Title: On Measurement Scales: Neither Ordinal nor Interval?

Abstract:
There is a received view on measurement scales. It includes both a classification of scales (nominal, ordinal, interval, and ratio) and a set of prescriptions regarding measurement inferences. This paper casts doubt on the adequacy of this received view. To do this, I propose an epistemic characterization of the ordinal/interval distinction, i.e., one in terms of researchers' beliefs. This novel characterization reveals the ordinal/interval distinction as too coarse-grained, and thus the received view as too restrictive of a framework for measurement research.

Contact Information: Cristian Larroulet Philippi, Department of History and Philosophy of Science and St. Edmund's College, University of Cambridge, Cambridge, UK; e-mail:cristianlarroulet@gmail.com. ORCID: 0000-0001-5793-4670

Acknowledgments: Thanks to Anna Alexandrova, Julia Staffel, and Jacob Stegenga for helpful comments and suggestions.

Funding: This paper is based on research that was funded by the Gates Cambridge Trust and ANID (Chile).

On Measurement Scales: Neither Ordinal nor Interval?

1. Introduction.

There is a received view on measurement scales. It includes both a classification of scales and a set of prescriptions regarding which measurement inferences are justified. According to this view, the measurement scales used by researchers may be classified as nominal, ordinal, interval, or ratio, depending on the information they provide. Nominal scales only represent equality among elements of the same category (as in the classification: 1=alive, 2=dead). Ordinal scales represent rank-order among elements (e.g., in an attitudinal question, options 5=strongly agree, 4=agree, 3=neutral, 2=disagree, 1=strongly disagree). Quantitative measurement, however, begins only with interval scales. Here the intervals (that is, the differences between subsequent levels of the scale) are equal in magnitude. For instance, the difference in temperature between 2ºC and 3ºC is the same as the difference between 4ºC and 5ºC. This equality marks the difference between interval and ordinal scales—unlike the case of temperature in the Celsius scale, we do not know if the distance between 'strongly agree' and 'agree' is the same as that between 'agree' and 'neutral'. Finally, ratio scales, such as length measured in centimeters, are distinguishable from interval scales in that they have a non-arbitrary zero.

This classification of scales is tied to a set of methodological prescriptions concerning the kinds of mathematical operations that may be applied to measurement results. These prescriptions are meant to ensure the correctness of inferences from measurements. For example, one prescription says researchers should not take averages of their results if they are measuring temperature with an ordinal scale (a "thermoscope"). This prescription entails researchers should not compare the average temperature of a group of places with that of another group in order to infer which one is on average hotter. Similarly, if measuring temperature with an interval scale, researchers should not compute ratios in order to infer that, say, place *a* is twice as hot as place *b*. These scale types and the associated prescriptions were first articulated by Stanley Stevens in his famous "permissible statistics" (1946). Later, the Representational Theory of Measurement (RTM) (e.g., Suppes and Zinnes 1963) provided formal foundations for both the standard classification of scales and the (un)justified mathematical operations. But the endorsement of the classification and prescriptions in research methodology goes well beyond the adherence to RTM, or for that matter to any specific theory of measurement. It is just the received view on measurement scales, usually presented as standard methodology in textbooks across the sciences.

A cursory look at many areas of the social and biomedical sciences, however, reveals that the prohibition on taking averages from ordinal scales has proved especially difficult to adhere to. Averages from ordinal scales are routinely used in psychology, sociology, economics, and medicine, despite methodologists frequently denouncing the practice as "impermissible." While we have seen a revival of the philosophy of measurement in the last years (Tal 2017), little has been said regarding the mismatch between practice and methodology that surrounds measurement scales. I address this lacuna here, casting doubt on the adequacy of the received view. Focusing on the ordinal/interval distinction, I argue that the received view is too blunt a tool to be a satisfactory guide to measurement.

After describing the scale classification and associated prescriptions of the received view (2), I raise the worry that the prescriptions may be overly restrictive if the classification is not exhaustive enough (3). In order to assess the relevance of this worry, I offer an epistemic (Bayesian) characterization of the ordinal/interval distinction, i.e., in terms of researchers' beliefs about intervals (4). This novel epistemic characterization reveals that the ordinal/interval distinction is too coarse-grained to appropriately represent all real-world measurement scales.

