

# Assessing the local identifiability of probabilistic knowledge structures

Luca Stefanutti · Jürgen Heller · Pasquale Anselmi · Egidio Robusto

Published online: 17 May 2012  
© Psychonomic Society, Inc. 2012

**Abstract** Given a collection  $Q$  of problems, in knowledge space theory Doignon & Falmagne, (International Journal of Man–Machine Studies 23:175–196, 1985) the knowledge state of a student is the collection  $K \subseteq Q$  of all problems that this student is capable of solving. A knowledge structure is a pair  $(Q, \mathcal{K})$ , where  $\mathcal{K}$  is a collection of knowledge states that contains at least the empty set and  $Q$ . A probabilistic knowledge structure (PKS) is a knowledge structure  $(Q, \mathcal{K}, \pi)$ , where  $\pi$  is a probability distribution on the knowledge states. The PKS that has received the most attention is the *basic local independence model* BLIM; Falmagne & Doignon, (British Journal of Mathematical and Statistical Psychology 41:1–23, 1988a, Journal of Mathematical Psychology 32:232–258, 1988b). To the best of our knowledge, systematic investigations in the literature concerning the identifiability of the BLIM are totally missing. Based on the theoretical work of Bamber and van Santen (Journal of Mathematical Psychology 29:443–473, 1985), the present article is aimed to present a method and a corresponding computerized procedure for assessing the local identifiability of the BLIM, which is applicable to any finite knowledge structure of moderate size.

**Keywords** Local identifiability · Probabilistic knowledge structures · Knowledge space theory · Basic local independence model

---

L. Stefanutti (✉) · P. Anselmi · E. Robusto  
Dipartimento di Psicologia Applicata, University of Padua,  
Via Venezia 8,  
35133 Padova, Italy  
e-mail: luca.stefanutti@unipd.it

J. Heller  
Fachbereich Psychologie, Eberhard-Karls-Universität Tübingen,  
Schleichstr. 4,  
72076 Tübingen, Germany

## Introduction

In the theory of knowledge structures (Doignon & Falmagne, 1985, 1999; Falmagne & Doignon, 2011), a *domain of knowledge* is a collection  $Q$  of problems or items (e.g., all problems that can be formulated in linear algebra), and the *state of knowledge* of an individual is the particular subset  $K \subseteq Q$  of problems that he or she is capable of solving. A *knowledge structure* is a collection of knowledge states, and in practical applications of the theory it represents all different knowledge states existing in a given population.

Knowledge structures are deterministic models of the organization of knowledge, and as such, are used to represent specific dependencies among the problems in  $Q$ . In the basic case, such dependencies are expressed in terms of a precedence relation on the collection  $Q$  of problems, called a *surmise relation* and interpreted as “failing  $p$  implies failing  $q$ ,” where  $p$  and  $q$  are two problems in  $Q$ . Knowledge structures corresponding to surmise relations are called *quasi-ordinal knowledge spaces*.

Introducing dependencies among the problems provides a way to decide whether or not an arbitrary subset of  $Q$  is a knowledge state. For instance, if the dependence is described by a surmise relation saying that “failing  $p$  implies failing  $q$ ,” then the existence of a knowledge state containing  $q$  and not containing  $p$  is excluded. Thus, a knowledge structure usually forms a strict subset of the power set on  $Q$ .

Individual assessment based on knowledge structures aims at recovering the knowledge state of a subject. Suppose that the problems in  $Q$  have been administered to a student who provided a response to each of them. The collection  $R$  of all problems in  $Q$  that were correctly answered by the student is called a *response pattern*. In an ideal situation, a student whose knowledge state is  $K$  provides a correct answer to an item  $q$  if and only if  $q$  is in  $K$ . This reflects

the idea that  $K$  is really the collection of what the student is capable of solving. However, this ideal situation is not very realistic. In some cases, the student might be able to guess the answer to a problem without knowing the correct solution process (this is the case, e.g., with multiple-choice problems, but not only with those problems). On the other hand, sometimes failure of a problem is a consequence of making a careless error or being inattentive. Given this variability, there is some chance that the response pattern provided by the student will differ to some extent from the underlying knowledge state. All of these effects must be taken into account when a knowledge structure is applied to empirical data.

For this reason, an appropriate probabilistic setting has been provided by Falmagne and Doignon (1988a, b). Probabilistic knowledge structures (PKS), which are the subject matter of the next section, are knowledge structures equipped with a probability distribution on the knowledge states. In a PKS, knowledge states are regarded as discrete latent classes that govern the response behavior of a student (Schrepp, 2005; Ünlü, 2011). At present a special case of this latent class model is the subject of our attention; it is called the *basic local independence model* (BLIM, for short). The BLIM imposes restrictions on the form of the conditional probabilities  $P(R | K)$  of the response patterns, given the knowledge states. These probabilities essentially depend on two parameters per item: a lucky guess probability  $\eta$  and a careless error probability  $\beta$ . The lucky guess parameter  $\eta_q$  of a problem  $q$  is regarded as the conditional probability that a correct answer is provided for  $q$ , given that  $q$  is not in  $K$ , whereas the careless error parameter  $\beta_q$  is interpreted as the conditional probability that a wrong answer is provided for  $q$ , given that  $q$  is in  $K$ .

A number of questions concerning the applicability of this model arise. Some of them, like parameter estimation, have already been answered, while other questions are still open. The parameters of the BLIM can be estimated by maximum likelihood via the expectation-maximization algorithm (see, e.g., Stefanutti & Robusto, 2009). Other estimation methods have also been explored by Heller and Wickelmaier (2011). The model fit can be tested by standard statistics, such as Pearson's chi-square or the likelihood ratio test.

