Synthese (2025) 206:32
https://doi.org/10.1007/s11229-025-05109-0

ORIGINAL RESEARCH

Measurement theory in the context of scientific enquiry

Davide Rizza[1]

Received: 13 November 2024 / Accepted: 3 June 2025

Abstract
The body of work on mathematical models of measurement known as 'measurement theory' is not yet adequately understood from a philosophical standpoint. Recent evaluations of measurement theory (in this paper I focus on Heilmann (Philosophy of Science 82:787–797, 2015); Philippi (Philosophy of Science 88:929-939, 2021); Tal (Perspectives on Science 29:701–741, 2021)) have dislocated it from the context of scientific enquiry, making it harder to appreciate its motivations and connection with experimental and theoretical work. This paper seeks to clarify what measurement theory is and does, offering a mathematically more nuanced account of its results than currently available in the philosophical literature.

1 Three dislocations of measurement theory

The second half of the xx century saw the rise of sustained interest, on the part of philosophers and mathematical modellers alike, in the mathematical dimension of measurement. For mathematical modellers, a fundamental motivation came from the need to introduce viable scaling methods into the social sciences, especially psychology.

This motivation led to a simple formal programme chronologically situated between Suppes (1951) and (Krantz et al., 1971), with Suppes and Zinnes (1963) its first comprehensive survey.

The formal programme in question may be outlined with the help of a simple example. Suppose an agent is asked to rank a finite set of items $A = \{a_1, \ldots, a_n\}$ according to preference, breaking ties. This ranking describes a finite linear order, designated by the symbol $\prec$. Each item in A can be assigned a positive rational num-

✉ Davide Rizza
d.rizza@uea.ac.uk

1 School of Politics, Philosophy, and Area Studies, Faculty of Humanities, University of East Anglia, Norwich NR4 7TJ, UK

Published online: 01 July 2025

ber in such a manner that the ordering of numbers according to magnitude reflects the agent's preference ranking.

To see how, pick a $a_1 \in A$ and assign it the positive rational r. Then pick a_2 and check whether the experimental subject prefers it to a_1. If a_2 is preferred, it is assigned $r + 1/2$, otherwise it is assigned $r/2$. Next, a_3 is taken into account. If a_3 is preferred to one but not the other of a_1, a_2, then it is assigned the arithmetical mean of the numbers assigned to a_1, a_2, and so on.

By viewing the assignment to each a_j of a unique number r_j as a function f with domain A and codomain $\mathbb{Q}$, we see that f satisfies the following equivalence:

$a_i \prec a_j$ iff $f(a_1) < f(a_j)$.

The last condition tells us that f is a *strong homomorphism* (a homomorphism only satisfies the left-to-right direction of the equivalence).[1] It is easy to see that, if G is an arbitrary, strictly increasing permutation of the rational numbers, then the composition $G \circ f$ is also a strong homomorphism.

In 1963, Suppes and Zinnes systematically gave formal conditions for a wide range of psychological traits that, if satisfied, would imply the existence of a strong homomorphism from each trait, described as a set-theoretic structure, to a structured numerical set.[2] Such strong homomorphisms are also known as representations or, more suggestively, measurement scales. The next step in the formal programme surveyed by Suppes and Zinnes consisted in describing the totality of available measurement scales in terms of 'scale transformations', i.e. structure-preserving bijections of the numerical codomain common to all homomorphic embeddings.

The programme articulated by Suppes and Zinnes may be called representational, on account of its exclusive interest in families of homomorphic embeddings from a class of structures defined by a set-theoretic predicate into a fixed numerical structure. (Adams, 1966) was the first philosophical attempt at critically evaluating such a formal approach to measurement, which that paper explicitly qualified as representational.

Since 1971, however, measurement-theoretic enquiries have integrated early work into a significantly more sophisticated and flexible body of mathematical results, which is largely, though not entirely, discussed in the three-volume work *Foundations of Measurement* (Krantz et al., 1971; Luce et al 1990).

The shift occurred between the early and later phase of measurement theory has not been accompanied by a satisfactory philosophical understanding of what the later measurement-theoretic results accomplish and of their function within scientific enquiry.

Recent interest in evaluating measurement theory and its significance has led to various, more or less critical, attempts at portraying this research programme. In

[1] In more general terms, a strong homomorphism strongly preserves relations, as well as functions and constants. If, for instance, $h : X \to Y$ is a strong homomorphism and X, Y are endowed with the respective n-ary relations R^X, R^Y, then for any $x_1, \ldots, x_n$, $R(x_1, \ldots, x_n)$ iff $R^Y(h(x_1), \ldots, h(x_n))$. Functions are preserved in an entirely analogous manner (for any k-ary function f^X on X, think of equations of the form $f^X(x_1, \ldots x_k) = x_{k+1}$ as relations on X) and the constants of X are sent by h into constants of Y.

[2] In the above toy example, the requisite formal conditions are the transitivity, trichotomy and irreflexivity of $\prec$.

this paper I consider three accounts and critical discussions of measurement theory, namely (Heilmann, 2015), (Philippi, 2021) and (Tal, 2021). These discussions illustrate three philosophical ways of dislocating measurement theory from its place within scientific enquiry. I find it convenient to qualify them a descriptive dislocation (Heilmann, 2015), a normative dislocation (Philippi, 2021) and a methodological dislocation (Tal, 2021).

The first dislocation consists in an account of measurement theory that severs it from its originating scientific motivations and regards it primarily as a cluster of purely formal results. The second dislocation consists in viewing measurement theory as a source of extrinsic norms imposed on the working scientist. The third dislocation consists in viewing measurement theory as the mathematical formulation of an implausible scaling methodology.

Because the papers referred to contain discussions of important aspects of measurement theory, but do not show a sufficiently nuanced appreciation of its leading problems, results, or their key consequences, it seems important to correct the critical evaluations they contain.

This paper is entirely devoted to providing a corrective to the dislocations described. I will do it in three stages, marked by pairs of consecutive sections. In Sects. 2 and 3, I clarify the central purpose of measurement theory and the range of strategies used in pursuing it, for the sake of correcting the descriptive dislocation effected in (Heilmann, 2015). In Sects. 4 and 5, I illustrate a classification programme from measurement theory in order to correct the normative dislocation effected in Heilmann (2021), and concerning the normative impact of a formal classification programme on scientific practice. In Sects. 6 and 7, I clarify what contact measurement theory makes with experimental work, especially in social science, and with theoretical work, especially in physical science, in order to correct the methodological dislocation effected in (Tal, 2021). In particular, I show that the methodological tenets presumed by Tal to be an inextricable part of measurement theory can be fully disassociated from it.

As the brief synopsis sketched suggests, this paper is not only a critical study of important contributions from the philosophical literature. It is also an attempt to offer a concise account of many sophisticated dimensions of measurement-theoretic work, which, to the best of my knowledge, are not yet appreciated in philosophical work.

2 Measurement theory: codification

The descriptive dislocation of measurement theory found in (Heilmann, 2015) can properly be discussed only against a sufficiently detailed, independently available, account of the central purpose of measurement theory. This purpose will be discussed in the present section. It is, in short, the symbolic codification of empirical content. As aptly noted by E. Adams:

> At least one very important reason for introducing measurement procedures is to provide systematic objective indices of phenomena. Numbers may enter into

the reports of applying measurement procedures but this is not essential to their function as indices (Adams (1966: 142).

The above quote makes two highly significant points. The first point is that the means of codification is not bound to be a homomorphism. The second point is that codes do not have to be numbers. What is of central importance is only that it should be possible to decode the empirical content that has been codified. Only in this sense can the codification provide 'systematic objective indices' of phenomena.

