Admissible statistics from a latent variable perspective

Zand Scholten, A.

Publication date
2011

Chapter 3

The Rasch Model as Additive Conjoint Measurement

To be sure, mathematics can be extended to any branch of knowledge [...] provided the concepts are so clearly defined as to permit accurate symbolic representation. That is only another way of saying that in some branches of discourse it is desirable to know what you are talking about.

James R. Newman

Abstract

The relation between Representational Measurement Theory (RMT) and Latent Variable Models (LVM) originating from psychometrics is investigated. RMT requires the empirical demonstration of adherence to axioms to ensure interval measurement. This deterministic theory, unable to accommodate error, seems incompatible with the imprecise measurement and fuzzy properties of psychology. Psychometrics, where messy variables are modeled using a probabilistic structure, fails to pose any direct, testable measurement level assumptions. These theories meet however in the Rasch model, a psychometric model considered by many to be a form of RMT. We investigate whether the Rasch model ensures quantitativeness by considering several problems associated with this claim. Although structurally equivalent, there are some valid reasons not to treat the models as interchangeable. These concern the interpretation of the concepts objects and empirical demonstration in RMT. The lack of continuity for many psychological properties, sequential item deletion and the use of probabilities as empirical objects is discussed. It is concluded that the problem in linking RMT and the Rasch model lies with the limited structure of psychological properties, not the statistical characteristics of the Rasch model itself.

3.1 Measurement Levels and Admissible Statistics

Measurement in the social sciences has always formed a controversial subject. Just after the turn of the nineteenth century, the status of measurement in the newly established field of psychology was immediately questioned by physicists such as Campbell (1940). According to the accepted view at that time, measurement requires that the property of interest has an additive structure that needs to be demonstrated through physical concatenation of objects. This is necessary since the goal of any measurement procedure is to assess the empirical validity of scientific laws, formulated in mathematical terms, involving numerical operations. The full use of such operations is warranted only if the empirical objects have the same additive structure as the numbers assigned to them (Campbell, 1920). A special committee (Ferguson et al., 1940), assembled for the express purpose of evaluating the feasibility of psychological measurement, was unable to settle whether any psychological property could be considered amenable to the requirements associated with this definition of measurement.

Stevens (1946, 1951) gave new impulse to the debate by turning the definition of measurement on its head. He proposed to let the numbers conform to whatever structure is present in the objects, instead of requiring that the empirical objects conform to the additive structure of numbers. For this purpose he redefined measurement as the assignment of numerals according to rule, and defined his famous levels of measurement – nominal, ordinal, interval, and ratio – as dependent solely on the question which rule was followed in assigning the numerals (e.g., determination of equality for nominal scales, determination of order for ordinal scales, etc.).

For each of these levels, different properties of the assigned numbers are relevant to the empirical structure. For example, if the empirical structure allows no more than the ordering of objects, then only the order of the assigned numbers is relevant. In this case the additive structure should be ignored and arithmetical operations that are based on the additive structure of numbers should be avoided. This restriction on numerical operations extends to a restriction on statistical operations. Statistical tests that make use of means for example, are therefore not permissible when measurements are only of the ordinal level.

Rarely explicit in Stevens' papers is the acknowledgement that some sort of similarity between the numerical assignments and the property measured should exist for his procedures to make sense. Without such similarity or isomorphism, however, his definition of measurement is completely vacuous, because the mere fact that numerals are assigned to rule is trivially satisfied in almost all cases, which makes it difficult to impossible not to be engaged in measurement. As a result of the fact that this requirement was left implicit, and because Stevens was vague on how to ascertain the structure of a property in the first place, it became common practice to just assume an interval level and be done with the measurement problem.

This should have changed with the advent of Representational Measurement Theory (RMT). With RMT (Krantz et al., 1971) the notion of representation of empirical structures through structure-preserving mappings into or onto numerical structures, which was already present in Stevens' theory, was fully formalized. In RMT the requirements to show that a property can be represented quantitatively (at the interval or ratio level) are specified in extenso. Interestingly enough however, RMT has had a less than spectacular impact on the social sciences. Applications of RMT concepts and actual testing of axioms has mostly remained reserved to psychophysics (Stevens, 1957; Luce, 1959a, 1959b; Iverson & Luce, 1998) and the field of subjective utility (Neumann & Morgenstern, 1953; Luce & Raiffa, 1957).