Indeed, (forced) application of the received view might lead to overly cautious methodological prescriptions. We need a subtler epistemic framework of measurement scales (5).

2. The Received View on Measurement Scales.

The received view defines scales by the uniqueness of their numerical assignments, which is in turn defined by set of transformations that preserve the information the scales give. These transformations are called "permissible" or "admissible" (Suppes and Zinnes 1963). Ordinal scales are defined as those that are unique up to order, which means that any order-preserving ("increasing monotonic") transformation is admissible. This expresses formally the intuitive idea—famously articulated in (Stevens 1946)—that these scales only provide information about the relative order of elements, but nothing more. Thus, any order-preserving transformation gives us the *same* information we already had.

Beyond providing information about order, the specific characteristic of interval scales is that their intervals are of equal magnitude. Here, any positive linear transformation (i.e., a transformation from x to y that satisfies: $y = a + bx$, $b>0$) is admissible. Any such transformation may change the magnitude that the scale assigns to 0 (if $a \neq 0$) and the absolute value of the intervals, but not the equality of the intervals. As well as having equal intervals, ratio scales have a natural 0. Thus, only positive similarity transformations are admissible (i.e., $y = bx$, $b>0$); otherwise, ratios would not be preserved.

The methodological prescriptions are based on these admissible transformations. The general form of the prescriptions is the following: when inferring claims from measurement results, only the claims that remain true under all admissible transformations are validly inferred. The justification for this general prescription lies in the fact that the information each scale gives is determined by what is common across their admissible transformations—all admissible transformations of a scale represent the phenomenon equally well. For example, somebody might (incorrectly) infer that place *a* is twice as hot as place *b* because the former is at 20ºC while the latter at 10ºC. However, in a Fahrenheit scale, *a*'s temperature is 68ºF, and *b* is 50ºF (*not* 68ºF/2=34ºF). Inferences such as 'here is twice as hot as there' are not validly made with these (interval) scales since the measurement comparison is sensitive to the admissible transformation used. This is why standard methodology rules them out.[1]

Let us see how this general prescription applies to ordinal scales. The paradigm example of an ordinal scale in the physical sciences is Mohs' scale of hardness for minerals. This scale uses the following rule to order minerals: if mineral *a* is able to scratch mineral *b*, then *a* is harder than *b*. It also assigns numbers from 1 to 10 in increasing levels of hardness, and each level is associated with a specific mineral. Because in ordinal scales the differences in magnitude between levels are not invariant across admissible transformations, mathematical operations like addition give results that are not invariant to admissible transformations. So, we cannot infer that groups of objects A and B have the same average hardness from the fact that their hardness levels in Mohs' scale are A={2,3,4} and B={3,3,3}. For if we apply the transformation $y = 2^x$, the averages now differ.

Note that the prescription allows inferences when they are invariant. Under which condition are inferences with averages from ordinal scales invariant? A mathematical concept helps stating this

[1] RTM conceives of this as an issue of "meaningfulness": if a claim is not invariant to admissible transformations, it is not (empirically) "meaningful" (Suppes and Zinnes 1963; Roberts 1985). That the issue at stake is better understood as one of valid inferences (versus of meaningfulness) is persuasively argued in (Michell 1990).

condition. Consider G^A and G^B to be the cumulative distributions of each group: $G^A(x)$ is the fraction of minerals in group A that are as hard or less hard than x (and similar for G^B). A well-known result in statistics and economics says: A's computed average is higher than B's computed average under any order-preserving transformation iff $G^A(x) \leq G^B(x)$ for all x and with a strict inequality over some values of x. The biconditional's right-hand side is called first order stochastic dominance (FOSD).[2] FOSD assures that no matter which order-preserving transformation is used, A's computed average would always be bigger than that of B. Of course, FOSD is a very strong condition. But nothing weaker can assure that the average comparison remains invariant under *all* order-preserving transformations.