The identifiability of the BLIM, which is the subject matter of the present article, is still an open question. Roughly speaking, a model is not identifiable if exactly the same predictions can be obtained starting from totally different parameter sets. Model identifiability is crucial to interpretation of the parameters. If different solutions exist for the parameters of the model, their values cannot be interpreted, and the model loses its explanatory power. This means that the model might still provide correct, or even excellent, predictions of the frequency distribution

over the response patterns, but the explanations that one obtains from an analysis of the parameter values can be totally misleading.

At the present time, not much is known about the identifiability of the BLIM. For instance, we don't know whether specific types of knowledge structures exist for which the model is identifiable, or whether other types exist for which it is not. In our view, the problem can be tackled by following two parallel routes. The first route is a purely analytic treatment, in which identifiability is systematically studied and understood on a formal basis. This requires spotting all of the different sources of nonidentifiability in the BLIM, which might be many. Much work remains to be done in this direction.

The other route, which is the one taken in this article, makes use of some analytic tools for testing the identifiability of the BLIM with respect to specific knowledge structures, either in concrete applications or in simulation studies. The proposed procedure is based on theoretical results by Bamber and van Santen (1985, 2000), applying them to the case of the BLIM.

The article is organized as follows: After a brief review of probabilistic knowledge structures and the BLIM model, the problem of assessing the local identifiability of the BLIM is introduced. A method based on analysis of the Jacobian matrix of the model's prediction function is then described. This method allows for detecting the parameters that are involved in specific tradeoffs, leading to nonidentifiability. In a subsequent section, the main tradeoff dimensions of the BLIM parameters are introduced and discussed. A theoretical result in this section shows that tradeoffs cannot occur among knowledge state probabilities. This part is followed by a section that deals with the problem of constructing the Jacobian matrix of the BLIM's prediction function in practice. This problem is not trivial, since such a matrix contains as many rows as the number of theoretically observable response patterns, which might be huge, even for knowledge domains of moderate size. Finally, two remaining sections describe a computerized MATLAB procedure for assessing the identifiability of the BLIM for arbitrary knowledge structures and parameter sets, provided that the set  $Q$  of items is of a moderate size. In principle, the procedure is also applicable to large-size item sets, but in that case the computational costs might become prohibitive.

### Probabilistic knowledge structures

A *knowledge structure* is defined as a pair  $(Q, \mathcal{K})$  in which  $Q$  is a nonempty set (assumed to be finite throughout this study), and  $\mathcal{K}$  is a family of subsets of  $Q$  containing at least  $Q$  and the empty set  $\emptyset$  (Doignon & Falmagne, 1985). While the set  $Q$  is called the *domain* of the knowledge structure, its elements are

referred to as *items* (or *problems*). The subsets in the family  $\mathcal{K}$ , which are called (*knowledge*) *states*, represent those subsets of items from the considered domain that an individual masters. In general, however, we may not assume that a person solves a problem if and only if the person masters it. In the case of a careless error, the person actually masters an item but does not solve it. In the case of a lucky guess, an item is solved without being actually mastered. These types of errors are handled within a probabilistic framework, which is based on dissociating the knowledge state  $K$  of a person from the actual given response pattern  $R$ . Let  $\mathcal{R} = 2^Q$  denote the set of all possible response patterns on the domain  $Q$ . Falmagne and Doignon (1988a, b) and Doignon and Falmagne (1999) defined a *probabilistic knowledge structure*  $(Q, \mathcal{K}, P)$  by specifying a (marginal) distribution  $P_{\mathcal{K}}$  on the states of  $\mathcal{K}$  and the conditional probabilities  $P(R | K)$  for all  $R \in \mathcal{R}$  and  $K \in \mathcal{K}$ . The marginal distribution  $P_{\mathcal{R}}$  on  $\mathcal{R}$  then is predicted by

$$P_{\mathcal{R}}(R) = \sum_{K \in \mathcal{K}} P(R|K) \cdot P_{\mathcal{K}}(K). \tag{1}$$

The probabilistic knowledge structure that has received the most attention is the BLIM, which satisfies the following condition. For each  $q \in Q$ , local independence assumes that there are real constants  $0 \leq \beta_q < 1$  and  $0 \leq \eta_q < 1$  such that for all  $R \in \mathcal{R}$  and  $K \in \mathcal{K}$ ,

$$P(R|K) = \left( \prod_{q \in K \setminus R} \beta_q \right) \cdot \left( \prod_{q \in K \cap R} (1 - \beta_q) \right) \cdot \left( \prod_{q \in K \setminus R} \eta_q \right) \cdot \left( \prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right) \tag{2}$$

### Assessing the local identifiability of the BLIM

After briefly introducing some general concepts related to identifiability in probabilistic models, in this section we will consider how these concepts may be applied to the basic local independence model. For a general framework to treat identifiability, we mainly refer to the work of Bamber and van Santen (1985, 2000). Within this framework, a model is regarded as a triple  $(D, f, O)$ , where  $D \subseteq \mathfrak{X}^m$  is called the model’s *parameter domain*,  $O \subseteq \mathfrak{X}^n$  is the model’s *outcome space*, and  $f: D \rightarrow O$  is the so-called *prediction function* of the model. The model’s prediction  $f(\theta)$  for a given parameter vector  $\theta \in D$  provides an outcome in  $O$ .

Formally, a model  $(D, f, O)$  is *identifiable* if its prediction function  $f$  is one-to-one, and it is *locally identifiable* at a given point  $\theta_0 \in D$  if  $f$  is one-to-one when restricted to points within some distance  $\varepsilon > 0$  from  $\theta_0$ . Bamber and van Santen (1985) showed that,

with  $m$  parameters in the model, if the maximum rank of the Jacobian matrix of the prediction function  $f$  of a model  $(D, f, O)$  is less than  $m$  everywhere in  $D$ , then the model is not identifiable. Similarly, but in a somehow opposite direction, Smith (1998) showed that, if the Jacobian matrix of the prediction function  $f$ , evaluated at a certain point  $\theta_0$ , has rank equal to the number of parameters in the model, then the model is locally identifiable at that point. We recall that the Jacobian matrix of the prediction function  $f$  is the matrix  $\mathbf{J}$  of the first partial derivatives of  $f$  with respect to the model parameters.