One of the reasons offered for this lack of interest from the broader field of psychology concerns the inaccessibility of the material due to its high level of abstraction and formalization (Cliff, 1992; Narens & Luce, 1993). In this context however, it is especially curious that RMT has failed to have a substantial impact on psychometrics, since this field is generally considered mathematically oriented. More importantly, psychometrics is concerned with formulating *measurement* models for psychological attitudes, abilities and personality traits.

One would expect a special interest in RMT from a field that seems to occupy itself with the same subject. So far however, RMT and psychometrics have only been linked successfully in one latent variable model, namely the Rasch model. This begs the question how RMT and latent variable models relate to each other and whether concepts from RMT can be incorporated in psychometric latent variable models. These questions will be addressed by first summarizing the key concepts and problems facing each of these approaches separately and then seeing how they come together in the Rasch model. Several conceptual problems raised by critics of the psychometric and the representational approach will be discussed.

3.2 Representational Measurement Theory

In RMT we start out by identifying a set of objects that can be compared on the property of interest. We subsequently specify one or more relations and possibly one or more operations that pertain to the objects. Together these elements form an empirical relational structure. This structure fully 'captures' the property of interest in terms of objects, the qualitative relations that hold between them and the operations that can be performed on them. The goal of measurement is to represent this empirical system consisting of qualitative relations and operations by finding a numerical system consisting of numerical relations and operations that has the same structure. Whether this can be done for a specific empirical structure is axiomatized in a representation theorem. How different but equally appropriate numerical representations are related to each other, i.e. what measurement level is associated with the empirical structure, is specified in a uniqueness theorem.

In modern day social sciences, at least in measurement theoretic circles, representational measurement theory is now the accepted view of measurement. The concept that numbers are convenient placeholders for more cumbersome qualitative relations is appealing. Social scientists have a greater insight into the empirical structures that might be available to them and are in principle able to rigorously test whether they have achieved a certain level of measurement. Also, the controversial concept of permissible statistics has been refined in terms of 'meaningfulness'. Where permissible statistics are concerned with invariance of test statistics, meaningfulness focuses on invariance of the truth value of a statistical statement with respect to the represented empirical structure, thereby transforming the problem into the legitimacy of the inference based on the statistic, not the legitimacy of the statistic itself.

3.2.1 Conceptual Issues in RMT

RMT carries with it some conceptual problems that render the representational approach problematic for psychology in general and for psychometrics in particular. Perhaps the most important issue concerns the deterministic nature of RMT and the associated lack of room for error. Representation theorems in RMT, whether they concern measurement at the nominal or ratio level, specify very strict axioms for the empirical relational structure. Psychological properties are often fuzzy in nature and difficult to assess through the haze of error that surrounds them. For such properties the requirements specified in these axioms are exceedingly likely to be violated. Transitivity, for example is a basic requirement of the weak order relation that figures prominently in many empirical structures. A relation $\succcurlyeq$ is transitive if for any objects a, b and c, if $a \succcurlyeq b$ and $b \succcurlyeq c$, then $a \succcurlyeq c$. In practice it is impossible to show transitivity for all objects, even for a property such as length. Due to the limited precision of our measurement instruments it is possible to find for example a set of rods where rod a extends at least as far as b, and b extends at least as far as c, for example if each pair seems to extend equally far, but where a does not extend at least as far as c (a structure known as a partial order).

In physics the lack of formal treatment of error in the theory is considered unproblematic, since the precision of measurement is generally very high. Critics of RMT have argued however that RMT is rendered useless when it comes to psychological measurement (Sijtsma, 2011). The structure of most psychological properties is little understood and any hope of filtering out systematic and random sources of noise to an acceptable degree seems far off. Under such circumstances one needs a way to deal with error. There have been some isolated efforts that employ probabilistic latent variable models to directly test RMT axioms (e.g., Karabatsos, 2001, 2005), but the application of these procedures seems limited for now. It has been suggested that RMT needs to be recast into the mapping of qualitative relations to distributions of random variables, instead of the random variables themselves, as is presently the case (Narens & Luce, 1993). What such a new approach to RMT will look like remains unclear however.