Before I proceed any further, I wish to clarify the use of the word 'codification' in this section. This term is chosen in a sense closely connected with that current in the branch of information theory known as 'coding theory'. What coding theorists study is the operation of sending and receiving messages transmitted through a noisy channel. In our context noise is not important, but the acts of coding and decoding are. To see why, let me illustrate the transmission of empirical information by means of a toy example.

Suppose the position of three islands in a bounded region at sea, known to sender and receiver, is to be transmitted. An encoding could be achieved by splitting a map of the relevant area into 35 sectors and sending a binary string whose thirty-five digits are 0's or 1's. Once received the string might be decoded as a drawing of a 5×7 grid in which the islands are positioned by the three occurrences of 1 in the given string. Decoding consists in finding the prime decomposition of the string's length and using the factors so obtained, in increasing order (corresponding to: rows; columns), to draw the map, marking the squares or 'pixels' that correspond to occurrences of 1.

The example just given is relevant for two reasons. First, because it suggests that we do not have to use a structure-preserving map to retrieve empirical content from information transmitted in numerical format. Second, because it indicates how we might view the strong homomorphisms used in measurement theory as *special cases* of encoders. Let me clarify this latter point by referring again to the example from Sect. 1: if someone was presented with the values $f(a_i), f(a_j)$ and knew that they came from a psychological experiment concerning preference rankings, knowing that f is an order-homomorphism would be enough to indicate that, by checking the relative ordering of the numbers $f(a_i), f(a_j)$, one could retrieve the preference ranking of a_i, a_j.

The two key points I wish to make in this section are: (i) that measurement theory works with codifications of empirical content, only some of which are obtained via strong homomorphisms; (ii) that measurement theory does not restrict its scale values to numbers.

Both points will be relevant to discussing (Heilmann, 2015), as we shall see. For now I only wish to stress that (i) is the reason why I prefer to adopt the denomination 'measurement theory' rather than the more common 'representational measurement theory' (often reduced to the acronym RTM). I see the latter as a subprogramme of measurement theory[3], but by no means as the whole of measurement theory.

[3] RTM has been so understood by its practitioners. For instance, (Narens, 2002a) gives abstract conditions under which an abstract notion of 'codifying content' specialises to invariance under scale transforma-

Codifications other than homomorphisms appear in various measurement-theoretic contexts. I describe three below.

1. **Homogeneous linear inequalities**: Krantz et al (1971: 59–70) contains a study of measurement contexts in which a linear model implies a finite system of equalities or inequalities in n unknowns. The numerical scaling goal is to solve such a system of equalities and inequalities, assumed to be of the form:

$$\sum_{j=1}^{n} \alpha_{ij} x_i > 0 \, i = 1, \ldots, m'; \; \sum_{j=1}^{n} \beta_{kj} x_i = 0 \, k = 1, \ldots, m'',$$

   with α_{ij}, β_{kj} integers. In such cases solvability depends on geometrical considerations involving the subspaces of $\mathbb{R}^n$ generated by the integer-valued vectors involved. No class of structures needs to be declared for scaling values to be obtained: solutions to the system of equalities and inequalities may be determined algorithmically. *A fortiori*, no representation theorem guaranteeing the embeddability of a class of structures into a numerical model is called for.

2. **Conjoint measurement**: Conjoint measurement, discussed in chapter 6 of (Krantz et al., 1971), arises when, instead of ranking individual items, as in the toy example from Sect. 1, an experimenter wishes to rank 'compound' items, which may be viewed as ordered tuples.
   A concrete illustration is afforded by the scaling of rats' problem-solving performance relative to food deprivation d, food reward k and prior learning h, as originally investigated in Hull (1952) and Spence (1956). Performance is then abstractly typified by a set of triples of the form (d, k, h). Comparisons of performance levels rank triples. If the relation $\precsim$ is the performance ranking, an abstract, qualitative model of performance may be described as a pair $\left\langle D \times K \times H, \precsim \right\rangle$, whose domain is a set of performance triples.

   As Krantz et al. (1971: 316–317) remarks, Hull (1956) suggests that performance level rankings are numerically codified by the condition:
   $(d, k, h) \precsim (d', k', h')$ iff $[\phi_D(d) + \phi_K(k)]\phi_H(h) \leq [\phi_D(d') + \phi_K(k')]\phi_H(h')$,
   where ϕ_D, ϕ_K, ϕ_H are three distinct measurement scales. If performance levels satisfy the formal conditions discussed in Krantz et al (1971: 348), they may be modelled as a conjoint structure $\langle D \times K \times H, \precsim \rangle$, which admits the above numerical encoding.

   The encoding relates a Cartesian product endowed with a binary relation to the ordered ring of the real numbers. It is worth noting that such a Cartesian product – referred to as a conjoint structure if it satisfies suitable conditions – is not

tion, a criterion associated with the representational approach. Other abstract notions of 'codifying content' are available. A range of them is discussed in (Narens, 2002b).

a structure in the model-theoretic sense. Its presentation as an ordering $\langle A, \preceq\rangle$ hides the fact that A results from the application of set-theoretic constructors (iterated powerset and union) to D, K, H. This is important because the internal set-theoretic structure of A enters its numerical codification.
When we regard $\langle A, \preceq\rangle$ as a model-theoretic structure, and set:

$$\Phi(d,k,h) = [\phi_D(d) + \phi_K(k)]\phi_H(h),$$

then Φ is an order-homomorphism between a conjoint structure and the structure $\langle \mathbb{R}, \leq\rangle$. Triples are treated as individuals, with no regard to their internal structure.
When, on the other hand, we consider the full ordered ring structure on the reals, we note that it allows us to define a six-place relation whose arguments are values of ϕ_D, ϕ_H, ϕ_K. This six-place relation is what provides an insight into the fact that the ordered triples in A have a component-wise structure. Since the conjoint structure carries no such relation, we are not in presence of a homomorphic embedding. Rather a triple of homomorphisms (which is not itself a homomorphism in the given context) is used to codify the behaviour of structured triples.
In this case, as already noted in Adams (1979: 216), we move beyond a purely representational standpoint.
More general polynomial functions of any finite number of scales are also possible. It is such functions that codify relations over tuples in conjoint measurement, not homomorphisms.

3. **Expected utility**: In chapter 8 of (Krantz et al., 1971), expected utility structures are discussed. These involve assigning utility values to gambles, ranked according to preference. The expected utility setup has no natural rendering in terms of structures and homomorphisms.
To see this, note that Krantz et al. (1971: 381) call an expected utility structure a 6-tuple of the form: $\langle X, \mathcal{E}, \mathcal{N}, \mathcal{C}, \mathcal{D}, \preceq\rangle$.

Here X is a sample space, $\mathcal{E}$ is its associated algebra of sets, $\mathcal{N}$ a subset of $\mathcal{E}$. If X is not finite, an expected utility 'structure' is not a structure in the model-theoretic sense, i.e. a set equipped with finitary relations, functions, and constants, because it carries a family of subsets closed under the infinitary operation of countable union.
Moreover, an expected utility codification is a pair (u, P) in which u assigns real numbers to $\mathcal{D}$ alone and P is a probability measure over the sample space $\langle X, \mathcal{E}\rangle$, which is not a structure in the sense specified.
Thus, if we take homomorphisms to be maps between structures in the ordinary sense, we cannot have them here because we do not have an ordinary structure to work with and we do not have a single embedding into a homomorphic real structure.
As in conjoint measurement, set-theoretic constructors play a role that is kept track of by the chosen numerical codification but not in terms of homomorphic mappings. To see this, we complete the description of an expected utility 'structure'. Its ordering $\preceq$, which we may suppose a weak linear ordering (intuitively,

behaving like $\leq$ on numbers), is defined only on a set of gambles $\mathcal{D}$, whose outcomes are the elements of $\mathcal{C}$, itself not in the range of the codification (u, P). If A is an event from a given sample space, the associated gamble $f_A \in \mathcal{D}$ specifies, for each 'outcome' $x \in A$, a payoff $f_A(x)$.