Both results suggest a way to assess the local identifiability of a model: (1) construct the Jacobian matrix  $\mathbf{J}$  of the prediction function  $f$  at some point  $\theta_0 \in D$ , where the rank of  $\mathbf{J}$  is maximum; (2) compare the rank of  $\mathbf{J}$  with the number  $m$  of parameters; (3) conclude that the model is identifiable if  $\text{rank}(\mathbf{J})$  equals  $m$ , and nonidentifiable otherwise.

In the procedure described above, the requirement that a vector  $\theta_0$  must be selected where the rank of the Jacobian matrix is maximum may be considered problematic. In this direction, however, a theorem by Bamber and van Santen (1985, Theorem B5) guarantees that, under very general conditions, the Jacobian matrix has maximum rank almost everywhere in the parameter space  $D$  of the model. In this case, if a point in  $D$  is picked at random, then the Jacobian matrix, evaluated at that point, will have maximum rank with probability 1. To be precise, the conditions require that the prediction function  $f$  be an analytic function, and that  $D$  be a subset of a connected open set and have positive Lebesgue measure (which holds, e.g., if  $D$  contains a non-empty open set).

### The Jacobian matrix of the BLIM

First, we define the parameter space  $D$  for a given BLIM  $(Q, \mathcal{K}, P)$ , which is the set of all admissible parameter combinations. Let  $\beta = (\beta_q)_{q \in Q}$  and  $\eta = (\eta_q)_{q \in Q}$  denote the parameter vectors of the item-specific careless error and guessing probabilities, respectively, and let  $\pi = (\pi_K)_{K \in \mathcal{K}^*}$  with  $\mathcal{K}^* = \mathcal{K} \setminus \{Q\}$  denote the parameter vector of independent state probabilities  $\pi_K = P_{\mathcal{K}}(K)$ ,  $K \in \mathcal{K}^*$ . Without loss of generality, we may assume that all state probabilities are nonzero, because otherwise we can simply eliminate the respective state from the knowledge structure. This leads to the parameter restrictions  $\pi_K > 0$  for all  $K \in \mathcal{K}^*$  and

$$\sum_{L \in \mathcal{K}^*} \pi_L < 1. \tag{T1}$$

Moreover, the probabilities of response errors are assumed to be nonzero. This means that for all  $q \in Q$  both the  $\beta_q$  and  $\eta_q$  are elements of the open real interval  $(0, 1)$ .

There may arise situations in which it is reasonable to assume that no guessing can occur—that is, where  $\eta_q = 0$  for all  $q \in Q$ . Considering a domain with items in open response format (e.g., mathematical problems in which subjects have to respond with a decimal number representing the result of a computation they have to perform) provides an example. In these situations, the  $\eta_q, q \in Q$ , are no longer free parameters, and thus are not part of the parameter space. In the general case, the parameter vectors  $\theta = (\beta, \eta, \pi)$  consist of  $m = 2 \cdot |Q| + |K| - 1$  components, with  $m$  characterizing the number of free parameters. In addition to Eq. T1, consider the parameter restrictions captured by the equivalent inequalities  $\eta_q < 1 - \beta_q, \beta_q < 1 - \eta_q$ . These restrictions mean nothing else but that a correct response is more likely if the item is mastered (i.e., it is contained in the knowledge state and no careless error occurs) than if it is not mastered (i.e., it is not contained in the knowledge state and a lucky guess occurs). Equivalently, an incorrect response is more likely if the item is not mastered (i.e., it is not contained in the knowledge state and no lucky guess occurs) than if it is mastered (i.e., it is contained in the knowledge state and a careless error occurs). This relation is at the bottom of the idea of a knowledge state, and forms the essence of any stochastic procedure for knowledge assessment that intends to uncover the knowledge state of an individual given the observed responses (Heller & Repitsch, 2011). Both restrictions are captured by the inequality

$$\beta_q + \eta_q < 1 \text{ for all } q \in Q. \tag{T2}$$

The parameter space D then is defined to consist of all vectors  $\theta = (\beta, \eta, \pi)$ , such that

$$D = \{ \theta \in (0, 1)^m \mid \theta \text{ satisfies Eqs. T1 and T2 } \}.$$

It follows that D is an open set that is convex, and thus connected.

If Eq. 1 is reformulated in terms of the parameters of the given BLIM, then the expression  $f_R(\theta)$  predicting the probability of the response pattern  $R \in \mathcal{R}^* = \mathcal{R} \setminus \{Q\}$  is a sum of the products of the parameters. This means that the prediction function

$$f(\theta) = (f_R(\theta))_{R \in \mathcal{R}^*}, \tag{3}$$

$\theta \in D$ , is an analytic function, and, together with D being a connected open set of positive Lebesgue measure, we can draw upon Theorem B5 of Bamber and van Santen (1985). This allows for considering the rank of the Jacobian matrix at any randomly selected point in the parameter space. Again, we restrict the set of response patterns in order to obtain probabilities  $\mathbf{p}_R = f_R(\theta)$  of observing a specific response pattern  $R$ —the *observables*, in the terminology of Bamber and van Santen (2000)—which are independent. With this convention at hand, the prediction function of

the BLIM can be rewritten in the following *nonredundant* form:

$$f_R(\theta) = \sum_{K \in \mathcal{K}^*} P(R|K)\pi_K + P(R|Q) \left( 1 - \sum_{K \in \mathcal{K}^*} \pi_K \right).$$

Analytic expressions can be derived for the entries of the Jacobian matrix of the prediction function of the BLIM. While its number of rows is the maximum number of independent observables  $2^{|Q|} - 1$ , its number of columns is the number of parameters  $m$ . A single entry of the Jacobian matrix has the general form

$$J_{Rj} = \frac{\partial f_R(\theta)}{\partial \theta_j},$$

where  $\theta_j$  represents the  $j$ th parameter in  $\theta$ .