Another conceptual problem somewhat related to the deterministic nature of RMT is its heavy dependence on infinite sets. Most useful measurement structures describe homomorphisms (structure-preserving mappings) of an infinite set of empirical objects to the also infinite real numbers. Given that the number of objects accessible to us is finite, this can pose a problem, depending on how strict one interprets the axioms of RMT. Batitsky (1998) advances the use of infinite structures as an argument against the empiricist view present in RMT. RMT is appealing in part because it is presented as if observation is the only epistemological link necessary between our theories about the world and the world itself. RMT is not commensurable with this view however and requires a realist view to be internally consistent.

Axioms concerning infinite sets for example are principally untestable through direct observation. Abstract representation axioms require the physically impossible comparison of an object with itself: a $\succcurlyeq$ a. Batitsky argues that from a strict empiricists' view, one cannot solve this problem without resorting to replication or duplication of an object which is either strictly impossible or changes the object so fundamentally, that the comparison is rendered meaningless. Another problem lies in the use of natural numbers in the Archimedean axiom, included in all representation theorems concerning infinite sets of objects. Put very generally, the Archimedean condition requires that no ratio between any object is infinitely small or large. In the additive case this means that for any object a no matter how large, and any object b no matter how small, there exists a positive integer n so that object b concatenated with itself n times is larger that object a: $nb \succcurlyeq a$. These natural numbers have no directly observable empirical counterpart however.

If violations of axioms are interpreted as error, as occurs with the intransitivity of length when the measurement instrument becomes exceedingly precise, then we accept something that is measurement-independent about the property. We consider a true and a measured value of the property, otherwise we would have to reject quantitative measurement of length. A minimal amount of realism is needed: the cognitive step necessary to accept that a numerical system represents an empirical system (since we know perfect adherence or perfect empirical demonstration is impossible) requires a realist position, "even when construed as involving no more than the belief in [the] theory's agreement with all measurable results"(Batitsky, 1998, p. 68). This shows that RMT is incompatible with strict empiricism. Although this argument seems sound, it does not disqualify RMT as valid theory of measurement if we are prepared to accept at least a minimal level of realism into our philosophical outlook on science.

A final problem that is relevant in this context is the lack of formal maturity of the concept of meaningfulness. Although several attempts have been made to fully formalize the concept in RMT and set-theoretic terms, this task remains unfinished (Robinson, 1965; Roberts, 1986; Marcus-Roberts & Roberts, 1987). For now the formal property of meaningfulness can be shown to be a necessary but not a sufficient condition for empirical significance. Meaningless statements that do not adhere to the formalized requirements of meaningfulness can be eliminated as empirically insignificant, but unfortunately we cannot conclude that a statement that does adhere to the formal requirements is guaranteed to be empirically significant.

3.2.2 Classical Theory of Measurement

It is important to note that although RMT is the generally accepted view on measurement in measurement theoretic and psychometric circles (Hand, 2004; Embretson & Reise, 2000; Bond & Fox, 2007), it is not the only view. According to Michell (1986, 1993, 1994, 1997) we should return to a more classical definition of measurement that goes back to Euclid. This definition asserts that measurement is equivalent to the establishment of ratios between quantities of some property. Quantitative properties are not constructed representations as in RMT, but quantitative structures that exists independently in the world, awaiting our discovery of them. The ratios between quantities, for example between some chosen unit and an object of interest, are identifiable with the real numbers. In RMT numbers exist aside from physical reality, they form a representation of reality that is assigned or constructed at an abstract level. In classical measurement theory the real numbers *are* in fact the relations between magnitudes and therefore the product of measurement. Michell maintains that the classical theory is preferable to RMT since it does not entail a vague ontological status of numbers and since it is not based on a set-theoretic foundation, which is associated with formal inconsistencies as pointed out in Gödel's incompleteness theorems (Michell, 1999).