No binary operations are written into an expected utility structure. Nonetheless, when A, B are disjoint events and the gambles f_A, g_B are in $\mathcal{D}$, the gamble $f_A \cup g_B$ is supposed to be in $\mathcal{D}$, too (see Krantz et al (1971: 376)). This gamble is defined on the event $A \cup B$ and it behaves like f_A over A and like g_B over B. Under the conditions given on the various set-theoretic objects making up an expected utility 'structure', it is possible to prove that the following equality, in which conditional probabilities 'weigh' utility values, holds:

$$u(f_A \cup g_B) = u(f_A)P(A|A \cup B) + u(g_B)P(B|A \cup B).$$

It is clear that the expected utility of $f_A \cup g_B$ is described by an operation in the ordered ring of nonnegative reals, which depends on $u(f_A), u(f_B)$. An expected utility 'structure' is not equipped with a binary operation on gambles, yet the fact that certain gambles can be spliced is kept track of by u, which codifies this information without mapping it homomorphically (u is only a homomorphism between $\langle \mathcal{D} \preceq \rangle$ and $\langle \mathbb{R}, \leq \rangle$).

The second fact left to illustrate in this section is that measurement-theoretic codifications can be carried out, even in a representational fashion, without using numbers or a numerical structure. An instructive example is discussed in Niederée (1987, 1992), which takes 'systematic objective indices of phenomena' to be certain model-theoretic types[4] realised by the elements of a given non-numerical structure.

Consider for instance a structure $\mathcal{X} = \langle X, \prec^X, \circ^X \rangle$, where $\prec$ is a linear order and $\circ$ a binary operation. If $\mathcal{X}$ satisfies the defining conditions of extensive structures[5], then it is homomorphically embeddable into the positive, additive reals.

Niederée proposes to fix $a \in X$ and to associate the elements of X with suitable subsets of $At(x/a)$, the set of atomic formulae with one free variable in the signature $\{\succ, \circ\}$ over the set of parameters $\{a\}$(i.e. only a may occur in the relevant formulae as a parameter).

More explicitly, $b \in X$ is measured by the subset of $At(x/a)$ whose elements it satisfies. Such a subset is called a 1-type over $\{a\}$. Because atomic formulae are, in the given signature, inequalities between multiples of a and multiples of b, type assignment is equivalent to a representation of b by successive rational approximations relative to the unit of measure a.

Niederée's approach does not stop at the example I have briefly described. Among classes of structures studied by measurement theorists, the difference structures dis-

[4] A type in the model-theoretic sense is simply a consistent set of formulae in some fixed first-order language. A type may be over a set of parameters, i.e. the formulae may contain names of distinguished objects from some specified domain.

[5] See e.g. Krantz et al (1971: 73) for a list of such conditions. An extensive structure may be viewed as a non-numerical model of an empirical variable like length. The ordering $\prec$ determines length-comparison and the operation $\circ$ combines lengths additively.

cussed in chapter 4 of (Krantz et al., 1971), additive conjoint measurement (in chapter 6 of Krantz et al., 1971) and the positive concatenation structures studied in chapter 19 of Luce et al (1990) are readily amenable to a type-theoretic treatment (in such cases we cannot only work with 1-types, but we need 2-types or n-types with $n > 2$, depending on the number of conjoint components). In general, there are structures that lend themselves to the model-theoretic scaling introduced by Niederée but do not admit of any numerical representations.[6]

Niederée's work does not only show that measurement values need not be numerical indices, but also that there can be distinct varieties of representational measurement. The one he proposes does not call for a preliminarily fixed numerical structure: measurement values are directly generated by an empirical variable, which structures them as linguistic reports that codify relevant empirical information.

What I hope to have highlighted so far is the measurement theorist's need to overcome the canonical representational goal of constructing homomorphisms into fixed numerical structures. With a greater awareness of this need in mind, it is possible sharply to identify the limitations and shortcomings of the philosophical account of measurement theory contained in Heilman (2015).

3 Descriptive dislocation

Heilmann's starting point is the reasonable observation that measurement theory may not be plausibly viewed as offering a complete account of measurement (see Heilmann (2015: 787)). This is clear in view of the fact that studies of calibration, error, and other themes crucial to the experimentalist or metrologist are not an integral part of mathematical work on measurement.

The initial observation, which rules out the tenability of a view of measurement theory as a *complete*, foundational framework for measurement is, however, escalated into a descriptive dislocation, i.e. the portrait of of formal work that is in the first instance self-contained and relatively insulated from scientific work to which it may later apply.

This is clear from Heilmann's insistence that measurement-theoretic work is capable of supporting theory construction, rather than having emerged for that purpose. According to Heilmann, measurement theory aids theorisation in mainly in two ways, i.e. by supplying various, structured notions of a scalable attribute ('specifying conditions for mapping' (Heilmann, 2015: 792)) or by seeking to relate numerical practices to underlying structural features ('backward engineering' (Heilmann 2015: 793).

[6] A purely illustrative example is the lexicographic ordering of the set of ordered pairs $A \times \{0, 1\}$, where A is the domain of a continuum in Cantor's sense (see section 4). For any $a \in A$, the pair $(a, 0)$, $(a, 1)$ defines a 'jump': no pair in $A \times \{0, 1\}$ lies strictly between these two endpoints. Because A is uncountable, there are uncountably many jumps. On the other hand, any substructure of the ordered reals has at most countably many jumps. It follows that the lexicographic order just defined is not embeddable into the ordered reals. It nonetheless admits a type-theoretic scale. Relative to its ordering as a continuum, A includes an order dense, countable subset B. The set of parameters $B \cup \{(a, 0) : a \in A\}$ guarantees the representation of the lexicographic ordering onto a set of quantifier-free types ordered by inclusion.

These functions are underpinned by a conception of (representational) measurement theory as a 'library of theorems' (Heilmann, 2015: 787) or a repertoire of formal results that may happen to be useful to the modeller.

The conception is unhelpful for three reasons. First, it suppresses the fact that, since at least Suppes and Zinnes (1963), representation theorems in measurement theory were sought with the explicit intention of formalising various notions of a scalable empirical variable pertinent to experimental settings in social science, especially psychology. If anything, experimental work has provided a library of scaling contexts for the measurement-theorist to work on.

Second, Heilmann remains silent on the possibility that measurement theorists may adopt codifications of empirical information that are not homomorphic embeddings. Implicit emphasis on representation theorems conceals the fact that the conditions of scientific enquiry may call for mathematical departures from the canonical notion of homomorphic embedding. That they have in fact called for such departures was shown in the previous section. Conjoint measurement has been in part motivated by behavioural experiments, while the economic theory of expected utility has led to non-representational, measurement-theoretic models.

Third, Heilmann seems entirely to neglect the measurement-theoretic concerns with the compatibility between a formal (representable) model and a finite data set. This compatibility problem arises when finite data, in the form of a list of inequalities, arises from an experimental setting. Typical situations are illustrated by factorial designs, an example of which is provided by the experiments on rats mentioned in the previous section. Those experiments involved three 'factors' i.e. the variables D, K, H, and generated finite lists of inequalities involving ordered triples of the factors.

