcal{R}_{\mathfrak{f}} = \mathcal{R}^{-1}_{J\times 0}$. Thus, for each $Y \in 2^X$, $\mathrm{test}(Y, \preceq_{\mathbb{D}}) \simeq \mathrm{test}(Y, \preceq_{\mathbb{J}})$. For any pair of generalizations $\mathcal{R}$ and $\hat{\mathcal{R}}$, $\hat{\mathcal{R}}$ is a perturbation of $\mathcal{R}$ if $\mathcal{R}_{\mathfrak{f}} = \hat{\mathcal{R}}_{\mathfrak{f}}$. Moreover, $\hat{\mathcal{R}}$ is a diverse perturbation if $\#\mathrm{total}(\hat{\mathcal{R}}) \leqslant \#\mathrm{total}(\mathcal{R})$.

4-P$^\flat$ For every $Y \subseteq X$ of cardinality 3 or 4, every testworthy Y-generalization of $\preceq_{\mathbb{J}}$ that is novel has a diverse perturbation that satisfies A0$^\flat$–A3$^\flat$.

The following result corresponds to claim 2 of GS03. Its proof is a consequence of mathematical induction and the combination axiom.

Lemma 1.2. *If $\mathcal{R}_{\mathfrak{J}}$ and $\hat{\mathcal{R}}_{\mathfrak{L}}$ are equivalent and the latter satisfies* A2*, then for every $J \in \mathfrak{J}$ and every rational number $q > 0$, we have $\mathcal{R}_{qJ} = \mathcal{R}_J$.*

The fact that $\preceq_{\mathbb{J}} \equiv \preceq_{\mathbb{D}}$ immediately implies that $\preceq_{\mathbb{J}}$ satisfies A0$^\flat$, A1$^\flat$ and A4$^\flat$ if, and only if, the corresponding axiom holds for $\preceq_{\mathbb{D}}$. In general, we have the following result, which then also yields the equivalence for the stability axiom.

Lemma 1.3 (proof on page 46)**.** *For $\mathcal{R}_{\mathfrak{J}} \equiv \hat{\mathcal{R}}_{\mathfrak{D}}$, $\mathcal{R}_{\mathfrak{J}}$ satisfies* A2$^\flat$–A3$^\flat$ *if, and only if, $\hat{\mathcal{R}}_{\mathfrak{D}}$ satisfies* A2–A3.

The matrix $\mathbf{v} : X \times \mathbb{T} \to R$ is a *representation of* $\preceq_{\mathbb{J}}$ whenever it satisfies

$$\begin{cases} \text{for every } x, y \in X \text{ and every } J \in \mathbb{J}, \\ x \preceq_J y \quad \text{if, and only if,} \quad \sum_{t \in \mathbb{T}} \mathbf{v}(x,t)J(t) \leqslant \sum_{t \in \mathbb{T}} \mathbf{v}(y,t)J(t). \end{cases} \tag{$\flat$}$$

We observe that, via the definition of case types, there exists a representation of $\preceq_{\mathbb{D}}$ that respects case equivalence if, and only if there exists a representation of $\preceq_{\mathbb{J}}$. The above translation and results imply that theorem 1 is equivalent to

Theorem 2. *Let there be given X, $\mathbb{T}^{\mathfrak{f}}$, $\preceq_{\mathbb{J}}$ and associated generalizations, as above. Then* (2.i) *and* (2.ii) *are equivalent.*

(2.i) A0$^\flat$–A4$^\flat$ *and* *4*-P$^\flat$ *hold for $\preceq_{\mathbb{J}}$ on X.*

(2.ii) *There exists a matrix $\mathbf{v} : X \times \mathbb{T} \to \mathbb{R}$ that satisfies both:*

(2.a) $\mathbf{v}$ *is a representation of $\preceq_{\mathbb{J}}$; and*

(2.b) *no row of $\mathbf{v}$ is dominated by any other row, and, for every three distinct elements $x, y, z \in X$, $\mathbf{v}(x,\cdot) - \mathbf{v}(z,\cdot)$ and $\mathbf{v}(y,\cdot) - \mathbf{v}(z,\cdot)$ are noncollinear (i.e. linearly independent).*

Moreover, $\mathbf{v}$ *is unique in the sense of* theorem 1 *part II, with* (2.ii) *replacing* (1.ii) *and $\mathbb{T}$ replacing $\mathbb{C}$.*

Appendix B. The proof of theorem 2

STEP B.1 (characterisation of A1$^\flat$–A3$^\flat$, *2*-diversity and novel generalizations). The following proposition corresponds to lemma 1 of GS03 and gives meaning to the statement "the arrangement generated by a generalization".

PROPOSITION 2. *For every* $Y \subseteq X$, A1$^\flat$–A3$^\flat$ *and 2*-diversity *hold for the* Y*-generalization* $\acute{\mathcal{R}}$ *if, and only if, there exists* $\acute{v}^{(\cdot,\cdot)} : Y^2 \times \mathbb{T} \to \mathbb{R}$ *such that,*

(i) *for every* $x, y \in Y$ *and* $J \in \mathfrak{J}$, $x \, \acute{\mathcal{R}}_J \, y$ *if, and only if,* $\langle \acute{v}^{(x,y)}, J \rangle > 0$; and

(ii) *for every* $x, y \in Y$, *there exists* $s, t \in \mathbb{T}$ *such that* $\acute{v}^{(x,y)}(s) < 0 < \acute{v}^{(x,y)}(t)$.

Moreover, $\acute{v}^{(x,y)}$ *is unique upto multiplication by a positive scalar,* $v^{(y,x)} = -v^{(x,y)}$ *and* $\acute{\mathcal{R}}$ *is novel if, and only if, for every* $t \neq \mathfrak{f}$, $\acute{v}^{(\cdot,\cdot)}(t) \neq \acute{v}^{(\cdot,\cdot)}(\mathfrak{f})$.

See proof on page 47. We refer to a matrix $\acute{v}^{(\cdot,\cdot)}$ that satisfies condition (ii) of proposition 2 as a 2*-diverse matrix*. We refer to a matrix $\acute{v}^{(\cdot,\cdot)}$ that satisfies proposition 2 as a 2-diverse (pairwise) matrix representation of $\mathcal{R}$.

STEP B.2 (for $\mathbb{T} < \infty$, characterisations of A4$^\flat$). A matrix $v^{(\cdot,\cdot)}$ that satisfies the conditions of the next lemma is a *conditionally-2-diverse (pairwise) representation.*

LEMMA 2.1 (conditionally-2-diverse representation). *Let* $\mathcal{R}$ *be a* Y*-generalization of* $\leq_{\mathbb{J}}$ *with* 2*-diverse matrix representation* $v^{(\cdot,\cdot)}$. *Then* A4$^\flat$ *holds on* Y *if, and only if, for every three distinct elements* $x, y, z \in Y$, $v^{(x,z)}$ *and* $v^{(y,z)}$ *are noncollinear.*

See proof on page 48. The following is a translation of observation 2.

PROPOSITION 3 (on A4$^\flat$ and A4$'^\flat$). *Let* $\mathcal{R}$ *be a* Y*-generalization* $\mathcal{R}$ *satisfying* A0$^\flat$–A3$^\flat$. *Then* A4$^\flat$ *holds if, and only if* A4$'^\flat$ *does. Morover,* $\#\text{total}(\mathcal{R}) \geqslant 4$.

See proof on page 48.

STEP B.3 (for $\mathbb{T} < \infty$, a characterisation of *4*-stability). The following Jacobi identity plays a central role in the proof of GS03.

DEFINITION. *For* $Y \in 2^X$, *the matrix* $v^{(\cdot,\cdot)} : Y^2 \times \mathfrak{T} \to \mathbb{R}$ *satisfies the* Jacobi identity *whenever, for every* $x, y, z \in Y$, $v^{(x,z)} = v^{(x,y)} + v^{(y,z)}$.

For any given Y-generalization $\mathcal{R}$, the *Jacobi identity holds for* $\mathcal{R}$ whenever it holds for some pairwise representation $v^{(\cdot,\cdot)}$ of $\mathcal{R}$. Moreover, in this case, $v^{(\cdot,\cdot)}$ is a *Jacobi representation.* Finally, if the Y-generalization $\mathcal{R}$ is improper and the Jacobi identity holds for $\mathcal{R}$, we simply say that the *Jacobi identity holds on* Y. Consider

k-Jac. *For every* $Y \subseteq X$ *with* $3 \leqslant \#Y \leqslant k$, *the Jacobi identity holds on* Y.

We will work with *3*-Jac and *4*-Jac in particular. The following lemma is the special case of theorem 3 of section C where $\mathbb{T}$ is finite. When $\mathbb{T}$ is finite, for every Y, the set of testworthy Y-generalizations that are novel is nonempty. In this case, *4*-stability implies A0$^\flat$–A3$^\flat$ via nonrevision of rankings (part 2 of definition 2).

Lemma 2.2. *For* $\preceq_{\mathbb{J}}$ *satisfying* A4$^\flat$, *4*-stability *holds if, and only if,* *4*-Jac *holds.*

Step B.4 (for $\#\mathbb{T} < \infty$, the concluding arguments in the proof of theorem 2). In step B.2 we showed that A1$^\flat$–A4$^\flat$ hold if, and only if, $\preceq_{\mathbb{J}}$ has a conditionally 2-diverse pairwise representation. In step B.3, we showed that, when $\preceq_{\mathbb{J}}$ satisfies A0$^\flat$–A4$^\flat$, a necessary and sufficient condition for *4*-Jac is *4*-P$^\flat$. In step B.5, via (mathematical) induction, we showed that conditionally 2-diverse Jacobi representations of Y-generalizations of $\preceq_{\mathbb{J}}$ such that $\#Y = 4$, can be patched together to obtain a conditionally 2-diverse Jacobi representation of $\preceq_{\mathbb{J}}$ (on all of X, regardless of cardinality). The fact that A0$^\flat$ is necessary for a Jacobi representation follows from GS03. As a consequence, A0$^\flat$–A4$^\flat$ and *4*-P$^\flat$ are necessary and sufficient for a conditionally 2-diverse Jacobi representation of $\preceq_{\mathbb{J}}$. The following argument then completes the proof of theorem 2.

Let $v^{(\cdot,\cdot)}$ be a (conditionally 2-diverse) Jacobi representation of $\preceq_{\mathbb{J}}$ and define $\mathbf{v} : X \times \mathbb{T} \to \mathbb{R}$ as follows. Fix arbitrary $w \in X$, and let $\mathbf{v}(w, \cdot) = 0$. Then, for every $x \in X$, let $\mathbf{v}(x, \cdot) = v^{(w,x)}$. Recalling that $v^{(w,x)} = -v^{(x,w)}$, note that, since the rows of $v^{(\cdot,\cdot)}$ satisfy the Jacobi identity, for every $x, y \in X$, we have

$$v^{(x,y)} = v^{(x,w)} + v^{(w,y)} = -\mathbf{v}(x, \cdot) + \mathbf{v}(y, \cdot).$$

To see that (2.a) holds note that, for every $J \in \mathbb{J}$, we have $x \preceq_J y$, if, and only if, $0 \leqslant \langle v^{(x,y)}, J\rangle$, if, and only if, $\langle v(x, \cdot), J\rangle \leqslant \langle \mathbf{v}(y, \cdot), J\rangle$. For (2.b), note that, since $v^{(\cdot,\cdot)}$ is a conditionally 2-diverse pairwise representation, for every $x, y \in X$,

$$0 \nleqslant v^{(x,y)} = -\mathbf{v}(x, \cdot) + \mathbf{v}(y, \cdot).$$

Finally, for every $z \in X$, we have, for every $\lambda \in \mathbb{R}$, $v^{(z,x)} \neq \lambda v^{(z,y)}$. Equivalently,

$$v(x, \cdot) \neq (1 - \lambda)\mathbf{v}(z, \cdot) + \lambda \mathbf{v}(y, \cdot).$$

Theorem 1 part II, on uniqueness, follows from lemma 2.3 and, without modification, part 3 of the proof of theorem 2 of GS03 (see page 23).

STEP B.5 (for arbitrary X and $\mathbb{T} < \infty$, the induction argument). The present step is closely related to lemma 3 and claim 9 of GS03. There the authors establish that, when *4*-diversity holds, 3-Jac is a necessary and sufficient condition for the (global) Jacobi identity to hold on X. GS03 relies on the fact that *4*-diversity implies linear independence of $\{v^{(x,y)}, v^{(y,z)}, v^{(z,w)}\}$ for every four distinct elements $x, y, z, w \in X$. In the present setting, where conditional-*2*-diversity only implies linear independence of pairs $\{v^{(x,y)}, v^{(y,z)}\}$, the Jacobi identity requires *4*-Jac.

