

In Order We Trust

Abstract

Copulas enter psychometrics with a deceptively simple promise: *separate dependence from measurement*. Sklar's theorem guarantees that any multivariate distribution can be decomposed into marginal distributions (what each variable is doing on its own) and a copula (how the variables couple or move together). This is usually presented as a technical convenience. Here we argue it is a *foundational measurement statement*: If a score is only defensible up to *strictly increasing transformations* (ordinal meaning), then the copula is not merely *a* tool for studying longitudinal association—it is, in a precise sense, the *maximal invariant* of dependence available under that measurement regime.

The core result is an invariance theorem: For continuous marginals, two joint distributions are related by coordinatewise strictly increasing transformations if and only if they share the same copula. Consequently, *every* dependence claim that is meaningful under ordinal admissible transformations must be a function of the copula. This places the classical “permissible statistics” debate (Luce & Suppes, 2002; Narens, 2002; Stevens, 1946) into direct contact with copula theory: The measurement level determines the transformation group; the transformation group determines what can be meaningfully inferred; and, under ordinal measurement, dependence reduces canonically to the copula.

A.1 Introduction: What Sklar's Theorem Implies About Measurement

Sklar's theorem is often introduced with an engineer's shrug: *Any* joint distribution can be written as a copula glued to marginal distributions. More formally: If H is a joint cdf with marginals $F_1, \dots, F_d$, there exists a copula C such that

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

Moreover, when the marginals are continuous it can be shown that C is unique (Fan & Patton, 2014; Sklar, 1959).

But the psychometric punchline is not the remarkable decomposition itself. It is the *invariance* built into the construction. Under strictly increasing transformations of each variable,

the marginals change but the copula does not—a fact emphasized in the copula literature as the margin-free character of dependence. This is the decisive bridge: Once dependence is defined only up to strictly increasing re-scalings of each margin, the “raw-score” coordinates become incidental, and the only coherent objects left are those that survive rank-preserving transformations. In that sense, copulas are not merely compatible with rank-based analysis—they are the dependence-theoretic core of it, a point made explicitly by (Schweizer & Wolff, 1981, p. 884): “The study of rank-statistics—insofar as it is the study of properties invariant under such transformations—may be characterized as the study of copulas and copula invariant properties.”

That invariance property is exactly the type of invariance that representational measurement theory treats as the criterion for meaningfulness. The key idea is that a numerical scale is not “a thing in itself,” but a representation of an underlying empirical relational structure. Consequently, the content of any statistical assertion is only as stable as the class of transformations that preserve that structure. In Luce and Suppes’ formulation, meaningful statements are those whose truth value is invariant under transformations that carry one admissible numerical representation of the same empirical structure into another—an invariance perspective inherited from geometry and dimensional analysis and made explicit in measurement theory. Put more concretely: Once we concede that an empirical attribute is secured only up to an ordinal interpretation, the admissible transformations are precisely the strictly increasing maps τ , and any numerical claim that can be altered by replacing ψ with $\tau(\psi)$ is, in the technical sense, not a claim about the attribute at all, but about an arbitrary coordinate choice.

Stevens’ familiar mean/median example is the canonical illustration. Suppose two groups have ordinal scores (1, 4, 5) and (2, 3, 6); their means are $10/3$ and $11/3$, respectively. One is tempted to conclude the first group is smaller on average. But because only order is meaningful, we are free to apply any strictly increasing recoding; under the equally admissible transformation that maps these values to (1, 5, 6) and (2, 2.1, 6.1), the means become $12/3$ and $10.2/3$, respectively. The reversal is not a paradox—it is a proof that mean comparisons rely on metric structure that ordinal measurement does not supply. By contrast, the median (and, more generally, quantiles and ranks) is invariant under strictly increasing transformations, and therefore supports assertions whose truth values survive the admissible re-expression of the scale. In this sense, meaningfulness is not a philosophical afterthought; it is the formal gatekeeper that determines which inferences are well-posed at a given measurement level—precisely the same gate-keeping role played by copula invariance for dependence.