To ask whether concrete behavioural data produced by such experiments is compatible with the formal model of a conjoint structure admitting a specified polynomial representation (e.g. the one proposed in Hull (1956)) is not a matter that can be resolved entirely by testing consistency of the data with the formal conditions that define a suitable class of conjoint structures.

There are situations in which:

> even if some conditions are true for the underlying data-generating process (e.g. weak ordering, solvability, Archimedean condition), and the others are true for a particular set of data (e.g. independence), this does not guarantee the existence of an appropriate numerical representation for those data. The existence of such a representation depends on the existence of solutions to a set of simultaneous linear or polynomial inequalities Krantz et al (1971: 426).

There is, in other words, an important difference between showing that an experimental setting may be viewed as a conjoint measurement structure and showing that a finite amount of factorial data may be embedded into a conjoint measurement structure. In the first case, the behaviour of the factors implies at once that they have distinct representations and that these representations admit of a specified polynomial combination, which yields a conjoint scale. In the second case, the behaviour of the factors is to be checked by looking for the numerical solvability of a system of inequalities.

This kind of check requires solving genuinely measurement-theoretic problems that are not representation problems. One of them, pertinent to conjoint measurement, is to find a solvability criterion for a set of polynomial inequalities: its solution is not a representational result, but a purely algebraic one on the completability of a finite order on a ring of polynomials $\mathbb{R}[Y]$, where Y is a finite set of unknowns associated with a factorial design (for a complete statement of the relevant theorem, see Krantz et al (1971: 447)).

Heilmann's account of measurement theory centres on representation theorems, seen as purely formal results that, independently given, may later be involved in the modeller's concerns. What I have tried to throw into relief is that the entire arc of development of measurement-theoretic unfolds in response to the concerns of scientific, in particular experimental, work.

Not only are representation theorems transcended in response to experimental needs; conditions are explicitly studied, under which non-numerical structures furnish plausible models of data. Heilmann's account of measurement theory as primarily formal loses sight of its intrinsic solidarity with experimental work. Measurement theory is not a formal framework that may be linked to modelling concerns: it is the modeller's mathematical work. When this work and its distinctive features are made visible again, the problem of deciding how measurement theory is to be connected to scientific work (what is known in the literature as the 'interpretation problem' of measurement theory) vanishes. This is because a description of how measurement theory *is* connected to scientific work can be supplied.

4 Measurement theory: classification

A second dislocation of measurement theory, other than descriptive, occurs when its results are regarded as external constraints on what modelling practices are acceptable. This perspective produces a normative dislocation of measurement theory. Its results are no longer co-regulated with other aspects of scientific enquiry, but stand aside as rigid criteria dictating what course scientific enquiry is to take.

Normative dislocation strikingly occurs in (Philippi, 2021). An adequate discussion of Philippi's work requires a preliminary discussion of certain important measurement-theoretic results, which constitute the main subject of this section.

The results in question concern the classification of scale types over continua, i.e. ordered relational structures that are order-isomorphic to $\langle \mathbb{R}, < \rangle$. For present purposes, we restrict attention to finite scale types. Given an empirical variable $\mathcal{X}$ that is a continuum in the sense just specified, its (finite) scale type is specified by a pair of integers (m, n), with $m \leqslant n$. These integers are called, respectively, the degree of homogeneity and the degree of uniqueness of the scale type.

Homogeneity and uniqueness, are, in these contexts, properties of the automorphism group of $\mathcal{X}$.[7] The continuum $\mathcal{X}$ is said to be m-point homogeneous if any two ordered m-tuples of elements of $\mathcal{X}$ are related by an automorphism in an order-preserving fashion.[8] The same continuum is said to be n-point unique if the only automorphism with at least n fixed points is the identity.

It is worth stressing that a finite scale type is defined intrinsically, i.e. independently of any numerical representation. It is only sensitive to how rich the automorphism group of an abstract structure is.

The connection with numerical scales may be understood by a simple example: an additive empirical variable like length is classically identified, up to isomorphism, with the additive continuum $\mathcal{R} = \langle \mathbb{R}^+, <, + \rangle$, where $\mathbb{R}^+$ is the set of positive reals. The automorphisms of the latter structure are the multiplications by a positive real constant, which correspond to the scale changes induced by a change of unit of measure. Thus, $\mathcal{R}$ is 1-point homogeneous.

Moreover, for any two scales f, g for length, seen as isomorphisms from an abstract additive continuum onto $\mathcal{R}$, the function $f \circ g^{-1}$ is an automorphism of $\mathcal{R}$. When f, g have the same unit of measure ($f(a) = 1 = g(a)$), $f \circ g^{-1}$ has one fixed point. In this event, $f \circ g^{-1}$ actually fixes every point, i.e. it is the identity automorphism. The fact that a scale is fixed by the specification of at least one value is thus the 1-point uniqueness of $\mathcal{R}$.

This lengthy technical digression was necessary to state a fundamental result on scale types established in (Narens, 1981b) and (Alper, 1987) (a comprehensive discussion appears in chapter 20 of Luce et al (1990)). Alper and Narens showed that, if the finite scale types of a continuum are of the form $(1, n)$, then $n \leqslant 2$.

What this means is that the decision formally to regard an empirical variable as a Dedekind-complete, dense linear order has a dramatic effect on the range of scales that it will support. Scales determined by m parameters, with $3 \leqslant m$, are automatically ruled out. The types of scale one may expect for a continuum, under the hypotheses cited, are only three, namely $(1, 1), (1, 2)$ and $(2, 2)$. The largest number of parameters determining a measurement scale over a continuum is 2. Intuitively, these parameters correspond to the choices of an origin and a unit.

Attempts at evaluating the philosophical significance of these results may be found in (Baccelli, 2020) and (Wolff, 2020). My main interest in the present section is to clarify what these results actually say and in what way they may be viewed as setting constraints on modelling practice. For these reasons I will not engage with the metaphysical discussion of scale types offered in (Wolff, 2020). (Baccelli, 2020) takes the classification of scale types to illustrate the fact that measurement theory is

[7]We recall here that, given a structure $\mathcal{X}$ with domain X, the automorphisms of $\mathcal{X}$ are the structure-preserving permutations of X. Such permutations determine a group relative to functional composition: the group identity is the identity permutation, which fixes every element of X, while the inverse of an automorphism α is the permutation α^{-1} that 'puts back' the elements of X moved by α. For a concrete example, any multiplication by a fixed positive real is an automorphism of $< R, <, + >$. In fact, the latter structure has no other automorphisms.

[8]This means that, if $a1 < ... < am$ is the ordering of the first m-tuple and $b1 < ... < bm$ that of the second, then there is an automorphism α such that $\alpha(a_i) = b_i, i = 1, \ldots, m$.

not exclusively concerned with constructing measurement scales: with this remark I concur.

The philosophical import of what I say about scale types here will fully emerge in the next section, when I discuss (Philippi, 2021) and its interpretation of results that involve the classification of 'admissible' scales. What I will have to say there depends on the brief analysis of known classification results to follow.

What the result obtained by Alper and Narens shows is that the ordered reals supply an extremely limited range of 1-point homogeneous abstract models of an empirical variable.[9]

This result does not primarily delimit, on purely formal grounds, which scales are acceptable, but directs the modeller away from Dedekind-complete ones, if she has reasons to doubt that she is working with an empirical variable whose 'degrees of freedom' are not allowed by continua.

Continua, after all, do not impose themselves as the unique resort open to the modeller, even though they are extremely helpful, e.g. because they enable the use of differential equations.