LEMMA 2.3 (Jacobi representation). *Let* $\leq_{\mathbb{J}}$ *have a conditionally-*2*-diverse representation* $u^{(\cdot,\cdot)}$*. Then* *4*-Jac *holds if, and only if,* $\leq_{\mathbb{J}}$ *has a Jacobi representation* $v^{(\cdot,\cdot)}$*. Moreover, for every Jacobi representation* $\mathbf{v}^{(\cdot,\cdot)}$ *of* $\leq_{\mathbb{J}}$ *there exists* $\lambda > 0$ *satisfying* $\mathbf{v}^{(\cdot,\cdot)} = \lambda v^{(\cdot,\cdot)}$.

See proof on page 49.

Appendix C. Statement and proof of theorem 3

Theorem 3. *For* $\leq_{\mathbb{J}}$ *satisfying* A0$^\flat$–A4$^\flat$*,* *4*-stability *holds if, and only if, the Jacobi identity holds (for some pairwise representation of* $\leq_{\mathbb{J}}$*).*

Via lemma 2.3, it suffices to show that *4*-stability is equivalent to *4*-Jac. Then by construction, *4*-stability and *4*-Jac are conditions that apply independently to each subset Y of X that has cardinality 3 or 4. Throughout the present section, $Y \subseteq X$ has cardinality 3 or 4 and $\mathcal{R}$ denotes the improper Y-generalization of $\leq_{\mathbb{J}}$.

Suppose there is no testworthy Y-generalization that is novel. In this case, *4*-stability holds vacuously on Y. The following lemma then confirms that *4*-diversity holds and theorem 2 of GS03 applies, so that *4*-Jac holds on Y.

Lemma 3.1. *If every testworthy* Y*-generalization of* $\leq_{\mathbb{J}}$ *is regular, then* $|\mathbb{T}| = \infty$ *and* *4*-diversity *holds on* Y*.*

Proof of lemma 3.1. Let $|Y| = 4$. (The proof for case where $|Y| = 3$ is similar and omitted.) Via lemma 3.2, the set of testworthy Y-generalizations is nonempty. Let $\hat{\mathcal{R}}$ be a testworthy Y-generalization, so that there exists J in $\mathbb{J}$ such that $\hat{\mathcal{R}}_{J\times 0}$ is total and equal to the inverse of $\hat{\mathcal{R}}_{\mathfrak{f}}$. Let $P \overset{\text{def}}{=} \hat{\mathcal{P}}_{\mathfrak{f}}$ so that $P \subsetneq Y^2$. Via A0$^\flat$–A4$^\flat$, proposition 2 applies, let $\hat{v}^{(\cdot,\cdot)}$ be the matrix representation of $\hat{\mathcal{R}}$ and let $\hat{v}^P$ denote the restriction of $\hat{v}^{(\cdot,\cdot)}$ to $P \times \mathbb{T}^{\mathfrak{f}}$.

Claim 3.1.1. *For every vector $\eta^P = \langle \eta^{(x,y)} \in \mathbb{R}_{++} : (x,y) \in P\rangle$, there exist $s, t \in \mathbb{T}$ such that $\hat{v}^P(s) = \eta^P$ and $\hat{v}^P(t) = -\eta^P$.*

Proof of claim 3.1.1. By way of contradiction, suppose there exists $\acute{\eta}^P \in \mathbb{R}^P_{++}$ such that, for every s in $\mathbb{T}$, $\hat{v}^P(s) \neq \acute{\eta}^P$. That is η^P such that, for every s in $\mathbb{T}$, there exists (x,y) in P such that $\hat{v}^{(x,y)}(s) \neq \acute{\eta}^{(x,y)}$. This property will suffice for the existence of a testworthy Y-generalization $\acute{\mathcal{R}}$ that is novel. Define $\acute{v}^{(\cdot,\cdot)} : X^2 \times \mathbb{T}^{\mathfrak{f}} \to \mathbb{R}$ as follows. For each (x,y) in P, let

$$\acute{v}^{(x,y)}(s) \overset{\text{def}}{=} \begin{cases} \acute{\eta}^{(x,y)} & \text{if } s = \mathfrak{f}, \\ \hat{v}^{(x,y)}(s) & \text{otherwise.} \end{cases}$$

For every (x,y) in P^{-1}, since (y,x) in P, take $\acute{v}^{(x,y)} = -\acute{v}^{(y,x)}$. Finally, since P is the asymmetric part of a total ordering, for every remaining (x,y) in Y^2, $x = y$, so let $\acute{v}^{(x,y)} = 0$. Observe that, by construction, for every s in $\mathbb{T}$, $\acute{v}^{(\cdot,\cdot)}(s) \neq \acute{v}^{(\cdot,\cdot)}(\mathfrak{f})$. This allows us to appeal to proposition 2 and take $\acute{\mathcal{R}}$ to be the novel generalization that $\acute{v}^{(\cdot,\cdot)}$ generates. Moreover, $\acute{\mathcal{R}}$ is also testworthy. For together $P = \hat{\mathcal{P}}_{\mathfrak{f}}$ and $\hat{\mathcal{P}}_{\mathfrak{f}} = \hat{\mathcal{P}}^{-1}_{J\times 0}$ imply that $\acute{\mathcal{R}}_{\mathfrak{f}} = \acute{\mathcal{R}}^{-1}_{J\times 0}$ since, for every (x,y) in P, we have

$$\langle \acute{v}^{(x,y)}, J \times 0\rangle = \langle \hat{v}^{(x,y)}, J \times 0\rangle < 0 < \acute{\eta}^{(x,y)} = \acute{v}^{(x,y)}(\mathfrak{f}).$$

This contradiction implies that, for every $\eta^P \in \mathbb{R}^P_{++}$, there exists s in $\mathbb{T}$ such that $\hat{v}^P(s) = \eta^P$. *Mutatis mutandis*, a repetition of the preceding argument by contradiction confirms that there exists t in $\mathbb{T}$ such that $\hat{v}^P(t) = -\eta^P$. □

Claim 3.1.1 implies that, when every testworthy Y-generalization is regular, the cardinality of $\mathbb{T}$ is equal to the cardinality of $\mathbb{R}^P$. We now show that *4*-diversity holds on Y. Let R denote an arbitrary total ordering of Y. We show that, for some K in $\mathbb{J}$, $\langle \hat{v}^{(x,y)}, K\rangle \geqslant 0$ if, and only if, (x,y) belongs to R. Claim 3.1.1

ensures that we can choose s in $\mathbb{T}$ such that, for some $0 < \epsilon < 1$

$$\hat{v}^{(x,y)}(s) = \begin{cases} 1+\epsilon & \text{if } (x,y) \text{ in } R \cap P, \\ 1-\epsilon & \text{if } (x,y) \text{ in } R^{-1} \cap P. \end{cases}$$

Via claim 3.1.1, take t in $\mathbb{T}$ such that, for every (x,y) in P, $\hat{v}^{(x,y)}(t) = -1$. Let $K := \delta_s + \delta_t$ in $\mathbb{J}^{\mathfrak{f}}$, so that $\langle \hat{v}^{(x,y)}, K\rangle = \hat{v}^{(x,y)}(s) + \hat{v}^{(x,y)}(t)$. By evaluating terms and observing that $\epsilon > 0$ we obtain

$$\langle \hat{v}^{(x,y)}, K\rangle = \begin{cases} (1+\epsilon) - 1 > 0 & \text{if } (x,y) \text{ in } R \cap P, \\ (1-\epsilon) - 1 < 0 & \text{if } (x,y) \text{ in } R^{-1} \cap P. \end{cases}$$

Since (x,y) in $R^{-1} \cap P^{-1}$ if, and only if, (y,x) in $R \cap P$ (and, similarly, (x,y) in $R \cap P^{-1}$ if, and only if (y,x) in $R^{-1} \cap P$), we appeal to $\hat{v}^{(x,y)} = -\hat{v}^{(y,x)}$ and obtain

$$\langle \hat{v}^{(x,y)}, K\rangle = \begin{cases} -(1+\epsilon) + 1 < 0 & \text{if } (x,y) \text{ in } R^{-1} \cap P^{-1}, \\ -(1-\epsilon) + 1 > 0 & \text{if } (x,y) \text{ in } R \cap P^{-1}. \end{cases}$$

Since P is the asymmetric part of a total ordering we conclude that, for every $x \neq y$, $\langle \hat{v}^{(x,y)}, K\rangle$ has the right sign. Finally, for $x = y$, $\langle \hat{v}^{(x,y)}, K\rangle = 0$. □

As a consequence of lemma 3.1, for the remainder of the proof of theorem 3 we work under the assumption that the set of testworthy Y-generalizations that are novel is nonempty. Throughout the sequel, $\mathcal{R}$ denotes an improper Y-generalization. Since A1$^\flat$–A4$^\flat$ hold, the pairwise representation $u^{(\cdot,\cdot)} : Y^2 \times \mathbb{T} \to \mathbb{R}$ of $\mathcal{R}$ is conditionally 2-diverse. Lemma 2.1 implies that $u^{(\cdot,\cdot)}$ has row rank $\mathbf{r} \geqslant 2$.

Lemma 3.2. *For every ranking $R \in \operatorname{total}(\mathcal{R})$, there exists a (testworthy) generalization $\acute{\mathcal{R}}$ with $\acute{\mathcal{R}}_{\mathfrak{f}} = R^{-1}$. Moreover, $\acute{\mathcal{R}}$ has a central arrangement with rank $\acute{\mathbf{r}} = \mathbf{r}$. Finally, 4-*Jac *holds for $\acute{\mathcal{R}}$ if, and only if, it holds for $\mathcal{R}$.*

Proof. Let $J \in \mathbb{J}$ such that $\mathcal{R}_J$ is total. Since Y is finite, so is the dimension of S_{++}. Take $L \in S_{++}$ such that $\mathcal{R}_L = \mathcal{R}_J$ and, for some rational $0 < \acute{\iota} < 1$, let $\acute{J} = (1-\acute{\iota})J \times \acute{\iota}$. Now, for every x, y in Y, let

$$\acute{\eta}^{(x,y)} := -\frac{1-\acute{\iota}}{\acute{\iota}} \langle u^{(x,y)}, J\rangle,$$

and let $\acute{u}^{(x,y)} \stackrel{\text{def}}{=} u^{(x,y)} \times \acute{\eta}^{(x,y)}$, so that $\langle \acute{u}^{(x,y)}, \acute{J}\rangle = 0$. Let $\acute{\mathcal{R}}$ be the associated Y-generalization, so that by construction $\acute{\mathcal{R}}_{\mathfrak{f}} = \acute{\mathcal{R}}_J^{-1}$. Since $\acute{J} \in \acute{H}_{++}^{\{x,y\}}$ for every distinct $x, y \in Y$, $\acute{\mathcal{H}}_{++}$ is central. This construction together with linearity of the inner product ensures that the Jacobi equations (27)–(29) hold for $u^{(\cdot,\cdot)}$ if, and only if, they hold for $\acute{\eta}^{(\cdot,\cdot)}$ and hence $\acute{u}^{(\cdot,\cdot)}$. Indeed, $\mathbf{r} = \acute{\mathbf{r}}$ holds for the same reasons. □

The case where Y has cardinality 3, follows from the next two lemmas.

Lemma 3.3. *If $Y = \{x, y, z\}$ and $\mathbf{r} = 2$, then 4-*Jac *and 4-*stability *hold on Y.*

Proof of lemma 3.3. Fix arbitrary $J \in \mathbb{J}$ such that $\mathcal{R}_J$ is total. We first apply Zaslavski's theorem to prove that the Y-generalization $\acute{\mathcal{R}}$ of lemma 3.2 (which, recall, is testworthy relative to J) satisfies $|\acute{\mathcal{G}}_{++}| = 6$.

Via lemma 3.2 the arrangement $\acute{\mathcal{H}}_{++}$ is central. Thus $\acute{\mathcal{H}}_{++} \equiv \acute{\mathcal{H}}$ and we may drop reference to the subscript $_{++}$. Via lemma 2.1, $\#\acute{\mathcal{H}} = 3$. Moroever, since every subarrangement of $\acute{\mathcal{H}}$ is central, for every $k = 0, 1, 2, 3$ there are $\binom{3}{k}$ ways to choose $|\mathcal{A}| = k$ hyperplanes from $\acute{\mathcal{H}}$. For $k < 3$, the rank of every subarrangement is k. For $k = 3$, the rank of the arrangement is $\acute{\mathbf{r}}$. Via lemma 3.2 $\acute{\mathbf{r}} = \mathbf{r} = 2$. Thus

$$|\acute{\mathcal{G}}| = \binom{3}{3}(-1)^{3-\acute{\mathbf{r}}} + \binom{3}{2}(-1)^{2-2} + \binom{3}{1}(-1)^{1-1} + \binom{3}{0}(-1)^{0-0} = 6. \tag{14}$$

For both 4-Jac and 4-stability, we require that every member of $\acute{\mathcal{G}}$ is associated with a total ordering. 3-Jac then holds because we have the conditions (A0$^\flat$-A3$^\flat$ and 3-diversity) to apply lemma 2 of GS03. 4-stability then holds since, for every testworthy $\hat{\mathcal{R}}$ that is novel and satisfies $\hat{\mathcal{R}}_{\mathfrak{f}} = \mathcal{R}_J^{-1}$ and A1$^\flat$-A3$^\flat$, $\acute{\mathcal{R}}$ is a diverse perturbation of $\hat{\mathcal{R}}$ that satisfies A0$^\flat$-A3$^\flat$.