In this classical theory of measurement the word quantification and measurement refer to the same thing. Classification and ordering – or nominal and ordinal measurement in RMT – are not considered measurement. The two theories concur however in the method in which quantitative structure needs to be demonstrated. Adherence to the axioms of an interval or ratio level RMT measurement structure will be an acceptable form of support for quantitativeness for a classical measurement theorist. The relevant axioms of RMT agree with the classical approach first axiomatized by Hölder (Hölder & Michell, 1997). We will therefore take a pragmatic stance, and consider what role axiomatic measurement theory in general, whether it be classical or representational, can play in psychometric models, given the difficulties associated with direct axioms testing, error and continuity.

3.3 Assumptions of Latent Variable Models

To loosely paraphrase Luce (1996), RMT essentially deals with the structure of measurable properties by way of axioms concerning empirical relations, whereas psychometrics deals with variability in measurable properties using the statistical random variable approach. In psychometric measurement models one or more latent variables are postulated that are assumed to cause or explain responses to items or problems. A Latent Variable Model (LVM) is data-oriented, aimed at providing a more parsimonious description of the data, possibly with more appealing characteristics than the raw scores. This does not necessarily mean however, that a LVM does not have anything to say on the structure of measurable properties. One could say that there are some structural assumptions hidden in the statistical assumptions that form the basis of at least some LVMs. We will consider a specific subset of LVMs that are characterized by the following assumptions concerning a set of item responses, denoted by a vector of random variables X and one latent variable denoted by θ:

- A probability distribution is specified for X.
- θ is unidimensional.
- θ is continuous.
- Local independence: X_i is independent, conditioned on θ.
- A link function is specified that relates the probability of each item response to θ.
- Monotonicity: the link function is non-decreasing.

Local independence means that the item responses are independent once the effect of the latent variable on the responses is accounted for. Monotonicity means that a higher value on the latent ability or trait will result in a higher probability of a correct response or endorsement of the item. These assumptions were distilled from several sources (Hambleton, Swaminathan, & Rogers, 1991; Embretson & Reise, 2000; Mellenbergh, 1994; Ellis & Junker, 1997). Most specify unidimensionality, local independence and monotonicity, some subsume unidimensionality under local independence, since the first follows from the second when there is one latent variable. Specification of a link function is often pooled together with monotonicity. Other types of LVMs are achieved by considering multiple or multidimensional latent variables, non-monotonic link functions or categorical latent variables.

It is important to note that this very general set of assumptions applies to Item Response Theory (IRT) Models that assume a nonlinear relation between the latent and observed level, but also to the linear model used in Factor Analysis. In any case these assumptions seem orthogonal to the axioms of RMT. Nothing is explicitly specified about the internal structure of the property or the measurement level associated with the observed or latent variables. Michell (1986, 1990, 1997, 2008b, 2009) points out that psychometricians structurally assume a continuous latent variable has an interval structure without ever testing this hypothesis. This reproach certainly has merit. Direct testing of these assumptions is not part and parcel of the general procedure followed when these models are applied to data. However, it is unfair to say that no attention has been given to this issue whatsoever. Psychometrics has been involved with the evaluation of assumptions that have an indirect, but nevertheless important bearing on the measurement pretensions of LVMs.

For example, Ellis and Junker (1997) examined what conditions the observed variables need to satisfy in order to allow a latent variable representation. It was shown that observed responses X satisfy a LVM model if and only if θ is tail-measurable, i.e. two extra conditions of conditional association and vanishing conditional dependence hold. Without going into details of these conditions, tail-measurability means that the latent variable can be estimated consistently from the observed responses even though a subset of the responses is removed. Their finding is also associated with a uniqueness statement, namely that such a LVM is unique up to a strictly increasing monotonic function, resulting in an ordinal scale.