First derivatives of  $f_R(\theta)$  with respect to each of the three types of parameters  $\beta, \eta$ , and  $\pi$  collected in  $\theta$  define the entries of the Jacobian matrix. Skipping the algebra that leads to these derivatives, we only provide the final results, which are based on the following collections, defined for every single item  $q \in Q$ :

$$\mathcal{K}_q^* = \{ K \in \mathcal{K}^* : q \in K \}$$

and

$$\overline{\mathcal{K}}_q^* = \mathcal{K}^* \setminus \mathcal{K}_q^* = \{ K \in \mathcal{K}^* : q \notin K \}.$$

Given any response pattern  $R \subseteq Q$ , the first derivatives of the prediction function  $f_R(\theta)$  with respect to each of the three types of parameters are, respectively:

$$\frac{\partial f_R(\theta)}{\partial \pi_K} = P(R|K) - P(R|Q)$$

$$\frac{\partial f_R(\theta)}{\partial \beta_q} = \begin{cases} \frac{1}{\beta_q} \left[ \sum_{K \in \mathcal{K}_q^*} P(R|K)\pi_K + P(R|Q) \left( 1 - \sum_{K \in \mathcal{K}^*} \pi_K \right) \right] & \text{if } q \notin R \\ -\frac{1}{1-\beta_q} \left[ \sum_{K \in \mathcal{K}_q^*} P(R|K)\pi_K + P(R|Q) \left( 1 - \sum_{K \in \mathcal{K}^*} \pi_K \right) \right] & \text{if } q \in R \end{cases}$$

$$\frac{\partial f_R(\theta)}{\partial \eta_q} = \begin{cases} \frac{1}{\eta_q} \sum_{K \in \mathcal{K}_q^*} P(R|K)\pi_K & \text{if } q \in R \\ -\frac{1}{1-\eta_q} \sum_{K \in \mathcal{K}_q^*} P(R|K)\pi_K & \text{if } q \notin R \end{cases}$$

Given some suitable parameter vector  $\theta$  and a fixed knowledge structure  $\mathcal{K}$ , a numerical computation of the Jacobian matrix  $\mathbf{J}_{\theta, \mathcal{K}}$  consists of an application of the three equations given above to every single response pattern  $R \subseteq Q$ .

### Testing local identifiability

A formal test of local identification of the BLIM for a given knowledge structure  $\mathcal{K}$  consists of deciding whether the condition

$$\max_{\theta \in \mathcal{D}} \text{rank}(\mathbf{J}_{\theta, \mathcal{K}}) = 2|Q| + |\mathcal{K}| - 1$$

holds true, where  $\mathcal{R}$  represents the parameter space of the BLIM and  $\mathbf{J}_{\theta, \mathcal{K}}$  is the Jacobian matrix of the prediction function evaluated at  $\theta$ , when the knowledge structure is  $\mathcal{K}$ . In practice, due to the already mentioned Theorem B5 of Bamber and van Santen (1985), it will suffice to pick some vector  $\theta_0$  at random from the parameter space of the BLIM and to test the condition

$$\text{rank}(\mathbf{J}_{\theta_0, \mathcal{K}}) = 2|Q| + |\mathcal{K}| - 1.$$

If this condition does not hold—that is, when  $\text{rank}(\mathbf{J}_{\theta_0, \mathcal{K}}) < 2|Q| + |\mathcal{K}| - 1$ —there are linear dependencies among some of the columns of  $\mathbf{J}_{\theta_0, \mathcal{K}}$  along one or more dimensions, implying that tradeoffs among some of the model parameters are allowed along such dimensions. The exact number of dimensions along which tradeoffs among parameters are allowed corresponds to the dimension of the null space of  $\mathbf{J}_{\theta_0, \mathcal{K}}$ . For  $m = 2|Q| + |\mathcal{K}| - 1$ , this number is simply

$$\text{null}(\mathbf{J}_{\theta_0, \mathcal{K}}) = m - \text{rank}(\mathbf{J}_{\theta_0, \mathcal{K}}).$$

If the aim is fixing an identification problem,  $\text{null}(\mathbf{J}_{\theta_0, \mathcal{K}})$  also gives the total number of parameters that should be removed from the model (e.g., by setting them to some constant value). However, it should be observed that neither the rank of the Jacobian matrix nor its null space dimension is informative about which *specific* parameters of the model are involved in tradeoffs. Suppose, for instance, that  $\text{null}(\mathbf{J}_{\theta_0, \mathcal{K}}) = 3$ . This indicates that exactly three parameters should be removed. However, removing any three arbitrary parameters will not always produce an identifiable model. This is because only parameters involved in tradeoffs should be removed.

### Detecting parameters that are involved in tradeoffs

Once it has been assessed that, for a given knowledge structure  $\mathcal{K}$ , the BLIM suffers from some identification problems, the next step is to try to fix the problem by setting some of the parameters to constant values. As stated in the previous section, only parameters involved in tradeoffs are good candidates, while the remaining parameters should be left free to vary. It is thus important to have a way to detect all those parameters that are not independent and, among them, those that should be removed from the model. There

exists a particular transformation of the Jacobian matrix that does the job: the reduced row echelon form of  $\mathbf{J}_{\theta, \mathcal{K}}$ .

Formally, a matrix  $\mathbf{R}$  is said to be in *row echelon form* (REF) if it satisfies the following conditions (see, e.g., Meyer, 2000):

- all rows with at least one nonzero element (nonzero rows) are above rows of all zeros;
- the leading coefficient (pivot) of a nonzero row is always strictly to the right of the pivot of the row above it.

Then,  $\mathbf{R}$  is said to be in *reduced row echelon form* (RREF) if it is in REF and satisfies the additional condition

- every pivot is 1 and is the only nonzero entry in its column.