So the profound reading of copulas in psychometrics is:

Sklar’s theorem Tells us that dependence can be represented independently of marginal scale choices.

Copula invariance Tells us that dependence representation is stable under strictly increasing transformations—precisely the admissible transformations for ordinal meaning.

Representational measurement theory Tells us that stability under admissible transformations is what meaning is (or, at least, the minimum requirement for meaningful inference).

In educational assessment, the familiar anxiety—“do we have an interval scale?”—is not a metaphysical hobbyhorse. It is the hinge on which the meaning of change swings. Braun (1988) makes this point clear: Large-scale reporting scales (e.g., NAEP’s Reading Proficiency Scale) can unquestionably support ordinal information, but it is much less clear whether they support interval interpretations, and if so in what sense. Psychometric opinion has long been that IRT-based scales warrant only ordinal claims in general, with the Rasch model sometimes treated as a partial exception only insofar as one is prepared to defend the non-admissibility of monotone transformations and the adequacy of the model for the comparisons at issue. This is the exact place where copula invariance becomes diagnostic: It is not merely telling us something about copulas; it is telling us what kinds of scale-dependent inferences we can responsibly make in the first place.

Seen through representational measurement theory, Braun’s critique is almost a worked example of meaningfulness in the technical sense: Statements are meaningful only to the extent that they are invariant under the admissible transformation group for the scale type. The classical Stevens-style illustration (mean vs. median under ordinal measurement) is not a side anecdote; it is the template. If ordinal admissibility means *strictly increasing transforms are allowed*, then any claim about growth that can be reversed by such a transform is not merely fragile—it is formally non-invariant, hence non-meaningful in the representational sense. Luce and Suppes frame invariance exactly this way: Properties of an empirical structure should be invariant under the transformations taking one numerical representation of the axioms to another. That invariance is not optional decoration; it is the criterion of interpretability.

- **If we do not even have order**, then we are not in a weak measurement world; we are in a different world—essentially nominal—where monotone invariance is not the right invariance group and dependence concepts based on ranking are ungrounded.
- **If we do have order**, then for dependence questions, copulas deliver the maximal invariant. Interval ambitions may still matter for other purposes, but they do not expand what is invariantly available about dependence via copulas.

The rest of this treatise expands and formalizes these ideas.

A.2 Measurement Level as a Transformation Group

Representational measurement begins with an empirical structure (objects and relations) and asks what numerical representations preserve those relations. For psychometrics, the minimal structure is typically an order—at least a *weak order* ($\preceq$) capturing “no less than” on an attribute. Luce and Suppes (2002) explicitly characterize measurement as representation of ordered structures and spell out the weak-order properties (transitivity and connectedness).

Once an empirical order is admitted, an *ordinal representation* is any $\psi : A \rightarrow \mathbb{R}$ such that

$$a > b \iff \psi(a) > \psi(b)$$

for strict order. With ties one uses a weak order formulation.

$$a \succeq b \iff \psi(a) \geq \psi(b),$$

The admissible transformations are the strictly increasing maps τ (or non-decreasing maps if ties are allowed). From the invariance standpoint, the admissible group of transformations $\mathcal{G}$ is not an annoyance; it is the *definition of what the numbers can mean*. A functional Γ is meaningful if and only if it is invariant under $\tau \in \mathcal{G}$, i.e. $\Gamma(\tau \circ \psi) = \Gamma(\psi)$. Meaningful statements must exhibit invariance appropriate to the underlying structure.

- If the inferential target is not invariant under admissible transformations, it is not “approximately meaningful”; it is undefined at the level of the structure (because it depends on representational artifacts).
- If it is invariant, it has a claim to meaning (though not automatically to scientific truth).

A.3 The Order Precipice: From Ordinal to Nominal

If a score does not support an ordinal interpretation—i.e., if it fails to preserve a defensible empirical order—then the entire logic of measurement changes character. The critical issue is which transformations are admissible. Under ordinal meaning, admissible transformations are strictly increasing maps—they preserve the only structure being claimed, namely order. If we drop that monotonicity requirement, two qualitatively different pathologies appear, both of which destroy ordinal meaning but in distinct ways.