It remains for us to show that every member of $\acute{\mathcal{G}}$ is associated with a total ranking of Y. In particular, since every associated ranking is CAR (see section E) it suffices to prove transitivity. For this, note that, via 3 four of the six members of $\acute{\mathcal{G}}$ intersect $S_{++} \times 0$ and are therefore associated with transitive rankings. Note that, since S_{++} is connected, the remaining two members are adjacent (separated by a single member of the arrangement). Take G to be one of the remaining members of $\acute{\mathcal{G}}$ and take $L \in G$. Define the affine path $\lambda \mapsto \phi(\lambda) = (1-\lambda)\acute{L} + \lambda\acute{J}$, where $\acute{J}$ belongs to the center $H^{\{x,y,z\}}$ of $\acute{\mathcal{H}}$. For λ sufficiently close but greater than one, $\acute{\mathcal{R}}_L = \acute{\mathcal{R}}_{\phi(\lambda)}^{-1}$ because $\phi(1) = \acute{J}$ belongs to the center. Thus, $\acute{\mathcal{R}}_L$ is transitive. □

Lemma 3.4. *If $Y = \{x, y, z\}$ and $\mathbf{r} = 3$, then neither 4-*stability *nor 4-*Jac *hold.*

Proof of lemma 3.4. When $\mathbf{r} = 3$, $u^{(x,y)}, u^{(y,z)}$ and $u^{(x,z)}$ are linearly independent, so that *3*-Jac fails to hold. We now confirm that *3*-stability also fails to hold. Via lemma 3.2, $\acute{\mathbf{r}} = \mathbf{r}$. We now apply eq. (14) with $\mathbf{r} = 3$:

$$|\acute{\mathcal{G}}| = (-1)^0 + 3(-1)^0 + 3(-1)^0 + (-1)^0 = 8.$$

Then there are $3! = 6$ members of total$(\acute{\mathcal{R}})$, and the two additional regions of $\acute{\mathcal{G}}$ are associated with intransitive CAR rankings. It remains for us to show that every Y-generalization $\hat{\mathcal{R}}$ with $\#\text{total}(\hat{\mathcal{R}}) = 6$ fails to satisfy A0$^\flat$.

Recall fig. 2b. If $\hat{\mathcal{R}}$ are the sentiments corresponding to this arrangement, then $|\hat{\mathcal{G}}| = 7$. This value is achieved by dropping the first term in eq. (14). That is, since $\hat{A}^{\{x,y,z\}}$ is the only member of $\hat{\mathcal{L}} - \hat{\mathcal{L}}_{++}$. We may reduce $|\hat{\mathcal{G}}|$ further by excluding one of the intersections of two hyperplanes such as $\hat{H}^{\{x,y\}} \cap \hat{H}^{\{y,z\}}$. That is $\hat{A}^{\{x,y\},\{y,z\}}$. In terms of eq. (14), this amounts to excluding one of the $\binom{3}{2} = 3$ central subarrangements. This would reduce $|\hat{\mathcal{G}}_{++}|$ to 6. To obtain $\hat{\mathcal{R}}$ satisfying A0$^\flat$, we would need to remove all $\binom{3}{2}$ central subarrangements of two hyperplanes. But this would reduce $|\hat{\mathcal{G}}_{++}|$ to 4. □

The case where Y has cardinality 4. Note that a failure of *3*-Jac on $Z \subset Y$ such that $|Z| = 3$ implies a failure of *4*-Jac on Y. And since the arguments for the case where $|Y| = 3$ account for the case where *3*-Jac fails, we henceforth assume that *3*-Jac holds on Y. That is, our conditionally *2*-diverse representation $u^{(\cdot,\cdot)}$ will now satisfy equations (27)–(29) with $\hat{\beta} = \beta$ if, and only if, *4*-Jac holds on Y.

First some some useful results that exploit *3*-Jac.

Proposition 4. *If $Y = \{x, y, z, w\}$ and 3-*Jac *holds for $u^{(\cdot,\cdot)}$, then, for every $\acute{v}^{(\cdot,\cdot)} = u^{(\cdot,\cdot)} \times \acute{\eta}^{(\cdot,\cdot)}$ with rank $\acute{\mathbf{r}}$, that satisfies 3-*Jac, $2 \leqslant \acute{\mathbf{r}} \leqslant 3$.

Proof of proposition 4. Via proposition 3 and lemma 3.3, $\mathbf{r} \geqslant 2$. Indeed the span of $\{u^{(x,w)}, u^{(y,w)}\}$ is two. Let S denote the span of $\{u^{(x,w)}, u^{(y,w)}, u^{(z,w)}\}$. Since $u^{(y,w)} = -u^{(w,y)}$ and $u^{(\cdot,\cdot)}$ satisfies *3*-Jac, equations (27)–(29) hold for $u^{(\cdot,\cdot)}$. (If *4*-Jac fails to hold, then $\beta \neq \hat{\beta}$, but the equations still hold.) Thus $u^{(x,y)}$, $u^{(y,z)}$ and $u^{(x,z)}$ all belong to S and $\mathbf{r} \leqslant 3$. Now note that above argument does not depend on the cardinality of $\mathbb{T}$, thus take $\acute{\eta}^{(\cdot,\cdot)}$ to satisfy *3*-Jac: indeed with the same parameters that feature in equations (27)–(29) for $u^{(\cdot,\cdot)}$. The preceding argument then generalizes *mutatis mutandis* to $\acute{v}^{(\cdot,\cdot)}$ and $2 \leqslant \acute{\mathbf{r}} \leqslant 3$. □

Proposition 5. *If $Y = \{x, y, z, w\}$, then $4 \leqslant |\mathcal{H}| \leqslant 6$ and these bounds are tight. If, moreover, 3-*Jac *holds for $u^{(\cdot,\cdot)}$ and $|\mathcal{H}| < 6$, then $\mathbf{r} = 2$.*

Proof of proposition 5. The upper bound $|\mathcal{H}| \leqslant 6$ follows because there are $\binom{4}{2} = 6$ ways to choose distinct pairs of elements from Y. Via lemma 2.1 at most the following equalities are feasible: $H^{\{x,y\}} = H^{\{z,w\}}$, $H^{\{x,z\}} = H^{\{y,w\}}$ and $H^{\{y,z\}} = H^{\{x,w\}}$. *Ad absurdum* suppose all three equalities hold, so that $|\mathcal{H}| = 3$. Via *2*-diversity, all six hyperplanes partition S_{++}, so w.l.o.g. suppose that $R = (x, y, z, w)$ and its inverse R^{-1} both feature in $\mathcal{R}$. Consider a convex path in S_{++} from G^R to $G^{R^{-1}}$. Since $H^{\{x,y\}} = H^{\{z,w\}}$, it follows that both support G^R. Thus, for $R' = (y, x, z, w)$, $G^{R'}$ is adjacent to G^R. Continuing along the convex path we find that $H^{\{x,z\}}$ supports $G^{R'}$. But then $H^{\{x,z\}} = H^{\{y,w\}}$ yields a contradiction of A0$^\flat$.

We now prove that *3*-Jac and $H^{\{x,y\}} = H^{\{z,w\}}$ together imply $\mathbf{r} = 2$. Consider equations (27)–(29) (so *3*-Jac holds, but *4*-Jac need not). Via (27), $S = \{u^{(x,w)}, u^{(w,y)}, u^{(x,y)}\}$ is 2-dimensional. Since $u^{(x,y)}$ and $u^{(z,w)}$ are collinear, $u^{(z,w)}$ belongs to S. Finally, equations (28) and (29) yield $u^{yz}, u^{(x,z)} \in S$. $\square$

Lemma 3.5. *If $Y = \{x, y, z, w\}$, $\mathbf{r} = 3$ and 3-*Jac *holds on Y, then 4-*stability *and 4-*Jac *both hold on Y.*

Proof of lemma 3.5. To see that *4*-Jac holds, appeal to the proof of lemma 3 of GS03: if *3*-Jac holds and *4*-Jac does not, then $\{u^{(x,w)}, u^{(y,w)}, u^{(z,w)}\}$ is linearly dependent. This is in contradiction of $\mathbf{r} = 3$.

We now verify that *4*-stability also holds. Since $\mathbf{r} = 3$, the contrapositive of proposition 5 implies that $|\mathcal{H}_{++}| = 6$. Via lemma 3.2, there exists a testworthy Y-generalization $\acute{\mathcal{R}}$ with an arrangement $\acute{\mathcal{H}}_{++}$ that is central and rank $\acute{\mathbf{r}} = \mathbf{r}$. Since $\acute{\mathcal{L}}_{++} = \acute{\mathcal{L}}$, we drop reference to $_{++}$. The rank of subarrangements with cardinality 4 or more is $\acute{\mathbf{r}}$. Let $\acute{\tau}$ denote the number of subarrangements $\mathcal{A}$ that have cardinality 3 and rank 2. Each of the other $\binom{6}{3} - \acute{\tau}$ subarrangements with cardinality 3 have rank $\acute{\mathbf{r}}$. All other subarrangements have rank equal to their cardinality.

$$
\begin{aligned}
|\acute{\mathcal{G}}| =& \binom{6}{6}(-1)^{6-\acute{\mathbf{r}}} + \binom{6}{5}(-1)^{5-\acute{\mathbf{r}}} + \binom{6}{4}(-1)^{4-\acute{\mathbf{r}}} \\
&+ \binom{6}{3}(-1)^{3-\acute{\mathbf{r}}} - \acute{\tau}\ (-1)^{3-\acute{\mathbf{r}}} + \acute{\tau}\ (-1)^{3-2} \\
&+ \binom{6}{2}(-1)^{2-2} + \binom{6}{1}(-1)^{1-1} + \binom{6}{0}(-1)^{0-0}
\end{aligned}
\tag{15}
$$

We claim that $\acute{\tau} = 4$. Each of the $\binom{4}{3} = 4$ subsets of Y that have cardinality 3 generates a subarrangement of cardinality $\binom{3}{2} = 3$. (For instance, $\mathcal{A}^{\{x,y,z\}} = \left\{\acute{H}^{\{x,y\}}, \acute{H}^{\{y,z\}}, \acute{H}^{\{x,z\}}\right\}$.) For such subarrangements, $\mathit{4}$-Jac implies a rank of 2. Arguments from the final step in the proof of proposition 5 confirm that every other subarrangement with cardinality 3 has rank 3. Equation (15) then implies

$$|\acute{\mathcal{G}}| = -1 + 6 - 15 + 20 - 4 - 4 + 15 + 6 + 1 = 24 = 4!.$$

The fact that $\acute{\mathcal{R}}$ satisfies A0$^\flat$ is an immediate consequence of $\mathit{4}$-Jac. $\mathit{4}$-stability then follows, for, if $\check{\mathcal{R}}$ is any novel, testworthy generalization that satisfies $\check{\mathcal{R}}_{\mathfrak{f}} = \mathcal{R}_J^{-1}$, then $\acute{\mathcal{R}}$ is a diverse perturbation of $\check{\mathcal{R}}$ that satisfies A0$^\flat$-A3$^\flat$. □

In the remaining case, where $Y = \{x, y, z, w\}$ **and** $\mathbf{r} = 2$**,** the proof is complicated by the fact that maximally-diverse generalizations have centerless arrangements. We begin by choosing $\acute{\eta}^{(\cdot,\cdot)}$ so as to construct $\acute{u}^{(\cdot,\cdot)} = u^{(\cdot,\cdot)} \times \acute{\eta}^{(\cdot,\cdot)}$ with $\acute{\mathbf{r}} = 3$.