The received view, then, offers a classification of scales in terms of admissible transformations, and a set of prescriptions about measurement inferences based on whether the measurement results are invariant across admissible transformations.

3. A Potential Problem for the Received View.

Why does the received view single out only the admissible transformations (and thus, the kinds of scales) that it does? If scales are defined by their admissible transformations, what stops us from having many other kinds of scales? Of course, many sets of admissible transformations can be considered between that of all order-preserving transformations and that of all linear positive transformations (Suppes and Zinnes 1963, 14) (see an example below).

The sets of admissible transformations considered are nested (Figure 1a). Order-preserving transformations include all positive linear transformations, which in turn include all positive similarity transformations. Importantly, there is a relation between the admissible

[2] FOSD is defined (and this result proved) in Hadar and Russell (1969).

transformations and what is necessary for conclusions to be invariant: the larger the set of admissible transformations, the stronger the condition for conclusions to be invariant (and thus validly inferred). For this reason, the conditions that ensure that conclusions are invariant are stronger for ordinal scales than for interval scales (e.g., FOSD is not needed to make average comparisons when working with interval scales). Given that there is a positive relationship between the size of the set of admissible transformations and the strength of the conditions necessary for results to be invariant, having scales that are defined by smaller sets of admissible transformations allows us to make valid inferences with weaker conditions. The possibility of having such scales makes the issue of what scales are (and are not) part of the standard classification more pressing.

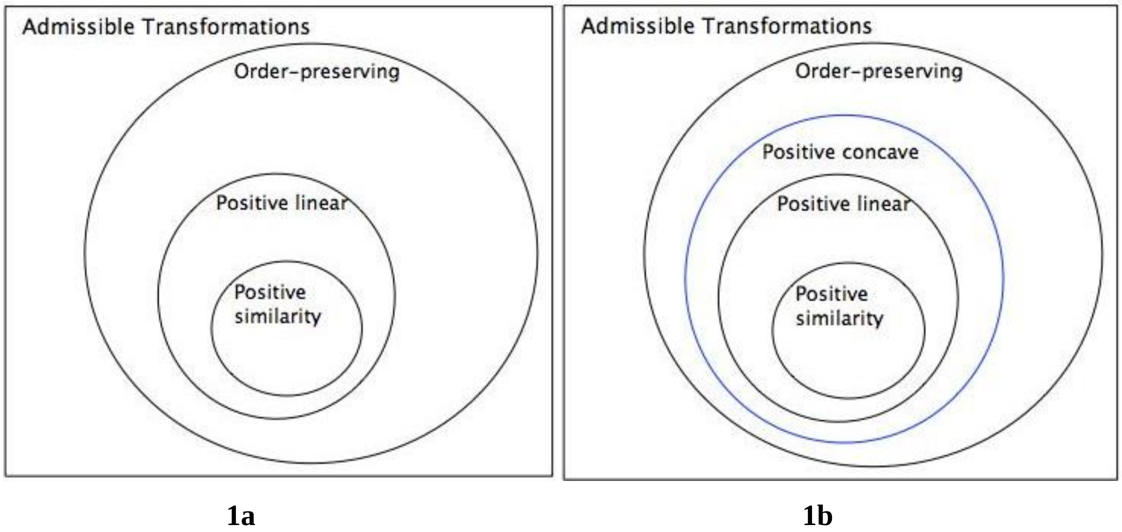

Figure 1. Sets of admissible transformations.

To illustrate, imagine there exists another kind of scale, the 'ordinal*', defined by the following set of admissible transformations: any positive concave transformation (i.e., $y = f(x)$, $f' > 0$, $f'' \leq 0$). This set of transformations is a subset of the set of order-preserving transformations, and it contains the set of positive linear transformations (Figure 1b). Thus, the conditions necessary for results to be invariant are weaker with ordinal* scales than with ordinal scales.