First, a transformation may be non-injective—it collapses distinct values. For example, a map that rounds or bins scores (or applies a step function) can send multiple distinct score levels to the same numerical label. This does not merely destroy order—it destroys distinctness. The result is a partition of the original scale into equivalence classes (“all x that map to the same $\tau(x)$ ”), i.e., categories. In the extreme, collapsing can reduce the scale to a small set of labels whose only stable content is membership in those categories.

Second, and more devastatingly, a transformation may be order-reversing on some subset—a re-ordering that scrambles the rank structure. Once such transformations are allowed, order ceases to be invariant—statements like “ a exceeds b ” can be made true or false by recoding alone. At that point, the admissible transformation class effectively expands toward all bijections on the realized score values (or on $\mathbb{R}$ if one allows unrestricted reparameterizations). Under that group, nothing about order can be defended as meaningful because any ordering can be permuted into any other ordering. What remains invariant are only those features that survive arbitrary relabeling: Equality/inequality of labels, counts in categories, and—if one considers joint classifications—contingency-table structure up to row/column permutations.

Relaxing monotonicity does not merely weaken measurement; it pushes one over an order precipice: Once rank structure is no longer protected, the measurement problem collapses back to nominal labeling, and any inference that depends on “higher” versus “lower” becomes representationally ill-posed. *Order is the first nontrivial measurement structure*—the minimal constraint that prevents a fall back into nominal equivalence.

Having established that monotonicity is the minimal constraint that keeps us from sliding back into nominalism, we can now state the corresponding positive fact: Once order is preserved, all admissible representations of that order are the same up to a strictly increasing re-expression. This is the representational measurement-theoretic heart of ordinal scaling. It says that ordinal meaning is not attached to any particular numerical coordinate system,

but to the equivalence class generated by strictly increasing transformations. In particular, if two scoring rules ψ and ϕ purport to represent the same strict empirical order, then they cannot disagree except by a monotone change of coordinates. Conversely, if no such monotone link exists, they are not representing the same order at all. The following lemma makes this precise and provides the formal hinge used repeatedly later: It characterizes exactly what it means, mathematically, for two numerical scales to be merely ordinally equivalent.

Lemma A.3.1 (Ordinal-to-Ordinal Transform–Strict Order Case). *Let $(A, >)$ be a set with a strict total order. If $\psi : A \rightarrow \mathbb{R}$ and $\phi : A \rightarrow \mathbb{R}$ both represent $>$, i.e.*

$$a > b \iff \psi(a) > \psi(b), \quad a > b \iff \phi(a) > \phi(b),$$

then there exists a strictly increasing function $\omega : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\phi = \omega \circ \psi.$$

Proof. Strictness implies both ψ and ϕ are injective. Define ω on $\text{Ran}(\psi)$ by $\omega(\psi(a)) := \phi(a)$. Injectivity of ψ makes ω well-defined. If $x_1 < x_2$ in $\text{Ran}(\psi)$, then $x_1 = \psi(b)$, $x_2 = \psi(a)$ for some $a > b$. Hence $\phi(a) > \phi(b)$, i.e. $\omega(x_2) > \omega(x_1)$. ■

It is worth noting the small but important complication that arises once the empirical relation is a weak order (allowing ties) and representations may collapse distinct empirical states into equal numerical values. In that case, the same construction still goes through, but it yields a non-decreasing (rather than strictly increasing) transformation ω . Strict monotonicity corresponds to injective scoring—no ties in the representation—which is seldom satisfied in practice and is often undesirable when scores are intentionally coarse.