It is known, for instance, that there certain abstract models of an empirical variable with numerical representations that cannot be Dedekind-completed: such is the case of certain positive concatenation structures (see Luce et al (1990: 53–54)). A modeller handling these structures is not subjected to the restrictions implied by the classification of Alper and Narens.

Much more dramatically, a modeller viewing measurement scales on the ordered rationals as more realistic than real-valued scales will not have ruled out any finite scale type (m, n), with $m < n$, by a deep result proved in (Cameron, 1989). In fact, Cameron's theorem indicates that Dedekind completion may prove a serious distorting factor in measurement: it collapses finite scale types with any degree of homogeneity and uniqueness onto types with uniqueness degree in $\{1, 2\}$. This situation is remarkable: it presents us with a case in which choosing a structurally richer mathematical description does not increase, but dramatically curtails, our ability to adjust to possible empirical variations.

Finally, there is no absolute norm forcing us to classify scale types on the basis of automorphism groups. A symmetry-based classification is certainly insightful, but it does not preclude the adoption of alternative classification criteria. A need for them has long been registered by measurement-theorists themselves:

> [T]here is little doubt that other principles of classification are needed because structures of these types range from highly regular ones with rich families of automorphisms [...] to structures with no automorphism except for the identity, which simply reflects their great irregularity (R.D. Luce et al., 1990: 122).

The foregoing reflections on the import of a symmetry-based classification of scale types for continua offer all the technical background required to examine the normative dislocation of measurement theory occurring in (Philippi, 2021).

[9] Alper showed that, if we drop the transitivity of the automorphism group, i.e. 1-point homogeneity, we can have any finite scale type of the form $(0, n)$.

5 Normative dislocation

A famous informal classification of measurement scales, due to the psychologist S.S. Stevens (see Stevens, 1946), distinguishes types of measurement scales by specifying, for each of them, a set of admissible scale transformations, which may be understood as groups of automorphisms of the ordered reals.

The largest such group contains the strictly increasing, monotonic transformations (this group singles out the 'ordinal' scales).[10] Subgroups of interest to Stevens consist of the positive linear transformations – associated with what Stevens called 'interval' scales – with domain $\mathbb{R}$, and the multiplications by a positive constant – associated with what Stevens called 'ratio' scales – with domain $\mathbb{R}^+$.

Stevens' classification comes reasonably close to the classification of finite scale types obtained by Alper and Narens under the assumption of 1-point homogeneity. Key differences are that: (i) Stevens' ratio and interval scales are closely related with the scale types (1, 1) and (2, 2) respectively, but they do not share the same domain: they can both be represented on $\mathbb{R}$ by taking the natural logarithm of ratio scale transformations, which are then analytically represented as translations by a positive constant; (ii) Stevens does not identify a group of scale transformations intermediate between (1, 1) and (2, 2).

Stevens' classification is the main reference in (Philippi, 2021) for what is called the 'received view' on admissible measurement scales: on this view, Stevens' classification should provide a normative grid against which the tenability of measurement practice is to be evaluated. More precisely, Stevens' classification is taken to require that, whenever an ordered empirical variable is studied, its measurement scale should be of ordinal, interval of ratio type, and cannot be of any other type. Second, once a scale type is fixed, only the numerical statements preserved under the corresponding group of automorphisms of the ordered reals can be considered empirically meaningful.

Philippi contrasts this highly prescriptive view with the requirements of scientific practice and observes that it is plausible to envisage research contexts in which the set of scale transformations may be neither ordinal, nor interval, nor ratio: for instance, such would be the case with the set of order-preserving functions that, beyond being strictly increasing, are twice differentiable and satisfy the conditions $f'(x) > 0$ and $f''(x) \leq 0$ over a fixed real domain (the examples to follow take this domain to be $\mathbb{R}^+$).

Philippi thus identifies a proposed set of scale transformations included in the set of strictly increasing, monotonic transformations, but different from the subgroups of multiplications by a real constant or the multiplicative version of interval transformations.[11] This choice does not only transcend Stevens' classification, but also departs from the mathematical classification given by Alper and Narens. Furthermore, Philippi's set of transformations is not in general a group but only a monoid, because it

[10] Stevens also considers the full symmetric group over a fixed numerical set, i.e. the group of all its permutations. For present purposes, we may ignore this aspect of his classification.

[11] On the positive reals, these are power transformations described by conditions of the form $x \mapsto \alpha x^{\beta}$, with $\alpha, \beta > 0$

does not contain the inverses of its elements (on $\mathbb{R}^+$, it contains the function defined by the condition $x \mapsto x^{1/2}$ but not its inverse on the positive reals, which has strictly positive second derivative).

Philippi's key point is that it would be unreasonable to rule out scaling practices that depart from the scale types recognised by Stevens. Philippi does also acknowledge the classification of scale types discussed in the previous section, in a brief mention of (Narens, 1981a) (see Philippi (2021: 934 n)). Philippi's reaction to this result is a rebuttal of what is recognised as an illegitimate normative stricture: the response to it recommended in the paper consists in abandoning a representational – which here means: measurement-theoretic – framework altogether. Philippi notes that:

> 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 on the inferences researchers can validly make with measurement results (Philippi: 934).

Such a sharp turn away from mathematical considerations, abandoned in favour of an explicit attention to researchers' beliefs, comes, it seems to me, from an impression of mathematical results as hinderances to scientific practice, which I find important to dispel. The classification of scale types that Philippi finds to be at odds with the plausibility of his proposed set of transformations is not a measurement-theoretic imposition. It is the consequence of certain choices, none of which is compulsory on mathematical grounds.

These choices consist in: taking measurement to be real valued (as opposed to rational valued, by Cameron, 1989); taking an empirical variable to be a continuum (as opposed to e.g. a suborder of a continuum or even a numerical structure lacking a Dedekind completion); allowing only highly regular scale types (with rich automorphism groups); choosing scale transformations to possess full group structure. All of the above choices can be dropped without leaving the domain of measurement theory: none of them is forced by it. The use of monoids of scale transformations that are not groups is already part of measurement-theoretic work, as shown in Narens (2002): 318–319.

Philippi is led to a dismissal of measurement theory because the theory was presented as a coercive intervention upon scientific work, which it is not. Philippi is certainly not wrong to suggest that the information of interest supposed to be captured by a measurement model should be responding to researchers' beliefs, but also heavily underestimates the requirements to which these beliefs are subjected when they have to be spelled out in terms of a mathematical model.

For instance, a researcher keen on assuming smoothness conditions like those described by Philippi, or making use of basic theorems of reals analysis that are equivalents of Dedekind completeness could not always do this consistently with a symmetry-based notion of scale type. This is all measurement-theoretic work alerts the modeller to.

Insofar as measurement theory keeps track of the constraints attending the adoption of certain formal assumptions on mathematical models, it can work as a useful instrument to prevent the researcher's beliefs, which Philippi evokes, from degenerating into a cognitively arbitrary employment of formalisms. We are not in presence of a normative straitjacket, but an aid to navigate the space of modelling possibilities.

6 Measurement theory: experimental and theoretical connections

The third type of dislocation mentioned in Sect. 1, it will be recalled, is methodological. This dislocation underlies (Tal, 2021). In order adequately to examine the essential critical points of Tal's contribution, a preliminary discussion of the connections between measurement-theoretic work, experimental work and scientific theorising is necessary. This section will offer such preliminary discussion, while the next section will directly address, in its light, Tal's critical examination of measurement theory.

The central point I wish to emphasise here is that different aspects of measurement theory acquire varying degrees of salience depending on the scientific research contexts to which they are connected.