Other examples of research into the structure of psychological properties are studies that attempt to ascertain the dimensionality of variables using confirmatory factor analysis or research where the comparative fit of latent class and latent trait models is assessed to determine whether a property should be viewed as a categorical or continuous variable (Quinlan, Maas, Jansen, Booij, & Rendell, 2007). Still another example concerns the investigation of ergodicity, or rather lack thereof, in psychological properties. A measurement instrument that is supposed to assess variation between persons on some quantitative property can only do so if the property has the same structure for all persons. When the intra-individual structure of a property is not consistent with the inter-individual structure, but a measurement instrument is based on the latter, attempts at quantification seem pointless. Differences between inter- and intra-individual structure are examined by comparing standard factor analysis based on group data with repeated measures of one person (Hamaker, Dolan, & Molenaar, 2005). As for direct testing of axioms however, there seems to be only one LVM that can be directly linked to RMT.

3.4 Rasch and Additive Conjoint Measurement

Largely in parallel with the development of RMT, several latent variable models were developed in psychometrics during the 1950s and 1960s, which have been used extensively in empirical situations: for instance, such models are routinely used in large-scale educational testing. It was soon apparent that one of these models, namely the Rasch model (Rasch, 1960), bears a striking resemblance to structures developed in RMT. In a Rasch model, a monotone transformation of the dependent variable (the item response probability) is an additive function of two independent variables, namely person ability and item difficulty. This is precisely the way that Additive Conjoint Measurement (ACM) pictures the situation (Luce & Tukey, 1964). Because the model is structurally equivalent to ACM, but also incorporates probabilities – so that we may deal with measurement imprecision in our model – it appears to give us the best of two worlds. Some have enthusiastically concluded that, with the Rasch model, psychologists are capable of true interval level measurement, comparable to that in physics (Perline, Wright, & Wainer, 1979; Embretson & Reise, 2000; Bond & Fox, 2007). Bond and Fox (2007) go so far as to decorate their book on Rasch models with the subtitle 'fundamental measurement in the social sciences'. It is suggested in such works that fitting a Rasch model to data automatically gives one an interval representation of person ability and item difficulty.

The Rasch model that ascribes observed item responses to the difference between person ability and item difficulty, was developed with the express purpose of obtaining interval level measurement in mind. Rasch's intention was to specify a model that allows for the determination of person abilities that do not depend on the specific items that were used for this determination and reversely, that allows for the determination of item difficulties that do not depend on the persons that were used. He termed this property 'specific-objectivity'. This property holds when the probability of answering an item correctly is given by the logistic function of the difference between person ability and item difficulty – any other function such as the normal ogive will fail to produce this property. Models using other functions or more parameters also fail to provide the raw sum score as a sufficient statistic for latent ability or item difficulty. The Rasch model is not only structurally different from other IRT models, it was also developed to be used very differently. Other IRT models are used to describe and fit data; when fit is poor, the model is adapted or discarded in favor of another model. In contrast, the Rasch model is more prescriptive. The data are required to fit the model and when they do not, items that show misfit are discarded until a satisfactory fit is obtained.

When data are generated by, or perfectly fit the Rasch model, the estimated abilities, item difficulties and probabilities of answering an item correctly will meet the axioms of ACM. The three most important axioms state that the item difficulties and person abilities , and the result of comparing these – i.e. the probabilities – should be weakly ordered; that the pairing of ability and difficulty should show single cancellation, also referred to as independence, which corresponds to specific-objectivity; and that the probabilities should also conform to double cancellation, which in very broad terms means that a simultaneous increase (or decrease) in ability combined with an increase (or decrease) in difficulty should result in a consistent change in the probability, always in the same direction no matter which ability or difficulty one starts out with.

3.4.1 Problems with the Rasch Model as ACM

Although data generated or perfectly described by a Rasch model result in adherence to the axioms of ACM, the conclusion that the underlying property is therefore measured on an interval level seems premature. There are several reasons why this claim is contentious. The Rasch model might bear an uncanny structural resemblance to the ACM structure of RMT, but its implementation does not sit well with some more fundamental elements of RMT. In RMT one is required to specify in advance what set of objects the axioms will apply to. For physical properties such as length, it is relatively easy to specify what sort of objects this set consists of (e.g. rigid rods of a certain material). Even for seemingly 'simple' psychological properties such as difficulty of items that assess arithmetic or spatial ability this is already much harder.