If  $\mathbf{R}$  is in RREF, its columns are called either *basic* or *nonbasic*. The basic columns are those containing the pivots, and there are as many of them as the rank of  $\mathbf{R}$ . A nice property of  $\mathbf{R}$  is that every nonbasic column is a linear combination of the basic columns to the left of it and, if  $\mathbf{R}$  is the RREF of some matrix  $\mathbf{A}$ , then exactly the same relationships can be found among the columns of  $\mathbf{A}$ , so that, if  $a_{ij}$  are the entries of  $\mathbf{A}$ ,  $r_{ij}$  the entries of  $\mathbf{R}$ , and  $k$  the index of any nonbasic column of  $\mathbf{A}$ , then the following linear relation holds true for any row  $i$  of  $\mathbf{A}$ :

$$a_{ik} = r_{k1}a_{i1} + r_{k2}a_{i2} + \dots + r_{k,k-1}a_{i,k-1}$$

and the general form of any nonbasic column  $k$  of  $\mathbf{R}$  is

$$\mathbf{r}_k = [r_{k1}, r_{k2}, \dots, r_{k(k-1)}, 0, \dots, 0]^T.$$

In particular, the transformation

$$\mathbf{r}_k^* = [-r_{k1}, -r_{k2}, \dots, -r_{k(k-1)}, 1, 0, \dots, 0]^T$$

spans the null space of  $\mathbf{A}$ , in the sense that  $\mathbf{A}\mathbf{r}_k^* = 0$ .

Bearing in mind the properties given above, if  $\mathbf{R}_{\theta, \mathcal{K}}$  represents the RREF of the Jacobian matrix  $\mathbf{J}_{\theta, \mathcal{K}}$ , then the nonbasic columns of  $\mathbf{R}_{\theta, \mathcal{K}}$  are just what we are looking for: Each of them represents a parameter that should be removed from the model.

Thus, a procedure for detecting parameters that should be removed from a nonidentifiable model consists of the following steps:

- compute the Jacobian matrix  $\mathbf{J}_{\theta, \mathcal{K}}$ ;
- transform  $\mathbf{J}_{\theta, \mathcal{K}}$  into its RREF version  $\mathbf{R}_{\theta, \mathcal{K}}$ ;
- find the set  $N$  of all of the nonbasic columns of  $\mathbf{R}_{\theta, \mathcal{K}}$  (i.e., those columns without a leading coefficient);
- each nonbasic column in  $N$  corresponds to a parameter that should be removed from the model.

It should be observed that the columns actually contained in  $N$  might depend on how the columns are ordered from left to right in the matrix  $\mathbf{J}_{\theta, \mathcal{K}}$ , in the sense that different

permutations of the columns in  $\mathbf{J}_{\theta, \mathcal{K}}$  might provide different results for  $N$ . Suppose, in fact, that a linear dependence exists among, say, three columns along a single dimension. In that case, any one of these three columns can be arbitrarily expressed as a linear combination of the other two.

Tradeoff dimensions

The BLIM essentially consists of three different types of parameters: lucky guesses  $\eta$ , careless errors  $\beta$ , and knowledge state probabilities  $\pi$ . Unidentifiability is a consequence of tradeoffs among these three types of parameters. It seems natural to distinguish tradeoffs that occur among parameters of the same type from tradeoffs that occur across different parameter types. To be more specific, three orders of tradeoffs can be recognized:

1. *first-order trade-offs*: Parameters involved in this type of tradeoff are all of the same type (e.g., a tradeoff among the lucky guess probabilities of different items);
2. *second-order tradeoffs*: These are tradeoffs involving parameters of two different types (e.g., a tradeoff between the careless error probability of some item and the probability of some knowledge state);
3. *third-order tradeoffs*: These involve parameters of all three types (e.g., one involving the lucky guess of an item, the careless error of another item, and the probability of some knowledge state).

These three classes of tradeoffs can be studied separately through a suitable decomposition of the Jacobian matrix into the three different submatrices  $\mathbf{P}$ ,  $\mathbf{B}$ , and  $\mathbf{E}$ , whose entries are, respectively:

$$\mathbf{P}_{RK} = \frac{\partial f_R(\theta)}{\partial \pi_K}, \mathbf{B}_{Rq} = \frac{\partial f_R(\theta)}{\partial \beta_q}, \mathbf{E}_{Rq} = \frac{\partial f_R(\theta)}{\partial \eta_q},$$

where  $K \in \mathcal{K}^*$ ,  $R \in \mathcal{R}^*$ , and  $q \in \mathcal{Q}$ . Thus,  $\mathbf{P}$  is the  $(2^{|\mathcal{Q}|-1} \times (|\mathcal{K}| - 1))$  matrix of the first derivatives of the prediction function  $f$  with respect to the knowledge state probabilities  $\pi$ ;  $\mathbf{B}$  is the  $(2^{|\mathcal{Q}|-1} \times |\mathcal{Q}|)$  matrix of the first derivatives of  $f$  with respect to the careless error parameters  $\beta$ ;  $\mathbf{E}$  is the  $(2^{|\mathcal{Q}|-1} \times |\mathcal{Q}|)$  matrix of the first derivatives of  $f$  with respect to the lucky guess parameters  $\eta$ . The whole Jacobian matrix is reconstructed by

$$\mathbf{J}_{\theta, \mathcal{K}} = [\mathbf{P}|\mathbf{B}|\mathbf{E}].$$

First-order tradeoffs among parameters of the same type occur when one or more submatrices are not of full rank. For instance,  $\text{rank}(\mathbf{B}) < |\mathcal{Q}|$  would indicate that tradeoffs among the  $\beta$  parameters of some of the items would be allowed. In this case, the total number of first-order tradeoff dimensions for such item parameters would be measured by  $\text{null}(\mathbf{B}) = |\mathcal{Q}| - \text{rank}(\mathbf{B})$ . Similar considerations apply to the two submatrices  $\mathbf{P}$

and  $\mathbf{E}$ . Therefore, the total number of first-order tradeoff dimensions for all parameter types is given by