With that nuance in place, the lemma delivers an immediate and policy-relevant implication about the perennial psychometric hope for a “true” interval scale. Suppose there exists an interval scale score ϕ for the same underlying empirical order $>$. Interval status means that alternative admissible representations of ϕ differ only by positive affine transformations $\phi \mapsto a\phi + b$ with $a > 0$, which are strictly increasing and therefore preserve $>$. Lemma A.3.1 implies that any ordinal score ψ that represents this same $>$ must relate to ϕ by a strictly increasing re-expression: There exists ω such that $\phi = \omega \circ \psi$ (in the strict-order case; non-decreasing in the weak-order case). Put plainly: If a globally interpretable interval scale exists and truly preserves the same ordering, it is already contained within the ordinal equivalence class—reachable from any valid ordinal representation by a monotone change of coordinates. The only way this conclusion can fail is if ψ and ϕ are not, in fact, representing the same empirical order. Equivalently, the absence of a monotone link between ψ and ϕ is proof that at least one of the ordinal claims is false with respect to the underlying ordering.

A.4 Copulas: Invariance as the Bridge Between Dependence and Measurement

A copula is the joint distribution of probability-integral-transformed margins (Nelsen, 2006).¹ If $\mathbf{X} = (X_1, \dots, X_d)$ is a random vector with continuous marginal cdfs F_i , then the transformed variables $U_i := F_i(X_i)$ are uniform on $[0, 1]$, and the copula C is the joint cdf of $\mathbf{U} = (U_1, \dots, U_d)$.

¹For a concise overview, see also the [Copula \(statistics\)](#) article on Wikipedia.

Two linked facts matter here.

Sklar decomposition (existence/uniqueness under continuity) As above: $H = C(F_1, \dots, F_d)$, uniquely in the continuous case.

Invariance under strictly increasing marginal transformations If $\tau_i : \mathbb{R} \rightarrow \mathbb{R}$ are strictly increasing, then the copula of $(\tau_1(X_1), \dots, \tau_d(X_d))$ equals the copula of $(X_1, \dots, X_d)$.

The second statement is where the measurement theory bites: *Strictly increasing transformations are exactly the admissible transformations for ordinal scales.* So copulas are not merely “distributional gadgets”; they are *dependence objects whose meaning matches ordinal admissibility*.

This is the formal connection to Stevens’ (1946) median/mean point. For ordinal measurement, “meaningful” univariate location is quantile-based (median, percentiles) rather than mean-based (Luce & Suppes, 2002). For multivariate dependence, the analogous claim is: *Meaningful dependence is copula-based* (rank-based), not Pearson-correlation-based (Schweizer & Wolff, 1981).

Taken together, Sklar’s decomposition and marginal-monotone invariance identify a clean division of labor: The marginals encode the scale-dependent aspects of each measurement, while the copula encodes exactly those features of joint behavior that survive admissible ordinal re-expressions. In the language of representational measurement theory, this means that copula-based dependence is not merely robust to monotone rescaling; it is the dependence content that remains meaningful when only ordinal structure is warranted. At this point the argument is still qualitative—copulas respect the right invariance—but the next step is to make it sharp—to show that copulas do not just provide an ordinal-invariant summary of dependence, but (under continuity) provide the maximal such summary, in the precise sense that any two joint laws indistinguishable by ordinally admissible transformations are characterized by the same copula.

A.5 Copula as the Maximal Ordinal-Invariant of Dependence

To formalize this maximality claim, it is convenient to treat admissible scale changes as a transformation group acting on the space of joint distributions. Ordinal measurement corresponds to coordinatewise strictly increasing reparameterizations. These transformations generate equivalence classes (“orbits”) of joint laws that differ only by admissible rescalings of the margins. The fundamental question is then the invariance question: What aspects of dependence are constant on each orbit and do they completely identify the orbit? The answer—under the usual continuity conditions—is that the copula is an invariant of this action and, moreover, a maximal one: Equality of copulas is equivalent to belonging to the same orbit. Establishing this formally turns the earlier interpretive slogan (“dependence separated from measurement”) into a theorem about the structure of ordinally meaningful dependence.