If a researcher wants to compare the average hardness of two groups of minerals, it makes a difference whether she has an ordinal versus an ordinal* scale. In the case of the former, she needs FOSD to hold. In the case of the latter, only a substantially weaker condition is necessary, called "second order stochastic dominance" (SOSD, see proof in Hadar and Russell 1969). Group comparisons that do not satisfy FOSD may satisfy SOSD. Hence, if a researcher is working with a scale that is in fact ordinal*, but which is deemed to be ordinal just because ordinal* is not part of the conceived possibilities, that researcher might be wrongly forbidden to make some inferences. If this were the case, the scales framework endorsed by the received view would be a defective guide for research—its (forced) application would classify valid inferences as 'invalid.' Policing too strict a (methodological) morality is, surely, an unwelcome result.

Whether the received view, which excludes any kind of scale between ordinal and interval, is a good guide for research depends among other things on whether the scales applied to actual measurement situations fall (for the most part, at least) neatly into those categories. If they do, researchers would not be working with a scale excluded in the received view, and thus nobody would be wrongly abiding to too strict a prescription. We should consider, then, whether the kind of scales singled out by the received view *just are* the ones researchers are likely to find themselves with in actual practice. If this is true, it would save the received view from the charge of being a defective guide. I will argue that this is unlikely.

The RTM-inspired way to assess whether this is the case would be to prove representational and uniqueness theorems about scales that lie between the ordinal and the interval, and verify whether the axioms required for those representational theorems are satisfied by the observed empirical systems (i.e., the phenomena) that the scales aim to measure. It has been persuasively argued, however, that it is not strictly speaking possible to verify whether the axioms are satisfied (because of cases not yet observed, some of which are not observable in practice) (Michell 1990, 31; Sherry 2011, 517ff). In this paper I offer a different route. Taking inspiration from the "epistemic turn" in the philosophy of measurement, I propose that we characterize the ordinal/interval distinction explicitly in terms of researchers' beliefs. This provides a more flexible way of thinking about scales; one that is less focused on the complete numerical representability of attributes abstractly considered (as in RTM), and more in the inferences researchers can validly make with measurement results.[3]

From an epistemic perspective, the ordinal/interval distinction reduces to beliefs about differences between intervals. An interval scale is a scale where the intervals are known to be of equal magnitude (and thus, inferences from averages are always valid). In the case of ordinal scales, things are not as straightforward. Ordinal is a scale that *only* informs about *order*. The 'order' part is easy to understand: all the intervals are known to be positive (so that, e.g., a '5' is strictly harder than a '4'). But what does the 'only' part entail for a researcher's belief about intervals' differences? It is not obvious. Clearly, it is not correct to say: 'If the intervals are known to be of different magnitude, a scale is known to be ordinal.' For example, if a researcher knows that all intervals are different, but also knows that no interval is three or more times larger than any other, then it is *not* the case that all order-preserving transformations are epistemically on a par. (Any order-preserving transformation that makes some intervals three or more times

[3] Narens' (1981) results cast doubt on whether there can be representational theorems for scales between ordinal and interval. My approach here offers a way of conceiving such scales that avoids the need for these representational theorems.

larger than other intervals should be ruled out.) Thus, the mere knowledge that intervals are not equal—which implies that the scale is not interval—is insufficient for saying that the scale is ordinal. It seems plausible to say that, the less equal the intervals of a scale are, the farther the scale is from being interval. But when are the intervals different enough for the scale to be ordinal? More generally, what beliefs about differences between intervals are constitutive of an ordinal scale? The tools of Bayesian epistemology can help model the ordinal/interval distinction.

4. A Bayesian Take on the Ordinal/Interval Distinction.

Under the received view, data from an interval scale reliably informs us about average differences, while data from an ordinal scale does not. In this line, one approach to model the ordinal/interval distinction is to take it as a case of "unreliable evidence" (see Howson and Franklin 1994). The idea here is that the computed average difference may be, depending on the kind of scale, more or less indicative of how the two groups actually compare to each other. Imagine a researcher interested in (dis)confirming hypothesis *H* ('group A is harder on average than group B'). Observing some positive evidence (*E*: A's average > B's average) may confirm hypothesis *H* more or less depending on how reliable the scale is (*K*) for inferring hypotheses like *H*.[4] We know that if the intervals are all equal (*K*=1), the scale is fully reliable: *E* entails *H* and is entailed by *H* (the likelihood is Pr(*E*|*H*&*K*)=1). The less equal the intervals are, the less indicative *E* is of *H*. This is because the numbers that the scale assigns to the different minerals (and, which determine whether *E* is the case) are less indicative of the actual relative degrees of hardness. One way of putting this is using the noise versus information analogy: the more