$$\text{null}(\mathbf{P}) + \text{null}(\mathbf{B}) + \text{null}(\mathbf{E}).$$

Second-order tradeoffs can be assessed by considering pairs of submatrices. For  $\mathbf{X}, \mathbf{Y} \in \{\mathbf{P}, \mathbf{B}, \mathbf{E}\}$ , there are in total three possible pairs of submatrices, each of which gives rise to a compound submatrix of the form  $[\mathbf{X} | \mathbf{Y}]$ . The number of second-order tradeoff dimensions contained in submatrix  $[\mathbf{X} | \mathbf{Y}]$  is obtained by subtracting the number of first-order tradeoff dimensions from the total number of tradeoff dimensions in  $[\mathbf{X} | \mathbf{Y}]$ . If we indicate with  $td_2([\mathbf{X} | \mathbf{Y}])$  the number of second-order tradeoff dimensions contained in  $[\mathbf{X} | \mathbf{Y}]$ , we can write

$$td_2([\mathbf{X}|\mathbf{Y}]) = \text{null}([\mathbf{X}|\mathbf{Y}]) - \text{null}(\mathbf{X}) - \text{null}(\mathbf{Y}),$$

which can also be expressed in terms of ranks. If  $\mathbf{X}$  and  $\mathbf{Y}$  have  $c$  and  $d$  columns, respectively, we can write

$$td_2([\mathbf{X}|\mathbf{Y}]) = c + d - \text{rank}([\mathbf{X}|\mathbf{Y}]) - c + \text{rank}(\mathbf{X}) - d + \text{rank}(\mathbf{Y}) = \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y}) - \text{rank}([\mathbf{X}|\mathbf{Y}]).$$

Finally, in order to count the number of third-order tradeoff dimensions, the whole matrix  $[\mathbf{P} | \mathbf{B} | \mathbf{E}]$  has to be considered. With  $td_3([\mathbf{P} | \mathbf{B} | \mathbf{E}])$  denoting this number, the requirement is that

$$\text{null}([\mathbf{P}|\mathbf{B}|\mathbf{E}]) = \text{null}(\mathbf{P}) + \text{null}(\mathbf{B}) + \text{null}(\mathbf{E}) + td_2([\mathbf{P}|\mathbf{B}]) + td_2([\mathbf{P}|\mathbf{E}]) + td_2([\mathbf{B}|\mathbf{E}]) + td_3([\mathbf{P}|\mathbf{B}|\mathbf{E}]),$$

from which, after some algebra, we obtain

$$td_3([\mathbf{P}|\mathbf{B}|\mathbf{E}]) = \text{rank}([\mathbf{P}|\mathbf{B}]) + \text{rank}([\mathbf{P}|\mathbf{E}]) + \text{rank}([\mathbf{B}|\mathbf{E}]) - \text{rank}(\mathbf{P}) - \text{rank}(\mathbf{B}) - \text{rank}(\mathbf{E}) - \text{rank}([\mathbf{P}|\mathbf{B}|\mathbf{E}]).$$

It is clear that identification of the BLIM strictly depends on the chosen knowledge structure  $\mathcal{K}$ . For a finite set  $\mathcal{Q}$ , the family of all possible knowledge structures is the collection of all subsets of the power set  $2^{\mathcal{Q}}$  containing at least the empty set and  $\mathcal{Q}$ . This collection can be theoretically partitioned into the set of all knowledge structures for which the BLIM is identifiable and the set of all those knowledge structures for which it is not. Considering knowledge structures belonging to the latter collection, they might differ from one another concerning the types of tradeoffs occurring among the model parameters. There might be a lot to study about similarities and differences in this respect.

In the rest of this section, a basic result is provided, which holds for all knowledge structures on a finite set  $\mathcal{Q}$ . It basically says that there are no first-order tradeoffs among the probabilities of the knowledge states. In this direction, the following lemma is useful.

**Lemma 1** Given any two  $n \times m$  matrices  $\mathbf{X}$  and  $\mathbf{Y}$ , it holds that  $\text{rank}(\mathbf{X} - \mathbf{Y}) \geq \text{rank}(\mathbf{X}) - \text{rank}(\mathbf{Y})$ .

*Proof* By the subadditivity property of the rank, it holds that  $\text{rank}(\mathbf{X} + \mathbf{Y}) \leq \text{rank}(\mathbf{X}) + \text{rank}(\mathbf{Y})$ . Therefore, we can write  $\text{rank}(\mathbf{X} - \mathbf{Y} + \mathbf{Y}) \leq \text{rank}(\mathbf{X} - \mathbf{Y}) + \text{rank}(\mathbf{Y})$ , and thus  $\text{rank}(\mathbf{X}) - \text{rank}(\mathbf{Y}) \leq \text{rank}(\mathbf{X} - \mathbf{Y})$ .

**Theorem 1** Let  $\mathcal{K}$  be any knowledge structure on the finite set  $Q$  and let  $\mathbf{J} = [\mathbf{B} \mid \mathbf{E} \mid \mathbf{P}]$  be the Jacobian matrix of the prediction function  $f$  corresponding to  $\mathcal{K}$ . Then  $\mathbf{P}$  is full rank.

*Proof* We first consider the case in which the knowledge structure  $\mathcal{K}$  equals the power set on  $Q$ : namely, when  $\mathcal{K} = 2^Q$ . Let  $\mathbf{J}^* = [\mathbf{B}^* \mid \mathbf{E}^* \mid \mathbf{P}^*]$  be the corresponding Jacobian matrix. For  $k = 2^{|Q|} - 1$ , define the  $k \times k$  matrix  $\mathbf{U} = [P(R \mid K)]$ ,  $R \in R^*$ ,  $K \in \mathcal{K}^*$ , and the  $k \times 1$  column vector  $\mathbf{v} = [P(R \mid Q)]$ . Then, assuming that  $\mathbf{v}$  is the last column of  $\mathbf{U}$ , we have the following equality:

$$[\mathbf{P}^* | 0] = \mathbf{U} - \mathbf{V},$$

where  $\mathbf{V} = [\mathbf{v} | \mathbf{v} | \dots | \mathbf{v}]$  is the  $k \times k$  matrix obtained by replicating vector  $\mathbf{v}$  for  $k$  times, and  $0$  is the  $k \times 1$  column vector of zeros. It is then clear that the rank of matrix  $\mathbf{V}$  is 1.