Since $\mathbf{r} = 2$, it follows that $u^{(x,z)}, u^{(y,z)}$ and $u^{(w,z)}$ form a linearly dependent set. Thus, for some $\pi, \rho \in \mathbb{R}$,

$$\pi u^{(x,z)} + \rho u^{(y,z)} = u^{(w,z)}. \tag{16}$$

Fix arbitrary $J \in \mathbb{J}$ such that $\mathcal{R}_J$ is total, and, as in lemma 3.2, w.l.o.g., we take $J \in G_{++}^{(x,y,z,w)}$. Then, for $\acute{\iota} = \frac{1}{2}$ and $\acute{J} = (1 - \acute{\iota})J \times \acute{\iota}$, let

$$\acute{\eta}^{(i,j)} = -\tfrac{1-\acute{\iota}}{\acute{\iota}}\langle u^{(i,j)}, J\rangle = -\langle u^{(i,j)}, J\rangle, \quad \text{for every } i, j \in \{x, y, z\}. \tag{17}$$

In the case where $\#Y = 3$, eq. (17) implies that the associated arrangement of hyperplanes $\acute{\mathcal{H}}_{++}$ is central. That is, recalling example 6, the associated positive intersection semilattice $\acute{\mathcal{L}}_{++}$ is isomorphic to $\acute{\mathcal{L}}$. The structure of $\acute{\mathcal{L}}_{++}$ is determined by the rank $\acute{\mathbf{r}}$ of $\acute{\mathcal{U}} = \left\{\acute{u}^{(x,y)}, \acute{u}^{(x,z)}, \acute{u}^{(y,z)}\right\}$. Via lemma 2.1, the rank of $\acute{\mathcal{U}}$ satisfies $2 \leqslant \acute{\mathbf{r}} \leqslant 3$.

$\mathit{4}$-Jac holds for the associated generalisation $\acute{\mathcal{R}}$ if, and only if, $\acute{\mathbf{r}} = 2$. Observe that, by construction, $\acute{u}^{(\cdot,\cdot)}$ satisfies the Jacobi identity if, and only if,

such that $\acute{\mathcal{G}}$ is maximal. This is because e Let $f : G_{++}^{(x,y,z,w)} \to \acute{G}_{++}^{(x,y,z,w)}$ be the mapping $L \to (1 - \acute{\lambda})L \times \acute{\lambda}$, where $\acute{\lambda}$ is the solution to $\acute{\eta}^{(y,z)} = -\frac{1-\acute{\lambda}}{\acute{\lambda}}\langle u^{(y,z)}, L\rangle$. In particular, substituting for $\acute{\eta}^{(y,z)}$ using eq. (17), we obtain $\frac{1-\acute{\lambda}}{\acute{\lambda}} = \frac{\langle u^{(y,z)}, J\rangle}{\langle u^{(y,z)}, L\rangle}$ and

$\acute{\lambda} = \frac{\langle u^{(y,z)}, L\rangle}{\langle u^{(y,z)}, J\rangle + \langle u^{(y,z)}, L\rangle}$. Since $J, L \in G^{(y,z)}$, all terms in the expression for $\acute{\lambda}$ are positive, so that $0 < \acute{\lambda} < 1$ and, via convexity of $\acute{G}^{(x,y,z,w)}_{++}$, f is well-defined. As the quotient of continuous functions of L (with the denominator $\langle u^{(y,z)}, J\rangle + \langle u^{(y,z)}, L\rangle$ bounded away from zero) f is continuous. Finally, $\lim_{L \to J} f(L) = \acute{J}$.

Via A4$'^{\flat}$ and $J \in G^{(x,y,z,w)}_{++}$, $\langle u^{(y,z)}, J\rangle \neq \langle u^{(z,w)}, J\rangle$ and both numbers are positive. Via $u^{(w,z)} = -u^{(z,w)}$, it follows that $\zeta = \frac{\langle u^{(w,z)}, J\rangle}{\langle u^{(y,z)}, J\rangle} < 0$ is the unique solution to

$$\langle \zeta u^{(y,z)} - u^{(w,z)}, J\rangle = 0.$$

Continuity of the map $L \mapsto \frac{\langle u^{(w,z)}, L\rangle}{\langle u^{(y,z)}, L\rangle}$ on $G^{(x,y,z,w)}_{++}$ and the fact that the latter set is open suffices for the existence of a sequence $(L_n : n = 1, 2, \dots)$, converging to J, such that, for every n, $L_n \mapsto \xi_n = \frac{\langle u^{(w,z)}, L_n\rangle}{\langle u^{(y,z)}, L_n\rangle}$ satisfies $\xi_n \neq \zeta$.

Next, take $(\epsilon_n : n = 1, 2, \dots)$ be the following non-zero real-valued sequence that converges to zero as $L_n \to J$:

$$\epsilon_n := \langle \xi_n u^{(y,z)} - u^{(w,z)}, J\rangle = \langle u^{(w,z)}, \tfrac{1-\acute{\lambda}_n}{\acute{\lambda}_n} L_n - J\rangle, \tag{18}$$

where $\frac{1-\acute{\lambda}_n}{\acute{\lambda}_n} = \frac{\langle u^{(y,z)}, J\rangle}{\langle u^{(y,z)}, L_n\rangle}$ is defined as in the definition of f above.

For every $(i, j) \in \{y, z, w\}^2 - \{(y, z), (z, y)\}$, let $\acute{\eta}^{(i,j)} = -\frac{1-\acute{\lambda}}{\acute{\lambda}}\langle u^{(i,j)}, L\rangle$. For $(i, j) \in \{x, y, z, w\}^2 - \{(x, w), (w, x)\}$, let $\acute{u}^{(i,j)} := u^{(i,j)} \times \acute{\eta}^{(i,j)}$. To complete the definition of $\acute{u}^{(\cdot,\cdot)}$, we appeal to the fact that, via *3*-Jac, $u^{(\cdot,\cdot)}$ satisfies equations (27)–(29). In particular, from these equations, take parameters α, β, and γ and let $\acute{\eta}^{(x,w)}$ be the (unique) solution to the Jacobi identity

$$\alpha\acute{\eta}^{(x,w)} = \gamma\acute{\eta}^{(x,y)} + \beta\acute{\eta}^{(y,w)} = -\langle \gamma u^{(x,y)}, J\rangle - \tfrac{1-\acute{\lambda}}{\acute{\lambda}}\langle \beta u^{(y,w)}, L\rangle. \tag{19}$$

For these parameter values, $\acute{u}^{(\cdot,\cdot)}$ also satisfies (27)–(29) of the proof of lemma 2.3. That is, for $\{x, y, z\}$, via eq. (17) and (29), $\gamma\acute{\eta}^{(x,y)} + \tau\acute{\eta}^{(y,z)} = \phi\acute{\eta}^{(x,z)}$, so that $\acute{u}^{(\cdot,\cdot)}$ satisfies (29). For $\{y, z, w\}$, Via **??** and (28), $\hat{\beta}\acute{\eta}^{(y,w)} + \sigma\acute{\eta}^{(w,z)} = \tau\acute{\eta}^{(y,z)}$, so that $\acute{u}^{(\cdot,\cdot)}$ satisfies (28). For $\{x, y, w\}$, via eq. (19), $\acute{u}^{(\cdot,\cdot)}$ satisfies (27).

Now note that, for every $L \neq J$,

$$\pi\acute{\eta}^{(x,z)} + \rho\acute{\eta}^{(y,z)} = \acute{\eta}^{(w,z)} + \epsilon \neq \acute{\eta}^{(w,z)}. \tag{20}$$

Then, via (20), for every $\epsilon \neq 0$, $\{\acute{u}^{(x,y)}, \acute{u}^{(y,z)}, \acute{u}^{(w,z)}\}$ forms a linearly independent set.

We now demonstrate that for the final triple $\{x, z, w\}$, the Jacobi identity holds if $\hat{\beta} = \beta$, and $\{\acute{u}^{(x,w)}, \acute{u}^{(w,z)}, \acute{u}^{(x,z)}\}$ has rank 3 otherwise.

First extract the parameters from equations (27)–(29) to obtain the matrix form

$$\begin{array}{cc} \begin{array}{cccccc} (x,w) & (y,w) & (x,y) & (w,z) & (y,z) & (x,z) \end{array} & \\ \left[\begin{array}{ccc|ccc} \alpha & -\beta & -\gamma & 0 & 0 & 0 \\ 0 & \hat{\beta} & 0 & \sigma & -\tau & 0 \\ 0 & 0 & \gamma & 0 & \tau & -\phi \end{array}\right] & \begin{array}{c} (27) \\ (28) \\ (29) \end{array} \end{array} \tag{21}$$

Since the triple $\{\acute{u}^{(i,z)} : i = x, y, w\}$ provides a basis for $\operatorname{span}(\acute{u}^{(\cdot,\cdot)})$, we will write all vectors in terms of this basis. To this end, we derive the reduced row echelon form of eq. (21). In particular, letting r_i denote the rows of the matrix, we perform the operation $r_1 \mapsto r_1 + \frac{\beta}{\hat{\beta}} r_2 + r_3$ to obtain

$$\begin{array}{cc} \begin{array}{cccccc} (x,w) & (y,w) & (x,y) & (w,z) & (y,z) & (x,z) \end{array} & \\ \left[\begin{array}{ccc|ccc} \alpha & 0 & 0 & \frac{\beta}{\hat{\beta}}\sigma & (1-\frac{\beta}{\hat{\beta}})\tau & -\phi \\ 0 & \hat{\beta} & 0 & \sigma & -\tau & 0 \\ 0 & 0 & \gamma & 0 & \tau & -\phi \end{array}\right] & \begin{array}{c} (27) \\ (28) \\ (29) \end{array} \end{array} . \tag{22}$$

In eq. (22), the fact that $\hat{\beta}$ (instead of β) that appears as a pivot in column 2, is a consequence of the fact that, in this derivation, we are choosing $\acute{v}^{(y,w)} = \hat{\beta}\acute{u}^{(y,w)}$. The other (relevant) rows of $\acute{v}^{(\cdot,\cdot)} : Y^2 \times \mathbb{T}^{\dagger} \to \mathbb{R}$ are $\acute{v}^{(x,w)} = \alpha\acute{u}^{(x,w)}$, $\acute{v}^{(x,y)} = \gamma\acute{u}^{(x,y)}$, $\acute{v}^{(z,w)} = \sigma\acute{u}^{(z,w)}$, $\acute{v}^{(y,z)} = \tau\acute{u}^{(y,z)}$ and $\acute{v}^{(x,z)} = \phi\acute{u}^{(x,z)}$. it The matrix of the equation that now follows, is invertible if, and only if, $(1 - \frac{\beta}{\hat{\beta}}) \neq 0$.

$$\begin{bmatrix} \acute{v}^{(x,w)} \\ \acute{v}^{(x,z)} \\ \acute{v}^{(z,w)} \end{bmatrix} = \begin{bmatrix} -\frac{\beta}{\hat{\beta}} & -(1-\frac{\beta}{\hat{\beta}}) & 1 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \acute{v}^{(w,z)} \\ \acute{v}^{(y,z)} \\ \acute{v}^{(x,z)} \end{bmatrix} \tag{23}$$

Thus, unless $\hat{\beta} = \beta$, we conclude that $\{\acute{v}^{(x,w)}, \acute{v}^{(x,z)}, \acute{v}^{(z,w)}\}$ has the same rank as $\{\acute{v}^{(w,z)}, \acute{v}^{(y,z)}, \acute{v}^{(x,z)}\}$ which, by construction, has rank 3 for every choice of $\epsilon \neq 0$.

Since $\hat{\beta} = \beta$ if, and only if, *4*-Jac holds for $\mathcal{R}$, we conclude that *4*-Jac holds for $\mathcal{R}$ if, and only if, it holds for $\acute{\mathcal{R}}$. It remains for us to show that, for ϵ sufficiently small $\acute{\mathcal{G}}$ is maximal. For if $\acute{\mathcal{G}}$ is maximal, then via lemma 3.3 and lemma 3.4, $A0^{\flat}$ holds if, and only if, $\operatorname{rank}\{\acute{v}^{(x,w)}, \acute{v}^{(x,z)}, \acute{v}^{(z,w)}\} = 2$.

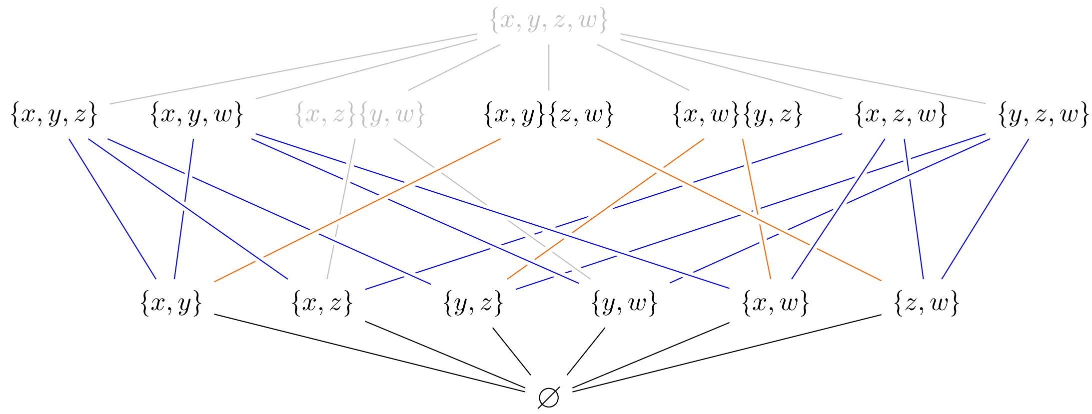


Figure 3: The intersection semilattices $\acute{\mathcal{L}}$ and $\acute{\mathcal{L}}_{++} = \acute{\mathcal{L}} - \{A^Y, \acute{A}^{\{x,z\}\{y,w\}}\}$ when $\#\mathbb{T} = 2$, $\hat{\beta} = \beta$ and ϵ is sufficiently small but distinct from zero.