For instance, measurement-theoretic work developed under the pressure of numerical practices in behavioural science will not rise to comparable salience when formally related to experimental physics. Similarly, the reconstruction of dimensional analysis from an underlying algebra of physical quantities (studied in chapter 10 of Krantz et al., 1971) will not play a major role in the social scientist's work.

Trivial as these remarks seem, they helpfully indicate that it is counterproductive to view measurement-theoretic work as providing a uniform foundation for measurement practices across the sciences. It is more realistic to think of measurement theory as a range of results that make contact with experimental or theoretical work in the sciences in a variety of distinct ways. A distorted picture of measurement theory is bound to arise from the decision to assign its more specialised contributions an unconditional significance.

While this decision may result, as we shall see, from a distinctive methodological interpretation of measurement theory, as in (Tal, 2021), it may also be the effect of traditional philosophical prepossessions lingering in measurement-theoretic work, not least the decision to understand this work as foundational, which pervades (Krantz et al., 1971) and Luce et al (1990).

Foundationalism evokes the idea that a primitive, and thus unique, basis of enquiry should be singled out. In light of this suggestion, it is easy to view 'foundational results' as the instruments of dogma: they proscribe what is illegitimate and establish an orthodoxy of concept and procedure. Philppi's understanding of a 'received view' of measurement as issuing proscriptions on scale types, as well as Tal's insistence on the narrow-minded methodological demands of measurement theory, result from understandable responsiveness to foundational associations evoked by casual philosophical remarks encountered in the writings of measurement theorists.

These remarks are of marginal significance relative to the actual mathematics of measurement theory. When the latter, its motivation and its consequences are straightforwardly examined, no substantial reason to associate foundational prejudice with this work will be found. An effective way to support the last claim consists in showing that measurement-theoretic work can be viewed as a complex of problem-solving approaches called forth by specific difficulties or questions independently occurring in experimental or theoretical work across various disciplines.

I begin by clarifying the connection between measurement theory and experimental work. It is fair to say that the original motivation for measurement theory was the problem of determining whether sound measurement practices could be established in psychology.

The problem was not purely methodological since scientific and practical goals were already driving experimental attempts at constructing numerical scales before the rise of the modern theory in the 1950's.[12]

Measurement theorists devoted their attention to numerical practices in which data is responses elicited from subjects under a controlled experimental setup. Psychophysics provides an especially apt example. In classical psychophysical measurement, experimental subjects act as 'measuring devices', either by estimating, say, the subjective loudness of a sound against a set benchmark, or by producing a prescribed loudness level, e.g. an average of two presented acoustic stimuli a, b. Measurement-theoretic work has been, to a large extent, an attempt to identify conditions on the responses elicited that ensure their codifiability on a numerical scale.

In order to understand how formal modelling interacts with experiment in such a context, it is instructive briefly to consider two examples. Take first the production of acoustic averages mentioned earlier. A setting is given in which, given a, b, a new 'average' stimulus aIb is produced by an experimental subject.

Because an averaging experiment produces a set of behavioural responses, it is of interest to ask whether combinatorial conditions on such responses may be found, which can both be tested and contribute to specifying the empirical information that may be numerically codified. What follows is an interplay between the formal isolation of conditions that support averaging scales and experimental checks that such conditions are satisfied.

The interplay occurs as an exchange between experimental practice and formal theory construction. In the example of acoustic averaging, an important question is to determine whether the response $a|b$, elicited by the presentation of a followed by the presentation of b, is equivalent to the response $b|a$, elicited by the presentation of b followed by the presentation of a.

Slightly more formally, this is to ask whether experiment supports the numerical codification of commutative or non-commutative 'averaging'. Since Pfanzagl et al (1973: 122), it is known that $b|a$ is consistently louder than $a|b$, so that the commutativity of can be finitely refuted and should not be included in the measurement-theoretic rendering of an 'averaging' operation.

While thus the experimental psychologist highlights certain constraints for the measurement-theorist, the latter, in the capacity of a modeller, is also in a position to offer guidance to experimental work. This dynamics is illustrated by my second example, the psychophysical theory of magnitude estimation axiomatised in (Narens,

[12] The many experiments carried out by Stevens on subjective loudness scales since the 1930's (see Stevens and Davies (1938)) were both an integral part of the wider physiological study of hearing and a response to practical issues raised by unreliable estimates of industrial noise that used Fechner's scale (see in this connection Churcher, 1935).

1996). Narens proposes abstractly to treat the responses of magnitude estimation experiments as triples $(x, \mathbf{p}, t)$, where t is a fixed modulus or benchmark stimulus, x a presented stimulus and $\mathbf{p}$ the numerical term or numeral (thus, not a number) with which a subject responds to a stimulus.

Against this framework, Narens gives conditions under which triples of the form $(x, \mathbf{p}, t)$ are described by numerical functions φ_t, parametrised by moduli. The key combinatorial condition supporting this construction is a form of commutativity, which it was left to experimenters to test. Successful testing is recorded in Ellermeier and Faulhammer (2000) and Ellermeier et al. (2003).

What the above examples show is that, in a specialised context like that of psychophysics, measurement theory can fruitfully interact with experimental settings in which the problem of scaling may be viewed as the problem of constructing a numerical codification of overt, non-numerical responses.

It evidently does not follow from this that all experimental practice should conform to the procedures measurement theorists were particularly interested in.

It is nevertheless true that measurement theorists have also studied formal models whose typical reference is not quantitative work in social science, but mathematical physics. Such is the case of extensive measurement, when related to canonical physical quantities like length, mass or time.

The purpose of such formal models cannot be to provide the experimental physicist with directions that preside over scale construction. Tal (2021: 720–721) offers compelling reasons from experimental practice to rule out such a role as a plausible use of extensive measurement.

To accept this conclusion is, however, not to reject extensive measurement. Whenever position or velocity enters a differential equation as a sufficiently smooth function of time, the latter is automatically conceived as a continuum endowed with extensive structure. It is within a theoretical, as distinct from an experimental, context, that extensive measurement proves its fruitfulness.

In a theoretical context, knowing that certain empirical variables carry structure that fixes their scalability (e.g. extensive scales over a continuum are of type $(1, 1)$) provides, among other things, valuable insight into the form of laws into which these variables may enter.

To clarify the last remark, I offer brief indications of measurement-theoretic work that tackled the question of the possible forms of laws involving scaled empirical variables. In a pioneering paper (Luce, 1959), Luce observed that, if x, y are extensive magnitudes conceptualised as continua and if, roughly, the scale changes of an independent variable x induce corresponding scale changes of a dependent variable y, then y, x are related by a power law.[13] (Luce, 1964) generalises this result to y depending on a finite family of independent, extensive variables, obtaining the analytical form of a product of powers ubiquitous in physics.

These results show how the analytical treatment of quantities constrains the forms of equations in which they may enter. In particular, they afford a generalisation of dimensional analysis. This generalisation does not stop at Luce's results but extends much further.

[13] That is, there are real numbers r, s such that $y = rx^s$, with s independent of scale choice for x, y.

For instance, Falmagne and Narens (1983) studied scenarios in which y depends on $x_1, \ldots, x_n$ and scale changes on the latter n-tuple are constrained by a nonempty relation.[14] In the same paper, Falmagne and Narens also considered the general situation in which the relationship between y and $x_1, \ldots, x_n$ belongs to a family of functions, thus seeking to address a problem already recognised in Luce (1959: 85), in which scale changes of an independent variable do not induce corresponding scale changes of the dependent variable.[15]

Moreover, Luce's work has been extended to a wider family of scalable variables in (Kim, 1990) and expanded from the standpoint of structural stability in (Yoshino, 1989), whose results allow one to derive information about finer constraints on scientific laws, e.g. bounds to the exponent in the formal power law derived by Luce.