For any  $q \in Q$ , define the  $2 \times 2$  matrix

$$\mathbf{U}_q = \begin{pmatrix} 1 - \eta_q & \beta_q \\ \eta_q & 1 - \beta_q \end{pmatrix}$$

Then, up to permuting rows appropriately,  $\mathbf{U}$  can be rewritten as

$$\mathbf{U} = \mathbf{U}_1 \otimes \mathbf{U}_2 \otimes \dots \otimes \mathbf{U}_n \tag{4}$$

where  $\otimes$  denotes the Kronecker product of matrices. The matrix  $\mathbf{U}_q$  is full rank if and only if it has a nonzero determinant that is, when

$$(1 - \eta_q)(1 - \beta_q) - \eta_q\beta_q \neq 0,$$

which gives

$$\eta_q + \beta_q \neq 1. \tag{5}$$

From assumption T2 concerning the parameter space of a probabilistic knowledge structure, we know that this condition holds true for all  $q \in Q$ , so that indeed  $\text{rank}(\mathbf{U}_q) = 2$ . Drawing upon the general properties of the Kronecker product, from Eq. 4 it follows that

$$\text{rank}(\mathbf{U}) = \prod_{q \in Q} \text{rank}(\mathbf{U}_q).$$

Therefore, we have  $\text{rank}(\mathbf{U}) = 2^{|Q|}$ . Then, for Lemma 1, we have

$$\text{rank}([\mathbf{P}^* | 0]) = \text{rank}(\mathbf{U} - \mathbf{V}) \geq \text{rank}(\mathbf{U}) - \text{rank}(\mathbf{V}) = 2^{|Q|} - 1.$$

However, since the size of  $\mathbf{P}^*$  is  $k \times k$ , its rank must be at most  $k$ , and adding a column of zeros to this matrix does not increase its rank; therefore, we have the upper bound

$$\text{rank}([\mathbf{P}^* | 0]) \leq 2^{|Q|} - 1. \tag{6}$$

Thus, we must have  $\text{rank}([\mathbf{P}^* | 0]) = 2^{|Q|} - 1$ . By removing column 0 from  $[\mathbf{P}^* | 0]$ , the rank does not change, so that we have  $\text{rank}(\mathbf{P}^*) = 2^{|Q|} - 1$ , and thus  $\mathbf{P}^*$  is full rank. We also observe that any matrix obtained by removing arbitrary columns from  $\mathbf{P}^*$  will also be full rank. Suppose now that  $\mathcal{K}$  is any knowledge structure included in  $2^Q$ . The corresponding matrix  $\mathbf{P}$  can then be obtained from  $\mathbf{P}^*$  by removing from it all columns that do not correspond to knowledge states. Thus,  $\mathbf{P}$  is full rank.

### Constructing the Jacobian matrix in practice

The number of rows in the Jacobian matrix of the BLIM prediction function equals  $2^{|Q|} - 1$ . In practice, the construction of this matrix is feasible only if the number of items is relatively small. To give an example, with  $|Q| = 20$  items, the number of rows of this matrix would be more than one million. From a computational point of view, this aspect represents a serious limitation of the proposed identification test, even with computers that are equipped with a huge amount of memory.

Nonetheless, there is a way to get around this computational problem, which is based on the fact that any maximal independent subset of rows from a matrix  $\mathbf{X}$  contains exactly  $\text{rank}(\mathbf{X})$  rows. Let  $\mathbf{J}_{\theta, \mathcal{K}}$  be the Jacobian matrix of the BLIM for some knowledge structure  $(Q, \mathcal{K})$ , and let  $\mathcal{M}$  be any maximal independent subset of rows from  $\mathbf{J}_{\theta, \mathcal{K}}$ . Now, if  $\mathbf{M}$  is any of the matrices whose rows are some permutation of the row vectors in  $\mathcal{M}$ , then it holds that

$$\text{rank}(\mathbf{J}_{\theta, \mathcal{K}}) = \text{rank}(\mathbf{M}).$$

The important fact about the matrix  $\mathbf{M}$  is that the number of its rows might be much smaller than that in  $\mathbf{J}_{\theta, \mathcal{K}}$ . More precisely, for any empirically testable model, this number will be at most  $2|Q| + |\mathcal{K}| - 1$ —that is, the total number of parameters in the model (columns in  $\mathbf{J}_{\theta, \mathcal{K}}$ ). Moreover, this number attains its maximum value only if  $\mathbf{J}_{\theta, \mathcal{K}}$  is full rank—that is, when the model is locally identifiable.

At this point, however, the problem arises how to construct the matrix  $\mathbf{M}$ , or, stated another way, how to find any of the maximally independent subsets of rows from  $\mathbf{J}_{\theta, \mathcal{K}}$ . Let  $\mathcal{J}$  denote the collection of all row vectors from  $\mathbf{J}_{\theta, \mathcal{K}}$ , and suppose that  $\mathcal{S} \subseteq \mathcal{J}$  is an independent subset. In this case, either of the following two conditions holds true:

1.  $\mathcal{S}$  is a maximal independent subset.
2. There exists  $\mathbf{r} \in \mathcal{J} \setminus \mathcal{S}$  such that  $\mathcal{S} \cup \{\mathbf{r}\}$  is independent.