By Zaslavski's theorem, it suffices to show that every member of the intersection lattice $\acute{\mathcal{L}}$, other than the center $\acute{A}^{\{x,y,z,w\}}$, has nonempty intersection with $\mathbb{R}_{++}^{\mathbb{T}^{\mathfrak{f}}}$. We now make explicit the dependence of the generalization $\acute{\mathcal{R}}$ on our choice of $L \in N_J$, though we do so indirectly via ϵ. Let $d^{\epsilon} = \max\{\mathrm{d}(\acute{J}, \acute{A}) : \acute{A} \in \acute{\mathcal{L}}^{\epsilon}\}$, where $\mathrm{d}(\acute{J}, \acute{A})$ is the minimum (Euclidean) distance between $\acute{J}$ and the (closed) linear subspace $\acute{A}$ of $\mathbb{R}^{\mathbb{T}^{\mathfrak{f}}}$. Note that for $\epsilon = 0$, we obtain a central arrangement of the form of lemma 3.2 with $d^{\epsilon} = 0$. Moreover, since the Euclidean metric is continuous in its arguments, and, for every ϵ, $\acute{\mathcal{L}}^{\epsilon}$ is finite, the map $\epsilon \mapsto d^{\epsilon}$ is continuous and $\lim_{\epsilon \to 0} d^{\epsilon} = 0$. Thus, for sufficiently small $\epsilon \neq 0$, every $\acute{A} \in \acute{\mathcal{L}}^{\epsilon}$ intersects $\mathbb{R}_{++}^{\mathbb{T}^{\mathfrak{f}}}$.

Remark. *We note that the above arguments apply without modification to the case where $\#\mathcal{H} = 4, 5$. Consider, for example, the Hasse diagram of fig. 3. That case arises when $u^{(x,z)}$ and $u^{(y,w)}$ are collinear, so that $A^{\{xz\}\{y,w\}}$ is a hyperplane of dimension $\#\mathbb{T} - 1$. Assuming the same construction, with $\epsilon \neq 0$, so that $\acute{u}^{(\cdot,\cdot)}$ has rank 3 and, via proposition 5, $\acute{u}^{(x,z)}$ and $\acute{u}^{(y,w)}$ are linearly independent. Thus $\acute{A}^{\{xz\}\{y,w\}}$ is of dimension $\#\mathbb{T}^{\mathfrak{f}} - 2 = \#\mathbb{T} - 1$. Thus, $\acute{A}^{\{x,z\}\{y,w\}} = A^{\{x,z\}\{y,w\}} \times \{0\}$ which belongs to the boundary of $\mathbb{R}_{++}^{\mathbb{T}^{\mathfrak{f}}}$. At $\epsilon = 0$, $\acute{A}^{\{x,z\}\{y,w\}}$ increases by one dimension and the upper two levels of the Hasse diagram collapse to equal $A^{\{x,y,z,w\}}$.*

Appendix D. Proofs of lemmas, observations and propositions

Proof of Observation 1. Fix $Y \subseteq X$ nonempty and $\acute{\mathcal{R}}$ regular. W.l.o.g., take $C \in \mathbb{D}^{\mathfrak{f}} - \mathbb{D}$, so that C contains at least one copy of $\mathfrak{f}$. For any $c \in C \cap [\mathfrak{f}]$, the fact that $\acute{\mathcal{R}}$ is regular implies that $c \sim^{\acute{\mathcal{R}}} c_1$ for some $c_1 \in \mathbb{C}$. The richness assumption ensures that we may choose c_1 from the complement of C. Then, since neither c nor c_1 belong to $C_1 \stackrel{\text{def}}{=} C - \{c\}$, $c \sim^{\acute{\mathcal{R}}} c_1$ implies $\acute{\mathcal{R}}_C = \acute{\mathcal{R}}_{C_1 \cup \{c_1\}}$. If c is the unique member of $C \cap [\mathfrak{f}]$, then the proof is complete. Otherwise, using the fact that C is finite, we may proceed by induction until we obtain a set C_n such that $C_n \cap [\mathfrak{f}]$ is empty and $D \stackrel{\text{def}}{=} C_n \cup \{c_1, \ldots, c_n\}$ belongs to $\mathbb{D}$. Part 1 of definition 1 then implies $\acute{\mathcal{R}}_D = \preceq_D \cap\, Y^2$, so that, since $\mathcal{R}$ is improper, $\acute{\mathcal{R}}_D = \mathcal{R}_D$. Finally, since $C \sim^{\acute{\mathcal{R}}} D$, $\acute{\mathcal{R}}_C = \acute{\mathcal{R}}_D$, as required. □

Proof of Corollary 1. This follows from theorem 3, lemma 2.3 and the fact that, via lemma 3.1, $\not{4}$-stability implies A0–A3 when $\#\mathbb{T} \leqslant \infty$. □

Proof of Lemma 1.1. We show that there exists a canonical embedding (a structure preserving injection) of $\text{nov}(Y, \preceq_{\mathbb{J}})$ into $\text{nov}(Y, \preceq_{\mathbb{D}})$. The fact that this map is also surjective follows from the fact that $\text{nov}(Y, \preceq_{\mathbb{D}})$ can be embedded in $\text{nov}(Y, \preceq_{\mathbb{J}})$ in precisely the same way. The proof that the two sets of regular generalizations are isomorphic follows via a similar argument plus the observation that every Y-generalization is either regular or novel.

Take $\mathcal{R} \in \text{nov}(Y, \preceq_{\mathbb{J}})$ and define $\hat{\mathcal{R}} = \langle \hat{\mathcal{R}}_C : C \in \mathbb{D}^{\mathfrak{f}} \rangle$ via the property: for each $C \in \mathbb{D}^{\mathfrak{f}}$, $\hat{\mathcal{R}}_C \stackrel{\text{def}}{=} \mathcal{R}_J$ if, and only if, $L_C = L_J$, where, as before, $t \mapsto L_C(t)$ counts the number of cases of type t in C and $L_J = \kappa_J J \in \mathbb{L}^{\mathfrak{f}}$ for some minimal $\kappa_J \in \mathbb{Z}_+$. Now, for any $\mathcal{R}' \neq \mathcal{R}$ in $\text{nov}(Y, \preceq_{\mathbb{J}})$, there exists $J \in \mathbb{J}^{\mathfrak{f}}$ such that $\mathcal{R}'_J \neq \mathcal{R}_J$. If we define $\hat{\mathcal{R}}'$ analogously, so that it is equivalent to $\mathcal{R}'$, then $\hat{\mathcal{R}}' \neq \hat{\mathcal{R}}$. As a consequence, the canonical mapping $\mathcal{R} \mapsto \hat{\mathcal{R}}$ is injective. If we can show that $\hat{\mathcal{R}}$ does in fact belong to $\text{nov}(Y, \preceq_{\mathbb{D}})$, then we have constructed the required embedding. The fact that $\hat{\mathcal{R}}$ satisfies 2 and 1 of definition 1 follows immediately from definition 2. The proof that item 3 of definition 1 holds is as follows. Take any $c, c' \in \mathbb{C}^{\mathfrak{f}}$ and $D \in \mathbb{D}^{\mathfrak{f}}$ such that $c \sim^{\star} c'$ and $c, c' \notin D$. First, observe that $D \cup \{c\} \sim^{\star} D \cup \{c'\}$, and moreover, for some $t \in \mathbb{T}^{\mathfrak{f}}$ we have $c, c' \in t$. Then, for every $t \in \mathbb{T}^{\mathfrak{f}}$, $|D \cup \{c\}| = |D \cup \{c'\}| = L$ for some $L \in \mathbb{L}^{\mathfrak{f}} \cap \mathbb{J}^{\mathfrak{f}}$. Thus $\hat{\mathcal{R}}_{D\cup\{c\}} = \hat{\mathcal{R}}_{D\cup\{c'\}}$, as required for $\hat{\mathcal{R}}$ to be a generalization of $\preceq_{\mathbb{D}}$. Finally, via definition 2, the definition of a novel generalization ensures that the induced equivalence relation $\sim^{\mathcal{R}}$ on $\mathbb{C}^{\mathfrak{f}}$ satisfies $c \not\sim^{\mathcal{R}} \mathfrak{f}$ for every $c \in \mathbb{C}$. Since $\sim^{\hat{\mathcal{R}}}$ inherits this property, $\hat{\mathcal{R}}$ is novel. □

Proof of Lemma 1.3. Fix $\mathcal{R}_{\mathfrak{J}} \equiv \hat{\mathcal{R}}_{\mathfrak{D}}$ and assume that $\hat{\mathcal{R}}_{\mathfrak{D}}$ satisfies A2. We show that $\mathcal{R}_{\mathfrak{J}}$ satisfies A2$^\flat$. Fix $x, y \in Y$ and $J \in \mathfrak{J}$ such that $x \mathrel{\mathcal{R}_J} y$ and $x \mathrel{\mathcal{R}_{J'}} y$. Fix $\lambda, \mu \in \mathbb{Q}_{++}$ and let κ be the smallest positive integer such that both $L \overset{\text{def}}{=} \kappa\lambda J$ and $L' \overset{\text{def}}{=} \kappa\mu J'$ belong to $\mathfrak{L}$. Then, by lemma 1.2, we have both $x \mathrel{\mathcal{R}_L} y$ and $x \mathrel{\mathcal{R}_{L'}} y$. Moreover, for D, D' such that $L_D = L$ and $L_{D'} = L'$, we have $x \mathrel{\hat{\mathcal{R}}_D} y$ and $x \mathrel{\hat{\mathcal{R}}_{D'}} y$ and, by A2, $x \mathrel{\hat{\mathcal{R}}_{D\cup D'}} y$. Finally, since $L_D + L_{D'} = \kappa(\lambda J + \mu J')$, one further application of lemma 1.2 yields $x \mathrel{\mathcal{P}_{\lambda J+\mu J'}} y$, as required for A2$^\flat$.

The proof that "A2 implies A2$^\flat$" is *mutatis mutandis* a special case of the above argument and ommitted. We now assume $\hat{\mathcal{R}}_{\mathfrak{D}}$ satisfies A2 and A3 and prove that $\mathcal{R}_{\mathfrak{J}}$ satisfies A3$^\flat$. Fix $x, y \in X$ such that $x \mathrel{\mathcal{P}_J} y$ for some $J \in \mathbb{J}$ and take any $J' \in \mathfrak{J}$. Then, by the construction of $\mathcal{R}_{\mathfrak{J}}$, there exists $L, L' \in \mathbb{L}$ such that $jJ = L$ and $j'J' = L'$ for some $j, j' \in \mathbb{Z}_{++}$. By lemma 1.2, $\mathcal{R}_L = \mathcal{R}_J$ and $\mathcal{R}_{L'} = \mathcal{R}_{J'}$. Moreover, by construction, for some D and D' such that $L_D = L$ and $I_{D'} = L'$, $\hat{\mathcal{R}}_D = \mathcal{R}_J$ and $\hat{\mathcal{R}}_{D'} = \mathcal{R}_{J'}$. We therefore conclude that $x \mathrel{\hat{\mathcal{P}}_D} y$, so that A3 implies the existence of $\kappa \in \mathbb{Z}_{++}$ and $\{D_l : D_l \sim^{\hat{\mathcal{R}}} D\}_1^\kappa$ such that $x \mathrel{\hat{\mathcal{P}}_{D_1\cup\cdots\cup D_\kappa\cup D'}} y$. Then, by the construction of $\mathcal{R}_{\mathfrak{J}}$, $x \mathrel{\mathcal{P}_{\kappa L_D + L_{D'}}} y$. Let $\nu \overset{\text{def}}{=} \frac{1}{\kappa j + j'}$ and take $\lambda = \nu j'$, so that $0 < \lambda < 0$ and $1 - \lambda = \nu\kappa j$. In fact, since $\lambda \in \mathbb{Q}$, we have

$$K \overset{\text{def}}{=} (1-\lambda)J + \lambda J' \in \mathfrak{J}.$$

Simplifying, we obtain $K = \nu(\kappa L + L')$. Since $\nu \in \mathbb{Q}_{++}$ and $\kappa L + L' \in \mathfrak{J}$, lemma 1.2 implies $\mathcal{R}_K = \mathcal{R}_{\kappa L + L'}$. This allows us to conclude that $x \mathrel{\mathcal{P}_K} y$. Finally, take any $\mu \in \mathbb{Q} \cap (0, \lambda)$. From basic properties of the real numbers, there exists $\xi < 1$ such that $\mu = \xi\lambda$ and, moreover, ξ is rational. Next, note that the definition of K implies $\xi(K - J) = \xi\lambda(J' - J)$. Adding J to each side of the latter and applying the definition of μ yields

$$(1-\xi)J + \xi K = (1-\mu)J + \mu J'.$$

Then, since $x \mathrel{\mathcal{P}_J} y$ and $x \mathrel{\mathcal{P}_K} y$, A2$^\flat$ implies $x \mathrel{\mathcal{P}_{(1-\mu)J+\mu J'}} y$, as required for A3$^\flat$.