[4] Just like in Howson and Franklin (1994), *K* is assumed to be probabilistically independent of *H*.

heterogeneous the intervals are, the larger the proportion of noise to information-about-degrees-of-hardness there is in the numbers that the scale uses. Arguably, there is a point in which intervals are believed to be wildly heterogeneous enough ($K \approx 0$) so that H and E are taken to be (for all purposes) probabilistically independent. In this case, there is no confirmation ($\Pr(H|E\&K) = \Pr(H|K)$).[5]

If we model the interval scale by a researcher that assumes (or assigns credence 1 to) K=1, how is an ordinal scale modelled? One option: the researcher assumes (or assigns credence 1 to) $K \approx 0$. Although at first sight plausible, there is something counterintuitive about this representation of an ordinal scale. It implies that the researcher takes for granted something quite specific about the intervals' differences (namely, that they are wildly heterogeneous). This plainly contradicts the idea that an ordinal scale gives *only* information about order, so that *nothing* is known about intervals' differences. Credence 1 in any other value of K faces the same problem.

Another possibility: the researcher assigns a uniform distribution to K: $K \sim U(0,1)$. Motivated by the principle of insufficient reason, the idea could be that the researcher has no reason to take as more likely any specific degree of heterogeneity between intervals than other degrees. Treating them on a par, the idea would go, requires believing $K \sim U(0,1)$. But, as it is well-known, the uniform distribution does not amount to an informationless assumption. For example, how is the

[5] This discussion simplifies in some regards the relationship between intervals' heterogeneity and the likelihood. It is true that, as suggested by the noise-to-information analogy, the more heterogeneous the intervals, the less one should trust average comparisons *in general*. However, heterogeneity can be increased in different ways, and not all of those ways affect all average comparisons equally. Once we fix the number of categories of the scale and the specific data observed, the specific intervals' heterogeneity that matters can be stated precisely (see Larroulet Philippi n.d.).

fact that the researcher assigns equal probability to, say, *K* being between 0.5-0.6 and between 0.6-0.7, compatible with her knowing *nothing* about intervals' differences?

So we have already a significant result. In this Bayesian framework, it is not clear how to represent an ordinal scale. No credence about *K* matches the informal description of an ordinal scale (namely, that which gives only information about order, so that nothing is known about intervals' differences). For Bayesians, at least, this result should raise some concerns about the suitability of the notion of an ordinal scale.

There is another possibility for (somehow still) representing ordinal scales in a standard Bayesian model. It involves giving up the goal of faithfully representing the researcher's beliefs (or ignorance, rather) about intervals' differences. We can black-box the beliefs (for a moment), and focus on representing faithfully the assumed *corollary* of having an ordinal scale. Ordinal scales, according to the received view, are such that positive evidence cannot be used to (dis)confirm *H.* So, taking that as a fixed point and reverse-engineering, we can now ask what should a researcher's beliefs be like for this prescription to be brought about by our representation? The only belief compatible with such prescription is to assign credence 1 to $K \approx 0$ so that there is no confirmation.

As argued above, these beliefs about the intervals do not match the common understanding of an ordinal scale. But unless we impose them on the part of the researcher, we just do not get the prescription that is supposed to hold for ordinal scales (i.e., that we cannot confirm *H*). Indeed, the apparently more reasonable (but still unsatisfactory) alternative of assigning a uniform distribution to *K* would have meant that positive evidence does provide some confirmation to *H*. Thus, it is only *certainty* about an extreme heterogeneity of intervals—to the point of having a *totally unreliable* measuring instrument—that is compatible with the prescription. In that sense, only this *certainty* about intervals' heterogeneity is a plausible representation of what it is for a researcher to have an ordinal scale.