Also, the process of sequential deletion to obtain adequate model fit appears inconsistent with RMT when items initially seem perfectly unbiased indicators of an ability, and no substantive reason for misfit can be identified. Suppose we were to construct a measurement instrument for length using paired comparison of people with wooden rods. If we discarded the rods that do not conform to our additivity axioms, we would be at serious risk of producing a measurement instrument that produces interval level measurement for only a subset of the objects that the axioms should hold for. Now if we could disqualify these wayward rods by pointing out that they were all rods made of freshly cut wood that was still highly flexible, we could legitimately adjust our demarcation of the set of objects. However, for psychological properties such an identification is often impossible, or an attempt to do so is not even made. When the focus is only on claiming interval level measurement for some property, and not on identifying factors that may confound the variable of interest, the use of the Rasch model can give us a false sense of accomplishment and can result in measurement instruments that falsely claim to represent some quantitative property.

Another problem in aligning the Rasch model with RMT is that quantitative structures, such as ACM, are infinite structures, thus one should always be able to supply an object that lies between two other objects with ever-increasing precision. Now although the number of wooden, sufficiently rigid rods on this planet is finite, we can imagine shaving a minute slice off the end of some rod to obtain the required object. When we ignore constraints of a pragmatic nature for a moment, it is easy to imagine obtaining a rod of any length. For psychological properties this is often intrinsically impossible. For example, what arithmetic item lies between the problems '1 + 1' and '1 + 2' in terms of difficulty? For the property of length it is immediately obvious how we obtain such an object, since we know the structure of the property. For most psychological properties this knowledge still eludes us.

3.4.2 Empirical Status of Probabilities in the Rasch Model

The empirical status of probabilities, that function as objects in the conjoint empirical relational system, also pose a source of contention concerning the claim that the Rasch model is an instantiation of ACM. Kyngdon (2008a) argues that there is no basis for this claim. He states that, from the RMT perspective, in order to call something conjoint measurement one has to prove that a mapping of an empirical relational structure into a numerical structure, with the right uniqueness properties, has been achieved. In RMT, this is done by checking whether certain axioms hold in the data and constructing the numerical representation from the ground up. In latent variable modeling in general, and Rasch modeling in particular, this is never done. Hence, Rasch enthusiasts are leaving something out; and, according to Kyngdon, this something is not just important, but essential.

To counter potential confusion in the reader, it is useful to note that there is no discussion about whether the properties of the mathematical structure as presumed in a Rasch model, i.e., about the way the parameters combine into item response probabilities, are similar to those of additive conjoint measurement; at the level of mathematical representation, the models are clearly equivalent (see, e.g., Borsboom, 2005). Kyngdon's argument is rather that the system of probabilities, on which the Rasch model operates and which is supposed to play the role of empirical relational structure, does not qualify as a bona fide candidate for such a structure from the perspective of RMT. Further, Kyngdon argues that merely fitting a Rasch model to a dataset is insufficient as a basis for claiming that fundamental measurement has been achieved, because in the process of model fitting as ordinarily executed, the relevant axioms are not explicitly tested. These issues are addressed in turn.

Kyngdon argues that the empirical relational system presumed in the Rasch model, and in latent variable models in general, does not qualify as an empirical relational structure from the RMT perspective. Kyngdon's point here is not as clear as possible, partly due to statements involving the claim that probabilities are Platonic realm pure sets of numbers, which are likely to confuse many readers. The point here is not really what probabilities are; the point is rather what probabilities are not. Probabilities, as Kyngdon thinks RMT must interpret them, are not themselves observable empirical entities and hence cannot function as the proper basis for an empirical relational system. More specifically, according to Kyngdon, RMT specifies that objects on which the empirical relational structure is defined should be 'spatio-temporally located', and this, he argues, is not a requirement that probabilities could meet. Kyngdon thinks that RMT must interpret probabilities as non-empirical Platonic realm stuff, i.e., as mathematical rather than empirical entities. And because the Rasch model decomposes a structure of probabilities (mathematical entities) into a set of person and item parameters (also mathematical entities), it really relates one numerical system to another, instead of relating an empirical relational system to a numerical one. Hence it fails to satisfy the demands of RMT. This reasoning is not foolproof however.