Conversely, we now assume that $\mathcal{R}_{\mathfrak{J}}$ satisfies A2$^\flat$ and A3$^\flat$ and prove that A3 holds. Take $D, D' \in \mathbb{D}$ such that $x \ \hat{\mathcal{P}}_D \ y$ and any other $D' \in \mathbb{D}$. Let $L = L_D$ and $L' = L_{D'}$. Then, by construction, $x \ \mathcal{P}_L \ y$ and, by A3$^\flat$, there exists $\lambda \in \mathbb{Q} \cap (0,1)$ such that $x \ \mathcal{P}_{(1-\mu)L+\mu L'} \ y$. Then, since μ is rational, $\mu = {}^{j}\!/_{k}$ for some $j, k \in \mathbb{Z}_{++}$. Let $q := (1-\mu)/\mu = (k-j)/j$ and let $\kappa = jq$, so that $\kappa = k-j$. The fact that $0 < \mu < 1$ ensures that $\kappa \in \mathbb{Z}_{++}$. To complete the proof, we show that $x \ \mathcal{P}_{\kappa L+L'} \ y$, for then the existence of $D_1, \ldots, D_\kappa$ such that $x \ \hat{\mathcal{P}}_{D_1 \cup \cdots \cup D_\kappa \cup D'}$ immediately follows. Together $x \ \mathcal{P}_{(1-\mu)L+\mu L'} \ y$ and lemma 1.2 imply $x \ \mathcal{P}_{qL+L'} \ y$. Similarly, together $x \ \mathcal{P}_L \ y$ and lemma 1.2 imply $x \ \mathcal{P}_{(j-1)qL} \ y$. Then, since $(j-1)qL + (qL + L') = jqL + L'$ and $\kappa = jq$, an application of A2$^\flat$ yields $x \ \mathcal{P}_{\kappa L+L'} \ y$, as required. □

Proof of Proposition 2. Let $\mathcal{Y} = \{Y_\alpha \subseteq Y : \#Y_\alpha = 2\}$ be the collection of distinct (unordered) pairs in Y. For every $Y_\alpha \in \mathcal{Y}$, A0 holds simply because Y_α is of cardinality two; moreover, on Y_α, *4*-diversity is equivalent to *2*-diversity. This allows us to apply theorem 2 of GS03.[14] (Part **??** of the present lemma follows from claim 5 of the proof of lemma 1 of GS03.) Thus, on Y_α, A1$^\flat$–A3$^\flat$ and *2*-diversity hold if, and only if, there exists of a matrix $v_\alpha^{(\cdot,\cdot)} : Y_\alpha^2 \times \mathfrak{T} \to \mathbb{R}$ with rows satisfying condition (i) of the present lemma. In particular, GS03 yields a matrix representation $v_\alpha : Y_\alpha \times \mathfrak{T} \to R$. For $x, y \in Y_\alpha$, take $v_\alpha^{(x,y)} = -v_\alpha(x, \cdot) + v_\alpha(y, \cdot)$. For condition (ii), note that via theorem 2 of GS03, the matrix v_α is 2-diversified if, and only if, *4*-diversity holds for $\acute{\mathcal{R}}$ on Y_α. Thus, for every $\lambda \in \mathbb{R}$, $v_\alpha(x) \nleqslant \lambda v_\alpha(y)$ if, and only if, *2*-diversity. Take $\lambda = 1$ and observe that this implies $v_\alpha^{(x,y)}(s) < 0 < v_\alpha^{(x,y)}(t)$ for some $s, t \in \mathfrak{T}$.

To extend this to arbitrary $Y \subseteq X$, note that, for some A', $\bigcup\{\mathcal{Y}_\alpha : \alpha \in A'\} = Y$. For an arbitrary function f, let $\operatorname{gr} f$ denote the graph of that function. To obtain the desired matrix $\acute{v}^{(\cdot,\cdot)} : Y^2 \times \mathbb{T} \to \mathbb{R}$, take

$$\operatorname{gr} \acute{v}^{(\cdot,\cdot)} = \bigcup\{\operatorname{gr} v_\alpha^{(\cdot,\cdot)} : \alpha \in A'\}. \tag{24}$$

It remains for us to prove the characterisation of novel generalizations. Fix arbitrary $t \neq \mathfrak{f}$. Then definition 2 implies the existence of $J \in \mathbb{J}$ and $L = J \times 0 \in \mathbb{J}^{\mathfrak{f}}$

[14] **?** explicitly prove their theorem 1 holds for the case of arbitrary X and $\mathbb{T}$. But although their theorem 2 is stated and proved in the first step (the case where $X, \mathbb{T} < \infty$) of the proof of their theorem 1, steps 2 and 3 of that proof apply equally to their theorem 2.

such that $\acute{\mathcal{R}}_{L+\delta_t} \neq \acute{\mathcal{R}}_{L+\delta_{\mathfrak{f}}}$. W.l.o.g., consider the case where, for some $x, y \in Y$, it holds that both $y \; \acute{\mathcal{R}}_{L+\delta_t} \; x$ and $x \; \acute{\mathcal{P}}_{L+\delta_{\mathfrak{f}}} \; y$. Equivalently,

$$\langle \acute{v}^{(x,y)}, L+\delta_t \rangle \leqslant 0 < \langle \acute{v}^{(x,y)}, L+\delta_{\mathfrak{f}} \rangle$$

which, via linearity of $\langle \acute{v}^{(x,y)}, \cdot \rangle$, we may rearrange to obtain

$$\acute{v}^{(x,y)}(t) \leqslant -\langle \acute{v}^{(x,y)}, L \rangle < \acute{v}^{(x,y)}(\mathfrak{f}). \tag{25}$$

Thus, $\acute{v}^{(x,y)}(t) \neq v^{(x,y)}(\mathfrak{f})$, as required for the lemma. □

Proof of Lemma 2.1. Let A4$^\flat$ hold on Y. Since $v^{(\cdot,\cdot)}$ is a 2-diverse pairwise representation, $v^{(x,z)}, v^{(y,z)} \nleqslant 0$. By A4$^\flat$, one of $G_+^{(x,z)}$ and $G_+^{(z,x)}$ contains both J, L such that

$$\langle v^{(y,z)}, L \rangle < 0 < \langle v^{(y,z)}, J \rangle.$$

W.l.o.g., suppose J, L belongs to $G_+^{(x,z)}$. Then $\langle v^{(x,z)}, \cdot \rangle$ is positive on $\{L, J\}$, so that, for every $\lambda \in \mathbb{R}$ $v^{(x,z)} \neq \lambda v^{(y,z)}$, as required.

Conversely, let $x, y, z \in Y$ be such that $v^{(x,z)}$ and $v^{(y,z)}$ are noncollinear. Then $H_{++}^{\{x,z\}} \neq H_{++}^{\{y,z\}}$, and there exists $L \in H_{++}^{\{x,z\}} - H_{++}^{\{y,z\}}$. W.l.o.g., therefore, suppose $L \in H_{++}^{\{x,z\}} \cap G_{++}^{(y,z)}$. Since $v^{(\cdot,\cdot)}$ is 2-diverse, there exists $s, t \in \mathbb{T}$ such that $v^{(x,z)}(s) < 0 < v^{(x,z)}(t)$. Noting that $L \in \mathbb{R}_{++}^{\mathfrak{T}}$, so that $v^{(x,z)} \neq \delta_s, \delta_t$, let ψ_s and ψ_t be the convex paths from L to δ_s and δ_t respectively. For sufficiently small $\lambda > 0$, $\langle v^{(y,z)}, \psi_{s'}(\lambda) \rangle$ remains positive for $s' = s, t$ and, moreover, since $L \in H_{++}^{\{x,z\}}$,

$$\langle v^{(x,z)}, \psi_s(\lambda) \rangle < 0 < \langle v^{(x,z)}, \psi_t(\lambda) \rangle.$$

Finally, since L has finite support, a finite sequence of perturbations of the elements of $\psi_s(\lambda)$ and $\psi_t(\lambda)$ yields (rational-valued) members of $\mathfrak{I}$ with the same properties, as required for A4$^\flat$. □

Proof of Proposition 3. When $X = 2$, A4$^\flat$ and A4$'^\flat$ are identical to *2*-diversity. Let $Y = \{x, y, z\} \subseteq X$ and let $\mathcal{R}$ denote the improper Y-generalization of $\preceq_{\mathbb{J}}$. We begin by assuming A4$^\flat$ and showing that $\#Y + 1 = 4 \leqslant \#\text{total}(\mathcal{R})$. Via lemma 2.1, there are three distinct hyperplanes $H_{++}^{\{x,y\}}, H_{++}^{\{y,z\}}$ and $H_{++}^{\{x,z\}}$ in the associated arrangement $\mathcal{H}_{++}$. Then, as in example 6, $A^{\varnothing} = S$ is the unique element of $\mathcal{L}_{++}$ that lies below each member of $\mathcal{H}_{++}$. Thus, via eq. (33), $\boldsymbol{\mu}(A^{\varnothing}) = 1$ and

$\boldsymbol{\mu}(A) = -\boldsymbol{\mu}(A^{\varnothing})$ for all three hyperplanes $A \in \mathcal{H}_{++}$. Thus, Zaslavski's theorem implies that $\#\mathcal{G}_{++}$ is bounded below by 4. Thus total$(\mathcal{R}) \geqslant 4$, and since, for every Y-generalization $\acute{\mathcal{R}}$, $\#\text{total}(\acute{\mathcal{R}}) \geqslant \#\text{total}(\mathcal{R})$, A4$'^{\flat}$ holds.

Conversely, suppose A4$'^{\flat}$ holds and, once again let $\mathcal{R}$ denote the improper Y-generalization of $\preceq_{\mathbb{J}}$, so that total$(\mathcal{R}) \geqslant \#Y = 3$. Now A4$'^{\flat}$ implies *2*-diversity, so that, via proposition 2, there exists a 2-diverse matrix representation with associated arrangement $\mathcal{H}_{++}$. It is not the case that $|\mathcal{H}_{++}| = 1$, for this would imply that $\#\text{total}(\mathcal{R}) = 2$. W.l.o.g., suppose $H^{\{x,y\}}_{++} \neq H^{\{y,z\}}_{++}$. Observe that A0$^{\flat}$ then implies $H^{\{x,z\}}_{++} \neq H^{\{x,y\}}_{++}$ and $H^{\{x,z\}}_{++} \neq H^{\{y,z\}}_{++}$. This implies that $v^{(x,y)}$, $v^{(y,z)}$ and $v^{(x,z)}$ are pairwise noncollinear. Finally, an application of lemma 2.1 then yields A4$^{\flat}$. □

Proof of Lemma 2.3. Note that, when $1 \leqslant |X| \leqslant 2$, *4*-Jac holds vacuously and $\preceq_{\mathbb{J}}$ has a Jacobi representation via proposition 2. For general X, the fact that *4*-Jac is necessary for $\preceq_{\mathbb{J}}$ to have a Jacobi representation follows simply because if the Jacobi identity holds on X, then it holds on every $Y \subseteq X$. For the sufficiency of *4*-Jac, we proceed by induction. As in lemma 3 and claim 9 of GS03, we assume that X is well-ordered.