A.5.1 A Group Action Formulation

Let $\mathcal{F}$ be the class of d -variate probability distributions on $\mathbb{R}^d$ with continuous marginals. Let

$$\mathcal{G} = \{\mathbf{g} = (g_1, \dots, g_d) : g_i : \mathbb{R} \rightarrow \mathbb{R} \text{ strictly increasing bijections}\}.$$

Define the action of $\mathcal{G}$ on $\mathcal{F}$ by: if $\mathbf{X} \sim P$, then $\mathbf{g} \cdot P$ is the distribution of $\mathbf{g}(\mathbf{X}) = (g_1(X_1), \dots, g_d(X_d))$. This is precisely changing the scale of each coordinate by an admissible ordinal transformation.

A functional I on $\mathcal{F}$ is an *invariant* of the $\mathcal{G}$ -action if $I(\mathbf{g} \cdot P) = I(P)$ for all $\mathbf{g} \in \mathcal{G}$. It is a *maximal invariant* if, additionally, $I(P) = I(Q)$ implies Q lies in the same orbit as P , i.e. $Q = \mathbf{g} \cdot P$ for some $\mathbf{g}$. The central example—and the one we prove maximal below—is the copula extraction map $P \mapsto C_P$, which takes values in the space $\mathcal{C}_d$ of d -dimensional copulas.²

A.5.2 The Main Theorem

Theorem A.5.1 (Copula = Maximal Invariant Under Strictly Increasing Marginal Transformations). *For $P \in \mathcal{F}$ with continuous marginals, let C_P denote its copula. Then:*

(Invariance) *For every $\mathbf{g} \in \mathcal{G}$,*

$$C_{\mathbf{g} \cdot P} = C_P.$$

(Maximality) *If $P, Q \in \mathcal{F}$ satisfy $C_P = C_Q$, then there exists $\mathbf{g} \in \mathcal{G}$ such that*

$$Q = \mathbf{g} \cdot P.$$

Equivalently, the orbit space $\mathcal{F}/\mathcal{G}$ is indexed by copulas: Each equivalence class of distributions under admissible ordinal recodings corresponds to exactly one copula.

Proof. (*Invariance*). This is standard in copula theory (Nelsen, 2006): Strictly increasing transformations change the marginals but preserve the joint distribution of the probability-integral-transformed variables, hence preserve the copula.

(*Maximality*). Suppose $C_P = C_Q = C$. Let P_i and Q_i be the i th marginal cdfs of P and Q . Because marginals are continuous, define

$$g_i(x) := Q_i^{-1}(P_i(x)),$$

which is strictly increasing. If $\mathbf{X} \sim P$, then $U_i := P_i(X_i)$ are i.i.d. uniform margins with joint cdf C , and $X_i = P_i^{-1}(U_i)$. Now set $Y_i = g_i(X_i) = Q_i^{-1}(P_i(X_i)) = Q_i^{-1}(U_i)$. Then $\mathbf{Y}$ has marginals Q_i and the same copula C , hence $\mathbf{Y} \sim Q$. So $Q = \mathbf{g} \cdot P$ (Fan & Patton, 2014). ■

A.5.3 The Measurement Corollary: All Meaningful Dependence is Copula-Measurable

Theorem A.5.1 becomes a measurement principle once we adopt the representational criterion of meaningfulness: An assertion about the attribute should not depend on the arbitrary choice of numerical coordinates used to represent it. In Luce–Suppes' and Narens'

²Familiar rank-based dependence measures such as Kendall's τ and Spearman's ρ are themselves functionals of the copula, hence invariants of the $\mathcal{G}$ -action. The copula C_P is *maximal* precisely because it captures all such invariants simultaneously: any $\mathcal{G}$ -invariant statistic—whether scalar (e.g., τ , ρ), vector-valued, or function-valued—must factor through C_P .

formulation, this is operationalized as an invariance requirement: If a class of transformations is admissible for the scale type, then any statistic or property purporting to describe the underlying attribute (rather than the representation) must be invariant under that class (Luce & Suppes, 2002; Narens, 2002). For ordinal measurement, the admissible transformations are the coordinatewise strictly increasing maps, i.e., precisely the group $\mathcal{G}$ acting on $\mathcal{F}$ in Section A.5.1.