Summing up, when modelling the received view on measurement scales from a standard Bayesian perspective, the most plausible interpretation of the ordinal versus interval distinction maps to the following distinction: researchers have certainty about the extreme heterogeneity of intervals versus researchers have certainty about the equality of intervals. How does this result bear on our assessment of the received view? Quite negatively—the ordinal/interval distinction is not (epistemically) fine-grained at all. The ordinal/interval distinction does not amount to two reasonably spaced categories, so that both might jointly capture the situation of most researchers working with scales in the ordinal/interval area. Rather, the distinction picks out two *extremes* of a continuum of possibilities regarding beliefs about intervals' differences. That actual researchers will never (or almost never) find themselves between being *certain of intervals' equality* versus being *certain of intervals' extreme heterogeneity*, is, on the face of it, extremely unlikely.

Bear in mind that any knowledge about plausible differences between intervals (that does not entail equality of intervals) is ruled out by the position being challenged. Think, for example, of any bounds that physical laws might suggest for plausible relative hardness of minerals, and thus for physically possible or likely differences between intervals of hardness scales. That kind of knowledge may rule out some levels of (substantial) heterogeneity without, of course, necessarily establishing equality of intervals. Thus, such knowledge needs to be assumed as non-existent if we are to say, as the received view assumes, that researchers have either ordinal or interval scales but nothing in between. That the absence of any such knowledge is required from the world (of researchers) for the received view to be an adequate framework puts pressure on it.

Moreover, focusing on the ordinal scale side of the continuum, it is doubtful that actual researchers will find themselves in such a doxastic state (let alone that scales will actually have extremely heterogeneous intervals). In order to get the prescription about averages within our epistemic representation, radically strong views about intervals' heterogeneity need to be held by researchers. What kind of evidence could they have for rationally settling on such strong beliefs is unclear to me. That actual researchers would rationally hold such beliefs in any given actual case is, then, unlikely. [6]

Granted, toy examples of ordinal scales can be produced by stipulation. But whether this resembles the situation of real-life researchers, working with scales developed out of background

[6] An alternative formal representation, which I cannot discuss here due to space constraints, is to use imprecise (versus sharp) credences. If the researcher knows nothing about intervals' differences, she can neither rule out any particular value of K nor consider all values as having equal density. This situation may be modelled by a set of probability distributions (versus a single distribution), one for each value of $K \in [0,1]$. Under this representation, the ordinal/interval distinction maps to the following distinction: the case where researchers *cannot rule out any* possible degree of interval heterogeneity versus the case where they can rule out *all* possible degrees of heterogeneity (except for no heterogeneity). Under this representation, then, we also have that the ordinal/interval distinction picks out two extremes of a continuum; in this case a continuum of possibilities regarding degrees of intervals' heterogeneity that researchers may rationally rule out. For the same reasons given above, it is unlikely that actual researchers will never find themselves in between these two situations. As before, any knowledge about plausible differences between intervals needs to be assumed as non-existent (otherwise *some* possible degrees would be ruled out, falling in between the two extremes). Thus, the ordinal extreme of this continuum is also unlikely to be instantiated, because of the *radical ignorance* it entails.

knowledge and experimental work, is a different matter. Indeed, given the kind of beliefs researchers must have about intervals' heterogeneity for a scale to be ordinal, finding a real ordinal scale might prove challenging. At least, a strong case can be made that the alleged "paradigm example" of an ordinal scale in the physical sciences—Mohs' scale—is neither ordinal nor interval. Friedrich Mohs himself believed he had a sense of how different the intervals were. With the exception of the last interval (9-10), he believed that the intervals of his scale were not that different so as to render the scale not fit for quantitative analysis. He was later on, to considerable extent, proved right on both counts (Tabor 1954; see discussion in Larroulet Philippi n.d.).