In the case that $|X| \leqslant 4$, we only need to show that $v^{(\cdot,\cdot)}$ is unique. W.l.o.g., we take the initial step in our induction argument to satisfy $|X| = 4$. Let $\mathbf{v}^{(\cdot,\cdot)}$ denote any other Jacobi representation of $\preceq_{\mathbb{J}}$. By proposition 2, for every distinct $x, y \in Y^2$, there exists $\lambda^{\{x,y\}} > 0$ such that $\mathbf{v}^{(x,y)} = \lambda^{\{x,y\}} v^{(x,y)}$. We need to show that $\lambda^{\{x,y\}} = \lambda$ for every distinct $x, y \in Y$. Let $Y = \{x, y, z, w\}$. By lemma 2.1, the set $\{v^{(x,y)}, v^{(x,z)}, v^{(x,w)}\}$ is pairwise noncollinear. Then, since the Jacobi identity holds for both $v^{(\cdot,\cdot)}$ and $\mathbf{v}^{(\cdot,\cdot)}$, we derive the equation

$$(1 - \lambda^{\{x,y\}})v^{(x,y)} + (1 - \lambda^{\{y,z\}})v^{(y,z)} = (1 - \lambda^{\{x,z\}})v^{(x,z)} \tag{26}$$

Suppose that $1 - \lambda^{\{y,z\}} = 0$. Then, either the other coefficients in eq. (26) are both equal to zero (and our proof is complete), or we obtain a contradiction of lemma 2.1. Thus, $1 - \lambda^{\{y,z\}}$ is nonzero and we may divide through by this term and solve for $v^{(y,z)}$. First note that, since $v^{(\cdot,\cdot)}$ is a Jacobi representation, $v^{(y,x)} + v^{(x,y)} = v^{(y,y)} = 0$. Then, since $v^{(y,x)} = -v^{(x,y)}$,

$$v^{(y,z)} = \frac{1 - \lambda^{xy}}{1 - \lambda^{yz}} v^{(y,x)} + \frac{1 - \lambda^{xy}}{1 - \lambda^{yz}} v^{(x,z)}.$$

We then conclude that both of the coefficients in the latter equation are equal to one. (This follows from linear independence of $v^{(y,x)}$ and $v^{(x,z)}$ together with the Jacobi identity $v^{(y,z)} = v^{(y,x)} + v^{(x,z)}$.) Thus, $\lambda^{\{x,y\}} = \lambda^{\{y,z\}} = \lambda^{\{x,z\}}$, as required. Repeated application of the same argument to the remaining Jacobi identities yields the desired conclusion, $\mathbf{v}^{(\cdot,\cdot)} = \lambda v^{(\cdot,\cdot)}$.

For the inductive step, take Y to be an initial segment of X. By the induction hypothesis, there exists a Jacobi representation $\mathbf{v}^{(\cdot,\cdot)} : Y^2 \times \mathbb{T} \to \mathbb{R}$ of the improper Y-generalization $\mathcal{R} = \preceq_{\mathbb{J}} \cap Y^2$ that is suitably unique.

Claim 3.5.1. *For every $w \in X - Y$ and $W \overset{\text{def}}{=} Y \cup \{w\}$, there exists a Jacobi representation $\hat{v}^{(\cdot,\cdot)} : W^2 \times \mathbb{T} \to \mathbb{R}$ of the improper W-generalization $\hat{\mathcal{R}}$.*

Proof of claim 3.5.1. Via lemma 2.1, there exists a conditionally 2-diverse pairwise representation $u^{(\cdot,\cdot)}$ of $\preceq_{\mathbb{J}}$. Fix any four distinct elements x, x', y, z in Y. Proposition 2 implies the existence of $\phi, \phi' \in \mathbb{R}_{++}$ such that $\phi u^{(x,z)} = \mathbf{v}^{(x,z)}$ and $\phi' u^{(x',z)} = \mathbf{v}^{(x',z)}$. Let $Z = \{x, y, z, w\}$ and $Z' = \{x', y, z, w\}$. Since *3*-Jac holds, there exist positive scalars $\alpha, \beta, \hat{\beta}, \gamma, \sigma$ and τ such that

$$\alpha u^{(x,w)} + \beta u^{(w,y)} = \gamma u^{(x,y)}, \tag{27}$$

$$\hat{\beta} u^{(y,w)} + \sigma u^{(w,z)} = \tau u^{(y,z)}, \text{ and} \tag{28}$$

$$\gamma u^{(x,y)} + \tau u^{(y,z)} = \mathbf{v}^{(x,z)}. \tag{29}$$

Moreover, *4*-Jac ensures that we may take $\beta = \hat{\beta}$. Since $u^{(\cdot,\cdot)}$ is conditionally 2-diverse, $\{u^{(x,y)}, u^{(y,z)}\}$ is linearly independent, and the linear system eq. (29) in the unknowns γ and τ has a unique solution. This, together with the induction hypothesis (which yields $\mathbf{v}^{(x,y)} + \mathbf{v}^{(y,z)} = \mathbf{v}^{(x,z)}$) implies that $\gamma u^{(x,y)} = \mathbf{v}^{(x,y)}$ and $\tau u^{(y,z)} = \mathbf{v}^{(y,z)}$. Similarly, for Z', *4*-Jac yields $\alpha', \beta', \sigma', \gamma', \tau' > 0$ such that

$$\alpha' u^{(x',w)} + \beta' u^{(w,y)} = \gamma' u^{(x',y)}, \tag{30}$$

$$\beta' u^{(y,w)} + \sigma' u^{(w,z)} = \tau' u^{(y,z)}, \text{and} \tag{31}$$

$$\gamma' u^{(x',y)} + \tau' u^{(y,z)} = \mathbf{v}^{(x',z)}. \tag{32}$$

As in the arguments involving γ and τ, the induction hypothesis yields $\gamma' u^{(x',y)} = \mathbf{v}^{(x',y)}$ and $\tau' u^{(y,z)} = \mathbf{v}^{(y,z)}$. We conclude that $\tau = \tau'$. Substituting for τ' in eq. (31) and appealing to linear independence of $\{u^{(y,w)}, u^{(w,z)}\}$ then yields the desired equalities $\beta = \beta'$ and $\sigma = \sigma'$.

As a consequence of the above argument, for every $y, z \in Y$, take $\hat{v}^{(y,w)}$ and $\hat{v}^{(w,z)}$ to be the unique vectors in $\mathbb{R}^{\mathbb{T}}$ that solve the equation $\hat{v}^{(y,w)} + \hat{v}^{(w,z)} = \mathbf{v}^{(y,z)}$. For every $y, z \in Y$, let $\hat{v}^{(y,z)} = \mathbf{v}^{(y,z)}$ and $\hat{v}^{(w,w)} = 0$. Then the matrix $\hat{v}^{(\cdot,\cdot)}$ with row vectors $\left\{\hat{v}^{(x,y)} : x, y \in W\right\}$ is a Jacobi representation of $\hat{\mathcal{R}}$. □

Our proof of claim 3.5.1 shows that the generalization to W holds for any initial subsegment of Y consisting of four elements. Our proof thereby accounts for the case where X is infinite and w is a limit ordinal. □

Appendix E. Online Appendix

For any pair of vectors $\acute{v}, J : \mathfrak{T} \to \mathbb{R}$ such that $J \in \mathfrak{J}$, the linear operator

$$J \mapsto \langle \acute{v}, J \rangle \stackrel{\text{def}}{=} \sum_{\{t : J(t) > 0\}} v(t) \cdot J(t).$$

is well-defined and real-valued by virtue of the fact that J has finite support.

In our proof, we build on GS03 to directly prove all results regardless of the cardinality of X and $\mathbb{T}$. To facilitate this approach, we first introduce the notion of an essentialization.

Let $\mathbb{R}^{\oplus\mathfrak{T}}$ denote the vectors in $\mathbb{R}^{\mathfrak{T}}$ that have finite support and observe that $\mathfrak{J} \subseteq \mathbb{R}^{\oplus\mathfrak{T}}$ is the dense subset of rational vectors. Given a vector $\acute{v} : \mathfrak{T} \to \mathbb{R}$, we associate the following subsets of $\mathbb{R}^{\oplus\mathfrak{T}}$: $\acute{H} = \{J : \langle \acute{v}, J \rangle = 0\}$, $\acute{G} = \{J : \langle \acute{v}, J \rangle > 0\}$, and $\acute{F} = \{J : \langle \acute{v}, J \rangle \geqslant 0\}$. For any finite collection $\acute{\mathcal{V}} = \{\acute{v}_1, \ldots, \acute{v}_n\}$ of such vectors, let $\acute{\mathcal{H}}$ denote the associated collection or *arrangement* of hyperplanes in $\mathbb{R}^{\oplus\mathfrak{T}}$. let $\acute{S} = \operatorname{span} \acute{\mathcal{V}}$ denote the $\acute{\mathbf{r}}$-dimensional linear span of $\acute{\mathcal{V}}$. Then $\acute{S}$ is a well-defined inner-product space in its own right and let $\langle \cdot, \cdot \rangle_{\acute{S}} : \acute{S}^2 \to \mathbb{R}$ denote the inner product. Let $\acute{p} : \mathbb{R}^{\oplus\mathfrak{T}} \to \acute{S}$ denote the orthogonal projection. Then observe that, for every $i = 1, \ldots, n$, and $J \in \mathbb{R}^{\oplus\mathfrak{T}}$, $\langle \acute{v}_i, J \rangle = \langle \acute{v}_i, \acute{p}(J) \rangle$. Moreover, note that, since $\acute{v}_i \in S$, $\langle \acute{v}_i, J \rangle = \langle \acute{v}_i, \acute{p}(J) \rangle_{\acute{S}}$. The *essentialization* $\mathcal{H}_{\acute{S}}$ of the arrangement $\acute{\mathcal{H}}$ is the arrangement we obtain by orthogonally projecting $\acute{\mathcal{H}}$ onto $\acute{S}$. That is, for every $H_{\acute{S}} \in \mathcal{H}_{\acute{S}}$, there exists $\acute{H} \in \acute{\mathcal{H}}$ such that $\acute{H} = \acute{p}^{-1}(H_{\acute{S}})$. In the literature on arrangements of hyperplanes, it is common to work with the essentialization of an arrangement by default. We therefore identify $\acute{\mathcal{H}}$ and $\mathcal{H}_{\acute{S}}$ and suppress reference to the latter subscript whenever no possible confusion may arise. The main benefit of essentializations is that they will allow us to work in finite dimensions whenever we consider a finite subset $Y \subseteq X$: regardless of the cardinality of $\mathfrak{T}$.

Our domain of interest is $\mathbb{R}_{+}^{\oplus\mathfrak{T}}$ and not the whole of $\mathbb{R}^{\oplus\mathfrak{T}}$. In order to be able to apply results from the literature on hyperplane arrangements without adjusting for boundaries, we find it useful to work within the relative interior $\acute{S}_{++}$ of $\acute{S}_{+} = \acute{p}(\mathbb{R}_{+}^{\oplus\mathfrak{T}})$. Observe $\acute{S}_{++}$ is an open subset of the $\acute{\mathbf{r}}$-dimensional linear space $\acute{S}$. For $\acute{H} \in \mathcal{H}$, take $\acute{H}_{++} = \acute{H} \cap \acute{S}_{++}$ to be the (strictly) *positive* null-space of $\langle \acute{v}, \cdot \rangle_{\acute{S}}$ and, for $0 \nleqslant \acute{v} \nleqslant 0$, $\acute{G}_{++}$ and $\acute{F}_{++}$ are, respectively, the open and closed half-spaces of $\acute{S}_{++}$ associated with $\acute{v}$. (For such $\acute{v}$, we also refer to $\acute{H}_{++}$ as a hyperplane in $\acute{S}_{++}$.) We refer to the non-negative counterpart of these sets as $\acute{H}_{+}, \acute{G}_{+}$ and $\acute{F}_{+}$. For any finite $Y \in 2^X$ consider the matrix $\acute{v}^{(\cdot,\cdot)} : Y^2 \times \mathfrak{T} \to \mathbb{R}$. For a given $x, y \in Y$ and row $\acute{v}^{(x,y)} : \mathfrak{T} \to \mathbb{R}$ of $\acute{v}^{(\cdot,\cdot)}$, the associated sets are $\acute{H}_{++}^{\{x,y\}}$, $\acute{G}_{++}^{(x,y)}$ and $\acute{F}_{++}^{(x,y)}$ respectively.

From the mathematics of hyperplane arrangements, the main result to which we extensively appeal is Zaslavsky's theorem. For any given generalization $\mathcal{R}$, Zaslavsky's theorem allows us to use information about the intersections of hyperplanes in the arrangement to identify $\#\text{total}(\mathcal{R})$. It does so by counting the collection $\mathcal{G}_{++}$ of open and connected subsets of $\mathbb{R}^{\mathfrak{T}} - \bigcup\{H_{++} : H_{++} \in \mathcal{A}\}$ are called the *chambers* or *regions* of the arrangement. In the present setting, each chamber corresponds to a a complete, antisymmetric and reflexive, but possibly intransitive, *CAR ranking* of the elements of Y. Every CAR ranking R can be succinctly represented as an ordered tuple as in examples 1 and 2. For instance, take $Y = \{x, y, z\}$ and $x \, R \, y \, R \, z$, then the corresponding tuple

$$l = \begin{cases} (x, y, z) & \text{if } R \text{ is transitive} \\ (x, y, z, x) & \text{if } R \text{ is intransitive.}^{15} \end{cases}$$

The notation generalizes without exception to sets of cardinality 4. For example (x, y, z, x, w) represents the CAR ranking that is intransitive over $\{x, y, z\}$ and such that w dominates every other member.