Corollary A.5.1 (Ordinal Meaningfulness $\Rightarrow$ Copula Sufficiency for Dependence). *Let Γ be any property of the joint distribution of $\mathbf{X}$ that is invariant under coordinatewise strictly increasing transformations. Then Γ depends on P only through the copula C_P .*

This is not merely a convenient summary result; it is the formal expression of the idea that, under ordinal admissibility, the copula exhausts the non-arbitrary content of dependence. In fact, Theorem A.5.1 is stronger than the usual copula invariance slogan. It states that copulas do not simply remain unchanged under admissible rescalings; rather, they index the equivalence classes of all joint laws under those rescalings. Consequently, any dependence claim that survives all ordinally admissible re-expressions must factor through C_P —equivalently, it must be expressible in terms of ranks/pseudo-observations. This perspective aligns with the dependence-measures literature, where (in the bivariate case) copula invariance and copula-functional characterizations of rank-based dependence measures are made explicit (Schweizer & Wolff, 1981).

At the same time, the corollary also clarifies what copulas do not provide. Because the copula is invariant to arbitrary monotone changes of each margin, it deliberately discards metric information about the marginals—any information that could distinguish two distributions within the same $\mathcal{G}$ -orbit. Thus, copulas deliver a maximal account of dependence that is meaningful at the ordinal level, but they leave open questions whose answers require more than order—questions about equal-interval meaning, additive change, or comparisons of differences across the scale. Those questions are not ruled out; they are additional commitments about the marginal scales, and making them meaningful requires restricting the admissible transformation class from “all monotone” to a smaller group (e.g., affine transformations for interval scales). This is exactly the point to which we now turn. What, if anything, is gained—and what must be assumed—when one asserts an interval/vertical scale beyond the ordinal structure already sufficient to identify ordinally meaningful dependence.

A.6 What Does an Interval/Vertical Scale Buy -- Beyond Copulas?

Here is the clean separation that prevents category error.

- The copula captures dependence structure invariant to monotone recodings. (Theorem A.5.1)
- Interval (and vertical) scaling concerns the marginals. It is the claim that the numerical representation supports more than order—specifically, that differences (and the comparisons of differences) are meaningful, which corresponds to restricting admissible transformations from arbitrary monotone maps to positive affine maps $x \mapsto ax + b$ with $a > 0$.

This restriction genuinely enlarges the set of meaningful statements, but it enlarges it in a very particular way. Under ordinal admissibility, statements about location must be quantile/rank-based and “distance” has no invariant meaning; under interval admissibility, a family of metric assertions becomes well-posed. Examples include claims of the form “the average change is larger in group A than group B,” “a one-unit increase has the same meaning at different points on the scale,” and “linear models describe changes with parameters that do not depend on arbitrary monotone rescalings.” In short, interval scaling buys you invariants of additive structure: differences are invariant up to a change of units, and linear-functional summaries (means, slopes, regression coefficients interpreted as differences) become meaningful because the admissible group has shrunk.

But this is exactly where the psychometric stakes rise. Declaring an interval scale is not merely choosing a convenient reporting metric; it is an empirical/axiomatic claim about the attribute and the measurement procedure: that there exists a representation in which differences correspond to a stable relational structure in the world, and that only affine re-expressions preserve that structure. Put in representational terms, one is asserting that the underlying empirical system has enough structure (often some form of additivity or concatenation/composition) to justify an affine uniqueness theorem. Without that structure, interval becomes a coordinate choice rather than a defended property of measurement.

The same point becomes sharper—indeed, more precarious—when interval is promoted to vertical scaling. A vertical (developmental) scale is not merely an interval scale at a single level; it is a claim that the same unit and same difference meaning hold across substantively different regions of performance (e.g., across grades, content standards, etc.). In invariance language, vertical scaling asks for a single affine admissibility class to hold globally across a heterogeneous domain. That is a much stronger demand than local ordinal comparability, and it is precisely the sort of demand that Braun and others approach with skepticism, especially in policy settings. Even if a scale admits ordinal interpretation robustly, there may be little warrant for the claim that equal numerical differences have equal substantive meaning across the full range.