This quick survey of lesser known measurement-theoretic work is meant to drive home the point that distinct measurement models function differently because they relate to different scientific contexts. Some of them afford a relatively direct aid to the social scientist engaged in experimental work while others cast light on the far-reaching effects of the analytical mathematisation of empirical variables, as pertinent to the mathematical physicist especially. Each of these functions is localised because it emerges from special scientific concerns and special choices of mathematisation.

The remarks just made enable us to consider (Tal, 2021) in close detail.

7 Methodological dislocation

In a rich and important 2021 contribution, Eran Tal has taken issue with measurement theory as a foundational framework presiding over experimental work and bound up with a strictly empiricist methodology. I have called this view of measurement theory the result of methodological dislocation.

The critical stance presented in (Tal, 2021) can now be reviewed in the light of what was observed the previous section.

Tal's leading concern is, as just noted, with the strictly empiricist foundational stance he believes inevitably to accompany measurement-theoretic work. Tal finds this stance openly at variance with at least measurement practice in physics and deems it philosophically harmful insofar as it contributes to propagating certain empiricist 'myths' concerning measurement.

Two myths seem to Tal inherent into the conceptual architecture of measurement theory. The first myth is that the structural traits of an empirical variable can be directly extracted from raw, non-numerical data, which are later subjected to numerical scaling. In short, the detection of qualitative structure *must* precede numerical assignment in measurement practice. The second myth is:

[14] If T_i is the set of scale changes associated with x_i, then the relation in question is a nonempty subset of $T_1 \times \ldots \times T_n$.

[15] This phenomenon, noted in Luce (1959: 91), was stressed in Rozeboom (1962: 545), which illustrates it with reference to the exponential law of radioactive decay. If the law is expressed as $q = ae^{-bt}$, with t time and q the quantity in grams of radioactive material, a scale change of t does not induce a change scale of q, which should be a multiplication by a positive constant.

> [...] the belief that establishing homomorphic mappings from independently abstracted data structures is not only sufficient, but also necessary for empirically detecting quantitative structure (Tal (2021: 715)).

The two myths are not implausibly framed against the type of measurement-theoretic result philosophers have especially been attentive to, namely a representation theorem asserting that any member of a set-theoretically definable class of structures can be homomorphically embedded into a prescribed real structure.

One may regard this type of result as suggesting, in formal terms, an account of measurement as concretely carried out. The account suggested proceeds through three fundamental stages: (a) the extraction of raw, non-numerical data from an experimental setting; (b) the testing of structural features directly exhibited by recorded data; (c) the construction of a homomorphism between the structured variable detected through (a) and (b) and a prescribed numerical structure.

Tal takes issue with this account of measurement, which is in fact evoked by occasional methodological statements made by measurement-theorists (see in particular the reference to Luce et al. (1990) discussed in Tal (2021: 714–715) and Tal's criticism of Krantz's sketch of the concatenation of periods using pendulums in Tal (2021: 722)).

It seems to me that there is no strong reason to regard measurement theory as the carrier of the procedural and methodological conception targeted by Tal's criticism. I do not disagree with the criticism as such, but with Tal's claim about what it shows.

I don't think this criticism shows that measurement theory essentially rests on an unacceptably unrealistic picture of measurement processes. What it shows is that the function and significance of measurement-theoretic work are misunderstood unless they be disassociated from an implausible foundational stance. I presume Tal believes the disassociation to be impossible. In the remainder of this section I shall endeavour to show that it can be made.

What was said in the preceding section already goes a long way toward supporting the last claim. Even focussing on representational measurement alone, it is possible to view it, where Tal indicates its experimental irrelevance, as the starting point of a theoretical study of real scalability and its impact on the formulation of mathematised laws. The theoretical relevance of measurement-theoretic results becomes invisible if one works with the initial hypothesis that they must be about experimental scale construction.

Section 6 was intended to indicate that scale construction is a dominant concern only in some specific contexts, typically coming from the behavioural sciences. It is in such contexts that scale construction based on overt non-numerical responses may be a meaningful practice. Here representational measurement is pertinent to experimental practice for two reasons: first, because it affords intrinsic (i.e. scale independent) presentations of traits whose features may be experimentally controlled or detected (recall the discussion of commutativity conditions in psychophysics); second, because it proves the existence of homomorphic embeddings in a fashion that may be suggestive of a scale construction procedure.

I must stress that what was just said is not a surreptitious way of indicating that empiricist prepossessions concerning measurement may after all hold some ground, but a direct way of noting that, relative to some actual experimental practices, it makes sense to think about scaling along lines indicated by the representational approach.

Representational measurement is thus not to be understood as methodological dogma, but as a mathematical response to *specific* investigations actually ongoing.

As Tal indirectly argues, the specific investigations concerned do not include most theory-laden physical measurement. Acceptance of this conclusion does not, however, automatically render measurement theory irrelevant to the physicists' concerns: it rather calls for relocating its relevance, as indicated towards the close of Sect. 6.

In order to complete the disassociation of measurement theory from the foundational perspective undermined by Tal's critique, it should now be sufficient to show that Tal's two myths can be rejected *on measurement-theoretic grounds alone*.

The first myth is that the structural traits of an empirical variable may simply be extracted from raw, non-numerical data. This myth is inconsistent with existing measurement-theoretic results. Some of them openly, some implicitly, allow data to be numerical and structured in the first instance: such data is then used to detect the intrinsic (non-numerical) features of an underlying empirical variable.

An illustration is offered by a measurement model discussed in Pfanzagl et al. (1973: 158–159). Motivation for this model comes from psychophysical experiments in which subjects are asked to assign numerical values to stimuli (e.g. sounds at varying sound pressure levels). Pfanzagl set-theoretically typifies this setup as a family $\mathcal{F}$ of real-valued functions on an ordered and connected set[16] $\langle A, < \rangle$ (which may be formally identified with an open real interval, if assumed to be a continuum).

If the given *numerical* functions satisfy a suitable list of conditions (see Pfanzagl et al., 1973: 158), then it is possible to use the elements of $\mathcal{F}$ to define a four-place relation on A, under which this set may be structured as a system of comparable intervals, giving rise to a real homomorphic embedding that codifies the subjective comparisons of stimuli in A. In this case, the isolation of a codifiable, non-numerical variable is not the starting point for numerical codification, but its outcome.

Many familiar measurement models can be used in the same way. We know, for instance, that positive-difference structures have a definable additive operation on intervals and that additive conjoint structures carry additive interval structure on each conjoint component. Since the formal conditions that define a positive-difference structure or an additive conjoint structure are obviously satisfied by numerical structures, the corresponding measurement theories can be used as ways of detecting underlying additive operations in presence of difference or conjoint numerical data.

The second myth, i.e. the necessity of homomorphic embedding for representation, only results from exclusive familiarity with single-embedding representations in measurement theory. As observed in Sect. 2, measurement-theoretic work includes scaling strategies that go well beyond proving the existence of a homomorphic embedding or do not depend on it.

Neither of Tal's myths is thus countenanced by measurement-theoretic results, mainly because they can be used in a more flexible manner than Tal seems to contemplate.

[16] Connectedness is intended relative to the interval topology.

A final word is worth spending on the fact that the measurement-theoretic stress on providing ‘coordinate-free’ or non-numerical presentation of empirical variables is not, as Tal suggests, the effect of empiricist allegiance. Its true basis is the need for an explicit isolation of the content that a certain numerical practice seeks to codify. This need may not be severely felt in the context of physical measurement, but it is natural in scientific work that is farther from having reached a comparably thorough degree of mathematisation.