The intersection semilattice of any arrangement $\mathcal{A}$ is the partially ordered (*by reverse inclusion*) set $\mathcal{L}$ of intersections of members of $\mathcal{A}$. The unique minimal element is obtained by taking the intersection $A^{\varnothing}$ over the empty subarrangement $\mathcal{A}^{\varnothing}$ of $\mathcal{A}$ to obtain the ambient space itself. That is $A^{\varnothing} = \mathbb{R}^{\mathfrak{T}}$ or $\mathbb{R}_{++}^{\mathfrak{T}}$, depending on whether we are considering the lattice $\mathcal{L}$ or the lattice $\mathcal{L}_{++}$ respectively. In GS03, as a consequence of *4*-diversity, $\mathcal{H}_{++}$ is always central. In our setting, it is only $\mathcal{H}$ that is guaranteed to be central. In general an arrangement is central, if, and only if, its intersection semilattice has a unique maximal element (**?**, proposition 2.3). Thus, if $\mathcal{H}_{++}$ is *centerless*, then $\mathcal{L}_{++}$ is a meet semilattice with multiple maxima: as in example 6. Extending our notation: if $Y = \{x, y, z, w\}$, then the unique intersection A^Y is the (nonempty) *center* of $\mathcal{A}^Y = \mathcal{H}$. By $A^{\{x,y,z\}}$, we mean the intersection over $\mathcal{A}^{\{x,y,z\}} \stackrel{\text{def}}{=} \{H^{\{i,j\}} : i \neq j \text{ in } \{x, y, z\}\}$. Finally, by $A^{\{x,y\}\{z,w\}}$, we mean the intersection over $\mathcal{A}^{\{x,y\}\{z,w\}} \stackrel{\text{def}}{=} \{H^{\{x,y\}}, H^{\{z,w\}}\}$.

Zaslavski's theorem provides two distinct methods for counting the number of regions in an arrangement. The first states that $\#\mathcal{G}$ *is equal to the sum of the absolute values of the Möbius function* $\boldsymbol{\mu} : \mathcal{L} \to \mathbb{Z}$ which is defined recursively via

$$\boldsymbol{\mu}(A) = \begin{cases} 1 & \text{if } A = A^{\varnothing} \\ -\sum\{\boldsymbol{\mu}(B) : A \subsetneq B\} & \text{otherwise.} \end{cases} \tag{33}$$

The above definition of Zaslavski's theorem is explicitly provided by **?**. Specialised to the present setting, the more common (see **???**) "rank" version of Zaslavski's theorem is

$$\#\mathcal{G} = \sum_{\substack{\mathcal{A} \subseteq \mathcal{H} \\ \mathcal{A} \text{ central}}} (-1)^{|\mathcal{A}| - \operatorname{rank}(\mathcal{A})},$$

where *central* means that $\bigcap\{H : H \in \mathcal{A}\}$ is nonempty, and $\operatorname{rank}(\mathcal{A})$ is the dimension of the space spanned by the normals to the hyperplanes in $\mathcal{A}$.

Example 6 (a comparison of $\mathcal{L}$ and $\mathcal{L}_{++}$)**.** *Let* $X = \{x, y, z\}$, $\mathbb{T} = \{s, t\}$, *and* $u^{(x,y)} = 1 \times -1$ *and* $u^{(y,z)} = 2 \times -1$ *denote vectors in* $\mathbb{R}^{\mathbb{T}}$*. Note that in this case* $S = \operatorname{span}\{u^{(x,y)}, u^{(y,z)}\}$ *coincides with* $\mathbb{R}^{\mathbb{T}}$*. We now apply the Jacobi identity and take* $u^{(x,z)} = u^{(x,y)} + u^{(y,z)} = 3 \times -2$ *and extend to the remaining pairs in* X^2 *using proposition 2. Since these vectors are pairwise noncollinear and* $\#\mathbb{T} = 2$*, the associated arrangement* $\mathcal{H}_{++} = \left\{H_{++}^{\{x,y\}}, H_{++}^{\{y,z\}}, H_{++}^{\{x,z\}}\right\}$ *consists of three pairwise disjoint lines that partition* $\mathbb{R}_{++}^{\mathbb{T}}$ *and* $\mathcal{G}_{++}$ *has cardinality* 4*. We now confirm this using Zaslavski's theorem.*

In the present setting, $A_{++}^{\varnothing} = \mathbb{R}_{++}^{\mathbb{T}}$*, and, via eq. (33),* $\boldsymbol{\mu}(A_{++}^{\varnothing}) = 1$*. Then, since* $\mathbb{R}_{++}^{\mathbb{T}}$ *is the unique element in* $\mathcal{L}_{++}$ *that (strictly) contains each hyperplane in* $\mathcal{H}_{++}$*,*

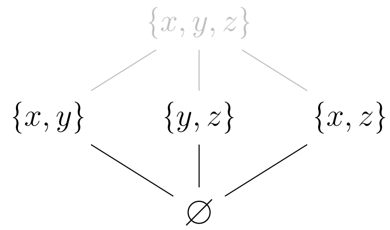


Figure 4: The intersection semilattice $\mathcal{L}_{++} = \mathcal{L} - A^{\{x,y,z\}}$.

eq. (33) *yields* $\boldsymbol{\mu}(A) = -\boldsymbol{\mu}(A^{\varnothing}_{++})$ *for each* $A \in \mathcal{H}_{++}$. *Now since the hyperplanes in* $\mathcal{H}_{++}$ *are pairwise disjoint, there are no further elements in* $\mathcal{L}_{++}$. *Thus*

$$|\mathcal{G}_{++}| = \sum_{A \in \mathcal{L}_{++}} |\boldsymbol{\mu}(A)| = 4.$$

In contrast, although the structure of $\mathcal{L}$ *is otherwise isomorphic to* $\mathcal{L}_{++}$, *since* $\{0\} \subset \mathbb{R}^{\mathbb{T}}$ *is a subset of every hyperplane in* $\mathcal{H}$, $\{0\}$ *is the center* $A^{\{x,y,z\}}$ *of* $\mathcal{H}$ *and the maximal element of* $\mathcal{L}$. *Via* eq. (33) *and the calculations of the previous paragraph, we obtain* $\boldsymbol{\mu}(A^{\{x,y,z\}}) = -\big(\boldsymbol{\mu}(A^{\varnothing}) - 3\boldsymbol{\mu}(A^{\varnothing})\big) = 2$. *Thus,*

$$\#\mathcal{G} = \sum_{A \in \mathcal{L}} |\boldsymbol{\mu}(A)| = 6 = 3!.$$

Remark 1 (The relationship between $\mathcal{L}$ and $\mathcal{L}_{++}$)**.** *Let* $\acute{\mathcal{R}}$ *be a* Y*-generalization with* 2*-diverse representation* $\acute{u}^{(\cdot,\cdot)}$. *Since, for every distinct* x *and* y *in* Y, $\acute{H}^{\{x,y\}}$ *contains the origin,* $\acute{\mathcal{H}}$ *is centered. As we see in* example 6, *this is not the case for* $\acute{\mathcal{H}}_{++}$ *where* $\acute{\mathcal{H}}_{++}$ *is centerless and each of its members is maximal in* $\acute{\mathcal{L}}_{++}$.

In GS03, *4*-diversity *guarantees that, for every* $Y \subseteq X$ *of cardinality* 2, 3 *or* 4, *the improper* Y*-generalization generates a centered arrangement in* $\mathbb{R}^{\mathbb{T}}_{++}$. *The fact that* $\mathbb{R}^{\mathfrak{T}}_{++}$ *is open in* $\mathbb{R}^{\mathfrak{T}}$ *ensures that the dimension of any* $L \in \acute{\mathcal{L}}$ *is equal to its counterpart* $L_{++} \in \acute{\mathcal{L}}_{++}$ *provided the latter exists. Thus,* $\acute{\mathcal{L}}_{++}$ *and* $\acute{\mathcal{L}}$ *are isomorphic if, and only if,* $\acute{\mathcal{H}}_{++}$ *is centered. For the same reason,* $\acute{\mathcal{G}}_{++}$ *and* $\acute{\mathcal{G}}$ *are isomorphic if, and only if,* $\acute{\mathcal{H}}_{++}$ *is centered.*

We now abstract a useful property from example 6.

Proposition 6 (polar opposite rankings)**.** *If* $\preceq_{\mathbb{J}}$ *satisfies* A1$^\flat$–A3$^\flat$ *and 2*-diversity, *then, for every* $Y \subseteq X$ *of cardinality* 3 *or* 4, *the improper* Y*-generalization* $\mathcal{R}$ *is such that, for some* $J, L \in \mathbb{J}$, $\mathcal{R}_J = \mathcal{R}_L^{-1}$ *belongs to* total$(\mathcal{R})$.

Proof. Fix $\#Y = 3$ or 4, via proposition 2, let $v^{(\cdot,\cdot)}$ denote the 2-diverse matrix representation of the improper Y-generalization $\mathcal{R}$. Let $\mathcal{H}_{++}$ denote the associated arrangement of hyperplanes. For every distinct $x, y \in X$, proposition 2 implies that $H_{++}^{\{x,y\}}$ intersects $\mathbb{R}_{++}^{\mathbb{T}}$. Then, similar to example 6, the $1 \leqslant n \leqslant \binom{\#Y}{2}$ distinct hyperplanes of $\mathcal{H}_{++}$ cut $\mathbb{R}_{++}^{\mathbb{T}}$ into at least $n+1$ regions. At least one pair G and G^* in $\mathcal{G}_{++}$ are therefore separated by all n distinct members of $\mathcal{H}_{++}$. Take $J \in G$, so that, for every distinct $x, y \in Y$, $\langle u^{(x,y)}, J\rangle \neq 0$. Thus $\mathcal{R}_J$ is antisymmetric, complete and, via A0$^\flat$, total. Next, take $L \in G^*$, so that since J and L are separated by every hyperplane in $\mathcal{H}_{++}$, $\mathcal{R}_J = \mathcal{R}_L^{-1}$. □

Example 7 (insufficiency of *2*-diversity)**.** *Let $X = [0,1]^2$ and let $\leqslant^{\text{lex}}$ denote the lexicographic ordering on X. Let $\mathbb{T} = \{s, t\}$, and, for each $J \in \mathbb{J}$, let*

$$\preceq_J = \begin{cases} X^2 & \text{if } J(s) = J(t)\,; \\ \leqslant^{\text{lex}} & \text{if } J(s) < J(t)\,; \\ (\leqslant^{\text{lex}})^{-1} & \text{otherwise.} \end{cases}$$

Recall that if $\preceq_J = X^2$, then $\preceq_J$ is symmetric and hence equal to $\simeq_J$. Thus, for every distinct $x, y \in X$, $H_{++}^{\{x,y\}} = \{J \in \mathbb{R}_{++}^{\mathbb{T}} : J(s) = J(t)\}$. Via proposition 2, $\preceq_{\mathbb{J}}$ has a two-diverse matrix representation $v^{(\cdot,\cdot)}$. But via lemma 2.1, below, conditional-*2*-diversity *fails to hold. The fact that $\preceq_{\mathbb{J}}$ fails to satisfy part* (2.a) *of theorem 2 follows from the fact that $\preceq_J$ is lexicographic for every J outside H.*

We now present the canonical intersection semilattices for the variety of cases that we consider in the proofs of the main paper.

In the intersection semilattice of fig. 5, an increase in level corresponds to a decrease in dimension: since $\acute{A}^{\{x,y,z\}}$ is nonempty, it is of dimension at least zero. Since $\acute{J}$ belongs to the interior of $\mathbb{R}_{++}^{\mathbb{T}^{\mathfrak{f}}}$ and $\acute{A}^{\{x,y\}\{y,z\}}$ is at least one-dimensional, $\acute{A}_{++}^{\{x,y\}\{y,z\}}$ is one-dimensional. Since $\acute{A}^{\{x,y,z\}} \subset \acute{A}^{\{x,y\}\{y,z\}}$, the latter set is nonempty whenever $\acute{\mathcal{H}}_{++}$ is central. The same, of course, applies to other members at the same level. Conversely, if $\acute{A}^{\{x,y\}\{y,z\}}$ is empty, then so is $\acute{A}^{\{x,y,z\}}$. We now use this to show that there is a unique form of Y-generalization $\hat{\mathcal{R}}$ such that $\#\hat{\mathcal{G}}_{++} = 6$ and, moreover, that any such $\hat{\mathcal{R}}$ fails to satisfy A0$^\flat$.

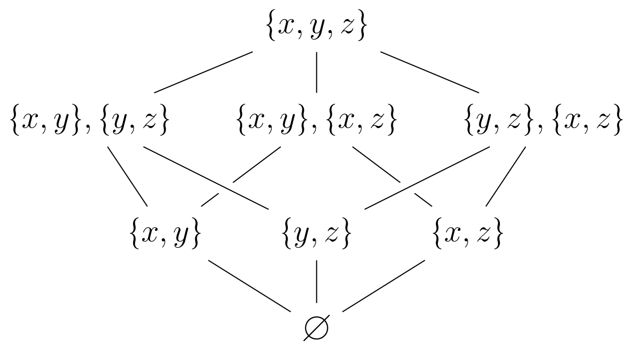


Figure 5: The intersection semilattice of a central arrangement for $\#Y = 3$ and $\mathbf{r} = 3$.