This framing also clarifies what interval/vertical scaling does not buy. It does not produce more dependence. Under ordinal admissibility, dependence content invariant to monotone recodings is already fully indexed by the copula (Theorem A.5.1). What interval/vertical scaling adds is not additional dependence structure but the license to interpret aspects of the marginals metrically—especially differences and linear trends—because the transformations that would otherwise distort those quantities are now declared inadmissible. Put differently: copulas answer the question, “What dependence survives admissible re-scaling?”; interval/vertical scaling answers the different question, “Which re-scalings are admissible in the first place?”

This is precisely why the Stevens (1946) debate keeps recurring in applied work today. Analysts routinely compute means, gains, and linear growth parameters on reported scores as if affine admissibility were granted by default. When conclusions shift under alternative (yet still order-preserving) score metrics, the field reenacts Stevens’ original “meaningful statistics” controversy in modern form, now expressed as disputes over IRT scaling model, vertical linking methodology, etc. Like with the original debate, the solution only comes when the debate is reframed as an invariance/meaningfulness problem: Once one asks which assertions remain unchanged under the admissible transformations of the scale, many apparent substantive disagreements collapse into differences between empirical claims and artifacts

of representation.

The crucial point for what follows is that moving from ordinal to interval (and especially to vertical) is not a harmless upgrade; it is an added hypothesis about the structure of the attribute and the adequacy of the measurement model. That added hypothesis is exactly where Joel Michell's critique enters. He argues that much of psychometrics has treated the interval/quantitative hypothesis as a methodological presupposition—often smuggled in via model fit or convenience—rather than as a substantive claim requiring explicit axioms and empirical support.

A.7 Conjoint Measurement and Michell's Critique as the Axiomatic Backstop

If one wants interval meaning, representational measurement theory offers an axiomatic path: Additive (conjoint) structures yield numerical representations unique up to affine transformations. This is the *empirical content* behind the interval claim. The price is that the axioms are strong, domain-specific, and testable in principle but often neglected in practice—exactly the kind of neglect that Michell (1997) criticizes.

A.7.1 Michell's Central Complaint: Psychometrics Often Treats “Assignment” as “Measurement”

Michell (1997) argues that much of twentieth-century psychometrics assimilated measurement to the procedural act of assigning numbers by rule, while downplaying (or bypassing) the classical scientific idea that measurement is the successful representation of a quantitative attribute—hence a substantive, falsifiable hypothesis about structure. On this view, to claim more than ordinal information is not merely to adopt a convenient scale, but to assert that the attribute itself supports additional relational structure (e.g., some form of additivity) that warrants a narrower admissible transformation group. Michell argues that Stevens' scale typology, whatever its pedagogical utility, inadvertently licensed a methodological drift. Instead of treating the interval hypothesis as a demanding claim requiring explicit axioms and empirical scrutiny, psychometrics often treated it as a pragmatic convention, rationalized post hoc by model fit or by the practical desirability of arithmetic operations. In Michell's framing, this is not a minor philosophical quibble; it is an epistemic reversal in which the conveniences of representation become mistaken for evidence about the structure of the attribute.

Whether or not one accepts Michell's full diagnosis, the invariance framework isolates a precise methodological core. The question is not, “are numbers being produced?” but “what equivalence class of representations is empirically warranted?”

- Claiming interval meaning is not a stylistic preference; it is a claim that the admissible transformation group has been restricted, and that this restriction has empirical/axiomatic support.
- If that support is absent, then procedures whose substantive conclusions can reverse under admissible monotone transformations—means on ordinal scores, difference-based

gains, linear growth parameters interpreted metrically— are not “approximately informative” about the attribute; they are representation-dependent in exactly the sense Stevens highlighted and Luce–Suppes formalized as a meaningfulness problem.