5. Conclusions.

I have cast doubt on the adequacy of the received view as a framework for guiding measurement by arguing that it is unlikely that the scales singled out by the received view *just are* the ones which researchers find themselves with in actual practice. When considered from the perspective of researchers' beliefs, the ordinal/interval distinction marks two extremes of a continuum. That all (or most) actual scientific scales lie in either extreme of the continuum is not self-evident. Indeed, for a scale to be ordinal, quite strong beliefs have to be in place.[7] Hence, it is unlikely that real-life researchers always have either ordinal or interval scales but never something in between.

Let me clarify that this does not necessarily amount to a critique of RTM. The correct interpretation of RTM (e.g., either as a complete theory of measurement or as a more modest project) is an open issue. And if RTM is interpreted merely as a (non-exhaustive) library of

[7] Or, complete ignorance, under the imprecise credences interpretation.

theorems (Heilmann 2015), the above cannot be a critique of RTM per se. Rather, it is a critique of what I have called the received view on measurement scales, which includes the claim that the actual measurement scales used by researchers may be smoothly classified in ordinal, interval, and ratio.

The thesis here defended raises an important methodological issue. Widespread acceptance of the received view has arguably led to the implicit endorsement of the following assumption: if a scale ranks correctly but it is not interval, then it is ordinal. This is altogether reasonable when there are no other options between ordinal and interval. But of course, we have seen that there may be other options available. Indeed, from an inferential perspective, it makes little sense to take them as the only two options. Arguably, Mohs' scale and (at least) several scales deemed 'ordinal' in biomedicine and the social sciences research contexts are merely known to be not-interval. Being wrongly classified as ordinal is no small problem. Researchers using these scales might wrongly be forbidden to make inferences (e.g., from computed averages). Since not-interval is compatible with being close to being interval (or close enough, depending on how strongly positive the evidence is), this prohibition might not be justified across the board. This methodological overstepping is, surely, an unwelcome result of the coarse-grainedness of the received view. And it may well explain part of the tension between practitioners and measurement methodologists mentioned at the beginning.

Looking forward, we need a subtler epistemic framework for measurement scales. This paper is only a first step in that direction. More fine-grained possibilities, and classifications better-aligned with researchers' epistemic predicaments, may ground more reasonable prescriptions. This would not only be better epistemology. It might also avoid some of the recurrent tensions between methodologists and practitioners on the status of their average comparisons.

References

Hadar, Josef, and William Russell. 1969. "Rules for Ordering Uncertain Prospects." *The American Economic Review* 59:25-34.

Heilmann, Conrad. 2015. "A New Interpretation of the Representational Theory of Measurement." *Philosophy of Science* 82:787-97.

Howson, Colin, and Allan Franklin. 1994. "Bayesian Conditionalization and Probability Kinematics." *The British Journal for the Philosophy of Science* 45:451-66.

Larroulet Philippi, Cristian. n.d. "Against Prohibition (Or, When Using Ordinal Scales to Compare Groups is OK)," unpublished manuscript.

Michell, Joel. 1990. *An Introduction to the Logic of Psychological Measurement*. Hillsdale, NJ: Erlbaum.

Narens, Louis. 1981. "On the scales of measurement." *Journal of Mathematical Psychology* 24:249-275.

Roberts, Fred. 1985. *Measurement Theory.* Cambridge: Cambridge University Press.

Sherry, David. 2011. "Thermoscopes, thermometers, and the foundations of measurement." *Studies in History and Philosophy of Science* 42:509–524.

Stevens, Stanley. 1946. "On the theory of scales of measurement." *Science* 103:667-680.

Suppes, Patrick and Joseph Zinnes. 1963. "Basic measurement theory." In *Handbook of mathematical psychology* (Vol. 1), ed. R. D. Luce, R. R. Bush, and E. H. Galanter, 1-76. New York: Wiley.

Tabor, David. 1954. "Mohs's Hardness Scale—A Physical Interpretation." *Proceedings of the Physical Society. Section B* 67:249-57.

Tal, Eran (2017). "Measurement in Science." *The Stanford Encyclopedia of Philosophy*. URL=<https://plato.stanford.edu/archives/fall2017/entries/measurement-science/>.