3.4.3 The Spatio-Temporal Location of Probabilities

Kyngdon's argument would seem to hinge on whether RMT does or does not inherently assume that probabilities are numerical, abstract, Platonic realm entities. Probabilities are certainly highly troublesome entities in any philosophical analysis, but there do exist interpretations of probability in which probabilities are considered to be real empirical entities: most notably, the propensity interpretation of probability (e.g., Hacking, 1965). In this interpretation, probabilities are real entities that could, in principle, form the basis for the construction of an empirical relational system—if only there were some way of noticing relations between them. And of course standard model-fitting techniques assume there is such a way, namely estimation. If one buys into this line of reasoning, and endows probabilities with a good deal of realism, there would appear not to be a major problem in the fact that Rasch models, and latent variable models in general, work on probabilistic structures.

Does RMT necessarily deny this possibility? Who the authoritative source is supposed to be remains unclear, but at least in the volumes on the Foundations of Measurement (Krantz et al., 1971), we do not find such claims. Philosophically speaking, these works are not very explicit at all. We are left with the rather vague notion that relations between objects should be noticeable, with no specification on who should do the noticing and how. This vagueness leaves open the possibility that different RMT protagonists may have very different ideas about this issue. In the work of Scheiblechner (1999), for instance, stochastic dominance relations between persons and items are defined in terms of probabilities. If one could notice these relations and use them to build up an empirical relational system, for instance through appropriate estimation techniques, would there necessarily be a problem? This is doubtful, especially given the fact that latent variable models are not used to model probabilities as such, or to map the real numbers into themselves, as Kyngdon appears to think. At least in the model formulation, probabilities are used instrumentally, namely to order persons and items, not real numbers, and it is this structure on which the measurement model is supposed to operate.

In all likelihood, Kyngdon would grant us the fact that persons and items are bona fide spatio-temporally located empirical objects, and also assume that he would acknowledge that these are the things that the empirical relational structure is supposed to be about. Even though the mathematical model consists of estimates and probabilities, which may or may not be interpreted as abstract Platonic entities, an applied researcher immediately superimposes spatio-temporally located, empirical counterparts onto these estimates and probabilities, namely persons and items. Certainly, we cannot stub our toe on a probability; we need to estimate it. But of course that is why we use statistical models. We assume the model describes a structure actually present in the world, and the statistical machinery gives us parameter estimates that are supposed to capture this structure, or at least approximate it. Naturally, it is hard to tell whether we have succeeded in doing this, but that is not significant here; the question is whether this is possible in principle. This reduces the issue to the question whether RMT should allow empirical relational systems to contain relations such as stochastic dominance or not, and whether model-fitting approaches could in principle uncover such relations.

3.4.4 Strict vs. Weak Representationalism

Let us divide the representational world into two sorts of people according to the answer they would give to this question: strict representationalists and weak representationalists. For the strict representationalist, noticeable means directly observable. For this representationalist, it is out of the question that any probabilistic model could be conjoint measurement, since probabilistic relations are not directly observable (whether they live in the Platonic realm or somewhere else is not important in this respect). For the weak representationalist, however, noticeable may mean noticeable in principle, or noticeable for an ideal observer, or perhaps even true. Such a weak representationalist may perhaps suppose that persons and items are actually ordered by probabilities, and capture this hypothesis in a latent variable model. It would seem that fitting a Rasch model, in this case, could lead to a bona fide measurement structure, provided that the generating model is properly specified and the statistical machinery works out as it should. This may not be in the spirit of RMT, and the position may lead to serious conceptual difficulties (Borsboom, 2005); but that would not clinch the argument here. The question is whether such weak representationalists would encounter insurmountable difficulties in doing empirical research and building measurements models, or whether we could point to a genuine inconsistency in their reasoning. This is doubtful. Given the serious difficulties of RMT to accommodate such a mundane thing as measurement error, if there are any representationalists among those involved in actual research, they should better be weak representationalists in order to get anything done at all.