In principle, it is not impossible to work with a notion of content that lacks a non-numerical reference.[17] It nonetheless remains important, and it is probably conducive to finer analyses, to be able to state, in non-numerical terms, what empirical content scales are supposed to codify.

Measurement theorists have pushed the analysis of content to a very refined level by investigating conditions under which numeral responses elicited by experimental subjects are legitimately treated as numerical (this is what Narens did in the 1996 work referred to in Sect. 6).

The concern here is strikingly not with the search for an empiricist foundation, but with control over content that appears in numerical format. One may operate with numbers rather freely, to the point of divorcing them from the study of content: to put it bluntly, one may multiply two figures of debt to obtain a credit, but this legitimate piece of arithmetic in $\mathbb{Z}$ will not provide any safe guidance to financial management.

In a slightly more sophisticated fashion, one may average data to which only ordinal significance is associated but then again the average may be manipulated at will by shrinking or stretching numerical differences between the ordered items. The concern of measurement theorists with non-numerical data is more sensibly understood as the consequence of a difficulty to which the social sciences seem especially exposed, namely the arbitrary manipulation of numerical indices, than as the effect of abstract methodological commitment.

8 Concluding remarks

In this paper I have attempted to show that, as soon as measurement theory is recognised as an integral part of scientific work, making distinct points of contact with experiment and theory, philosophical criticism resting on its dislocation from scientific work no longer applies.

More interestingly, what were originally presented as critical points against measurement theory can be viewed as consequences of measurement-theoretic work itself (the conclusion arose in the previous section with respect to Tal’s ‘myths’). This result is not paradoxical: dislocations portray a picture of measurement theory as foreign to scientific work and at variance with the latter’s proceedings. When relocation is effected, the consistency between measurement-theoretic and broader scientific work is simply made discernible again.

[17] After Stevens’ fashion, fix a numerical measurement structure $\mathcal{R}$ with domain $\mathbb{R}$, a group G of automorphisms of $\mathcal{R}$(which may not be its full automorphism group) and declare content to be, for each positive integer n, the family of sets $A \subset \mathbb{R}^n$ that are fixed by G.

It is, from this standpoint, possible to list some alleged criticism of measurement theory in its defence, as indirectly showing that certain accounts of measurement-theoretic work are actually inconsistent with its motivation and scientific function.

It does not follow, of course, that measurement theory is shielded from all possible philosophical criticism, but only that meaningful criticism could not rest on an unreliable picture of its object.[18] There is no compelling reason for philosophers to scrutinise measurement theory while neglecting a fairly large amount of work in this area or glossing over the subtler implications of what mathematical work is reviewed. I hope that this paper may contribute to emancipating philosophical debates on mathematical aspects of measurement from such unnecessary restrictions.

Acknowledgements I would like to thank three anonymous referees for encouraging comments and penetrating criticism of this paper's initial draft.

Declarations

Competing interests I declare that the research work connected with this paper has no competing interests other than the research funding whose details are included in this submission.

References

Adams, E. W. (1966). On the purpose and nature of measurement. *Synthese*, *16*, 125–169.

Adams, E. W. (1979). Measurement theory. P. D. Asquith & H. E. KyburgEds., *Current research in philosophy of science*(207–227).East Lansing.

Alper, T. (1987). A classification of all order-preserving homeomorphism groups of the reals that satisfy finite uniqueness. *Journal of Mathematical Psychology*, *31*, 135–154.

Baccelli, J. (2020). Beyond the metrological viewpoint. *Studies in the History and Philosophy of Science*, *80*, 56–61.

Cameron, P. J. (1989). Groups of order-automorphisms of the rationals with prescribed scale type. *Journal of Mathematical Psychology*, *33*, 163–171.

Churcher, B. G. (1935). Noise measurements for engineering purposes. *Transactions of the American Institute of Electrical Engineers*, *54*, 55–65.

Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluations of axioms fundamental to stevens ratio-scaling approach: I. loudness production. *Perception and Psychophysics*, *62*, 1505–1511.

[18] It may be useful to note that measurement theorists working on the representational approach have not failed to indicate its limitations or unsatisfactory aspects (Luce and Narens (1994)) or to explore alternative, non-representational frameworks (Narens, 2002b).

Ellermeier, W., Narens, L., & Dielmann, B. Perceptual ratios, differences, and the underlying scale. In B. Berglund & E. Borg (eds) *Fechner Day 2003. Proceedings of the 19th annual meeting of the International Society for Psychophysics*. Stockholm: International Society for Psychophysics, pp. 71–76. (2003)
Falmagne, J. C., & Narens, L. (1983). Scales and meaningfulness of quantitative laws. *Synthese*, *55*, 287–325.
Heilmann, C. (2015). A new interpretation of the representational theory of measurement. *Philosophy of Science*, *82*, 787–797.
Kim, S.-R. (1990). On the possible scientific laws. *Mathematical Social Science*, *20*, 19–36.
Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement volume i: additive and polynomial representations*.Wiley.
Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement volume III: representation, axiomatization and invariance*.Wiley.
Luce, D. R. (1959). On the possible psychophysical laws. *Psychological Review*, *66*, 81–95.
Luce, D. R. (1964). A generalization of a theorem of dimensional analysis. *Journal of Mathematical Psychology*, *1*, 278–284.
Luce, R. D., & Narens, L. (1994). Fifteen problems concerning the representational theory of measurement. P. HumphreysEd., *Patrick suppes: scientific philosopher, vol. 2: philosophy of physics, theory structure, measurement theory, philosophy of language, and logic*(219–245).Kluwer Academic Publishers.
Narens, L. (2002b). *Theories of meaningfulness*.Lawrence Erlbaum Associates
Narens, L. (2002a). A meaningful justification for the representational theory of measurement. *Journal of Mathematical Psychology*, *46*, 746–768.
Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109–129.
Narens, L. (1981b). A general theory of ratio-scalability, with remarks on the measurement-theoretic concept of meaningfulness. *Theory and Decision*, *13*, 1–70.
Narens, L. (1981a). On the scales of measurement. *Journal of Mathematical Psychology*, *24*, 249–275.
Niederée, R. (1987). On the reference to real numbers in fundamental measurement: A model-theoretic approach. E. Roskam & R. SucksEds., *Progress in mathematical Psychology*(3–23).North Holland.
Niederée, R. (1992). What do numbers measure? A new approach to fundamental measurement. *Mathematical Social Sciences*, *24*, 237–276.
Pfanzagl, J., Baumann, V., & Huber, H. (1973). *Theory of measurement* (2nd). Springer Veralg.
Philippi, C. L. (2021). On measurement scales: neither ordinal nor interval? *Philosophy of Science*, *88*, 929–939.
Rozeboom, W. W. (1962). “On the possible psychophysical laws”: Comment. *Psychological Review*, *69*, 552.
Stevens, S. S., & Davies, H. (1938). *Hearing: Its psychology and physiology*.Wiley
Stevens, S. S. (1946). On the theory of the scales of measurement. *Science*, *103*, 677–680.
Suppes, P., & Zinnes, J. (1963). Basic measurement theory. D. LuceEd., *Handbook of mathematical psychology*(1–76).Wiley.
Tal, E. (2021). Two myths of representational measurement. *Perspectives on Science*, *29*, 701–741.
Wolff, J. E. (2020). *The metaphysics of quantities*.Oxford University Press.
Yoshino, R. (1989). On the possible and stable psychophysical laws. *Journal of Mathematical Psychology*, *33*, 68–90.