This is the formal point at which Michell and copulas meet. Copula methods do not solve the problem of quantitateness. They avoid pretending it has been solved. When the defensible admissible group is monotone, the copula isolates what remains invariant—and therefore what can be treated as meaningful—about dependence without smuggling in unearned metric structure.

A.8 Conclusion: In Order We Trust

With these pieces in place, the implications for growth become hard to ignore. Growth claims are, in the first instance, claims about how performance at one occasion relates to performance at another—that is, about the joint law of $(X^{(t)}, X^{(t+1)})$ (possibly conditional on covariates). If we restrict ourselves to what is meaningful under ordinal admissibility, then the dependence component of that joint law must be expressed in an invariance-respecting language. The copula framework supplies exactly that language: It represents the maximal ordinal-invariant content of the joint distribution, while keeping any stronger (interval/vertical) assumptions explicitly confined to the marginal scale. This reframes the usual growth discussion as a dependence problem first, and only secondarily as a question about metric change. Copula-based growth is not a branding exercise, it is the measurement-theoretic consequence of taking ordinal meaning seriously. That is exactly the separation-of-concerns the Sklar factorization is trying to teach: *Dependence lives in C ; measurement choices live in $X^{(t)}$.*

The foundational message can be stated with almost obnoxious economy:

1. **Meaningfulness is Invariance.** A claim is meaningful if and only if it is invariant under the admissible transformations of the measurement scale (Luce & Suppes, 2002).
2. **Ordinal $\Rightarrow$ Monotone Freedom.** If only order is warranted, strictly increasing re-expressions are admissible; absent order, one reverts to nominal labeling.
3. **Copulas are the Ordinal Dependence Object.** Under monotone marginal transformations the copula is invariant and, with continuous marginals, maximal. Thus, every ordinally meaningful dependence claim factors through the copula (Nelsen, 2006; Schweizer & Wolff, 1981).
4. **Interval Adds Metric Marginals, not Dependence.** Interval/vertical scaling licenses affine-invariant marginal comparisons (differences, means, slopes) only with further axiomatic/empirical support (Braun, 1988; Schulz & Nicewander, 1997). It does not enlarge the ordinal-invariant dependence content already captured by the copula.

So the title is not metaphor: *Order* is the trust boundary. If order fails, measurement collapses toward nominal labeling and the invariance regime that makes copulas meaningful is unavailable. If order holds, copulas are not one more modeling framework—they are the canonical dependence object implied by representational measurement theory.

References

- Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, 28(4), 42–51. <https://doi.org/https://doi.org/10.1111/j.1745-3992.2009.00161.x>
- Braun, H. I. (1988). A new approach to avoiding problems of scale in interpreting trends in mental measurement data. *Journal of Educational Measurement*, 25(3), 171–191.
- Fan, Y., & Patton, A. J. (2014). Copulas in econometrics. *Annual Review of Economics*, 6, 179–200. <https://doi.org/10.1146/annurev-economics-080213-041221>
- Genest, C. (2021). A tribute to Abe Sklar. *Dependence Modeling*, 9(1), 200–224. <https://doi.org/doi:10.1515/demo-2021-0110>
- Luce, R. D., & Suppes, P. (2002). Representational measurement theory. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., Vol. 4, pp. 1–41). Wiley.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355–383. <https://doi.org/10.1111/j.2044-8295.1997.tb02641.x>
- Narens, L. (2002). *Theories of meaningfulness*. Lawrence Erlbaum Associates.
- Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). Springer.
- OECD. (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do*. <https://www.oecd.org/pisa/publications/pisa-2018-results/>.
- Schulz, E. M., & Nicewander, W. A. (1997). *Grade equivalent and IRT representations of growth* (ACT Research Report 97-2). ACT. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR97-02.pdf
- Schweizer, B., & Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9(4), 879–885. <https://doi.org/10.1214/aos/1176345528>
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680. <https://doi.org/10.1126/science.103.2684.677>