Even from such a weak point of view, however, there is absolutely no guarantee of success, and Kyngdon is certainly right to point this out. Fitting statistical models is wrought with difficulties; especially troublesome is ascertaining whether the model fits the data or not. In this respect we should be very careful. When we want to draw strong conclusions, such as that we have achieved 'fundamental measurement', based on a model-fitting exercise alone, there is bound to be trouble. In such a case, one may conclude that one has conjoint measurement, based on a fitting Rasch model, when in fact there is no additive conjoint structure present. This would occur when a Rasch model fits the data adequately, but at the same time the relevant axioms of the additive conjoint model are violated. Kyngdon supposes that this may regularly happen, referring to Wood (1978), who demonstrates this by imagining that we let people toss a coin several times and mark the heads or tails as if they were responses to items. However, although such data would not reject a Rasch model, the experiment should result in flat item response functions, which give constant probabilities at 0.5; as a result, each response pattern would be equally likely. Presumably, this would be noticed in the analysis. The example does show that the failure to reject a model does not guarantee that the axioms of conjoint measurement are supported; however, it is not entirely clear what to make of this given that the example is rather contrived (one also wonders what an axiomatic approach would lead to in this case). Moreover, it would seem to us that serious violations of the Rasch model, such as intersecting item response functions, will be picked up by decent statistical methods (e.g., likelihood ratio tests) if power is sufficiently high.

3.5 Conclusion

If the Rasch model is 'true', axioms imposed from additive conjoint measurement, for instance as used in Karabatsos' (2001) approach, will ordinarily be satisfied. However, it is an open question how sensitive different goodness of fit statistics are to deviances from the model structure that would lead to the violation of such axioms. For this reason, researchers who aim to use the Rasch model and interpret it in terms of fundamental measurement may do well to use direct axiom checking approaches such as those developed by Karabatsos (2001), as Kyngdon suggests. Given the general difficulty of ascertaining the fit of latent variable models, such methods should certainly be hailed as important additions to the statistical methods currently in place.

However, since the empirical checks that such methods perform are implied by the Rasch model whether one thinks it is an RMT structure or not, these methods can be used regardless of one's philosophical inclinations on the relations between Rasch models and additive conjoint measurement. Whether there is a difference in effectiveness of axiom checking versus conventional model fitting is an interesting question of some importance, but that importance is not of a theoretical nature. Perhaps under Karabatsos' (2001) approach the Rasch model will be rejected more often, and perhaps not. Whatever turns out to be the case, it will have no relevance for the conceptual question whether the Rasch model could instantiate additive conjoint measurement or not.

The fundamental question of whether the Rasch model may or may not be interpreted as a probabilistic species of RMT does not depend on which model-checking approach is best. It depends on how far one is prepared to stretch one's interpretation of both Rasch models and RMT. Kyngdon shows that one has to stretch quite far; the present discussion has demonstrated this more or less in vivo, by executing some of the theoretical manoeuvres necessary for connecting Rasch models to RMT. Hopefully, this will caution researchers to consider Rasch models and RMT to be exchangeable terms; they certainly are not. At least, not without considerable stretching.

The question of the empirical status of probabilities put forth by Kyngdon (2008a) is one of two serious objections raised against the claim of interval level measurement attached to the Rasch model (see the next chapter for a discussion of the second objection). The problematic status of probabilities seems overshadowed however by the problem of the intrinsically discrete nature of many psychological properties, the lack of clear definition or demarcation of the set of objects that display these properties and the risk associated with sequential item deletion in the Rasch model. The problem in linking RMT and the Rasch model lies more with the limited structure of psychological properties than with the statistical characteristics of the Rasch model. These problems should receive more attention: A critical look should be taken, not at the statistical model itself, but at the method with which it is applied and at the properties it is applied to.