Given the observation above, a maximal independent subset of rows from  $\mathbf{J}_{\theta, \mathcal{K}}$  can be constructed iteratively. We first recall that the row vectors in  $\mathcal{J}$  are in a one-to-one correspondence with the response patterns in  $\mathcal{R}^*$ . Moreover, for  $Z = \{0, 1, \dots, 2^{|\mathcal{Q}|-2}\}$ , let  $h: Z \rightarrow 2^{\mathcal{Q}} \setminus \{Q\}$  be a bijection. For instance,  $h$  can be regarded as the lexicographic order, so that for any  $z \in Z$ ,  $h(z)$  is the collection of all digits that are set at 1 in the binary representation of  $z$ . Therefore, for example,  $h(0) = \emptyset$ ,  $h(1) = \{1\}$ ,  $h(2) = \{2\}$ ,  $h(3) = \{1, 2\}$ , and so forth. But other choices for the bijection  $h$  might be considered as well, of course.

For  $\mathcal{S} \subseteq \mathcal{J}$ , let  $[\mathcal{S}]$  denote any of the matrices whose rows are a permutation of the row vectors in  $\mathcal{S}$ , and let  $m = 2|\mathcal{Q}| + |\mathcal{K}| - 1$  be the total number of parameters in the model. Then, an algorithm for constructing the matrix  $\mathbf{M}$  is as follows:

```

 $\mathcal{S} \leftarrow \emptyset, i \leftarrow 0$ 
while  $\text{rank}([\mathcal{S}]) < m$  and  $i \leq 2^{|\mathcal{Q}|-2}$  do
     $R \leftarrow h(i)$ 
     $\mathbf{r} \leftarrow \left\{ \frac{\partial f_R(\boldsymbol{\theta})}{\partial \theta_j} \right\}_{1 \leq j \leq m}$ 
    if  $\text{rank}([\mathcal{S} \cup \{\mathbf{r}\}]) > \text{rank}([\mathcal{S}])$  then
         $\mathcal{S} \leftarrow \mathcal{S} \cup \{\mathbf{r}\}$ 
    end
     $i \leftarrow i + 1$ 
end
 $\mathbf{M} \leftarrow [\mathcal{S}]$ 

```

A few remarks about this algorithm are in order. At the outset, the collection  $\mathcal{S}$  is initialized to the empty set. Then the `while` loop is entered and repeated until one of the following two conditions is satisfied:

- (1) The rank of  $[\mathcal{S}]$  equals  $m$ .
- (2) All response patterns in  $\mathcal{R}^*$  have been examined.

In every single step  $i > 0$  of the while loop, a new response pattern  $R = h(i)$  is produced, and the corresponding row vector  $\mathbf{r}$  is added to  $\mathcal{S}$  only if it increases the rank of  $[\mathcal{S}]$ . Now suppose that  $\mathcal{S}$  is an independent subset. If the condition  $\text{rank}([\mathcal{S} \cup \{\mathbf{r}\}]) > \text{rank}([\mathcal{S}])$  holds true, then  $\mathcal{S} \cup \{\mathbf{r}\}$  must be an independent subset, too. From this observation and the fact that at the outset the collection  $\mathcal{S}$  is empty, it follows by induction that at each step  $i > 0$ , the collection  $\mathcal{S}$  is an independent subset.

Concerning termination of the algorithm, suppose that at step  $i > 0$  Condition 1 is met first. Then, regardless of which response patterns have been examined so far, the rank of  $[\mathcal{S}]$  equals  $m$ , which is the maximum value that can be attained by  $\text{rank}([\mathcal{S}])$ .

Thus,  $\mathcal{S}$  must be a maximal independent subset of  $\mathcal{J}$ . On the other hand, suppose that Condition 2 is met (thus,  $i > 2^{|\mathcal{Q}|-2}$ ). In this case, all possible row vectors in  $\mathcal{J}$  have been taken into consideration. Therefore, there is no row vector  $r \in \mathcal{J}$  such that  $\mathcal{S} \cup \{\mathbf{r}\}$  is an independent subset, meaning that  $\mathcal{S}$  is maximal.

With the purpose of testing the local identifiability of the BLIM, the matrix  $\mathbf{M} = [\mathcal{S}]$  can be used in place of  $\mathbf{J}_{\theta, \mathcal{K}}$ . The conclusions are the same as with  $\mathbf{J}_{\theta, \mathcal{K}}$ : Namely, if  $\text{rank}(\mathbf{M}) < m$ , then the model is not locally identifiable. It is worth noticing that in this case, the number of rows of  $\mathbf{M}$ , which equals its rank, will be less than the number of columns, which equals the total number  $m$  of parameters. Thus, the difference between the columns and the rows of this matrix gives the total number of tradeoff dimensions.

With the aim of providing a simple example of how the algorithm works, consider the simple knowledge structure  $\mathcal{K} = \{\emptyset, \{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}\}$ , for which the BLIM happens to be locally identifiable. With respect to this knowledge structure, the BLIM contains in the whole  $4 + 4 + 3 = 11$  free parameters.

At the outset,  $\mathcal{S}$  is empty. At the first step of the algorithm, the empty response pattern is evaluated, and the row of the first derivatives of the probability of this pattern with respect to the 11 model parameters is computed. This row is added to  $\mathcal{S}$ , and the matrix  $[\mathcal{S}]$  happens to be rank 1. At the second step, the response pattern  $\{1\}$  is evaluated, and a corresponding row is added to  $\mathcal{S}$ , which now contains two row vectors. After this change, the rank of  $[\mathcal{S}]$  increases by 1, meaning that  $\mathcal{S}$  is an independent set. Therefore, the algorithm moves to the third step, where response pattern  $\{2\}$  is evaluated. The procedure repeats, and the rank increases each time by 1 up to the response pattern  $\{2, 4\}$ . At this point,  $\mathcal{S}$  contains 11 row vectors, but the rank of  $[\mathcal{S}]$  is 10. Thus, the last column of  $\mathcal{S}$  is removed, and the new response pattern  $\{1, 2, 4\}$  is evaluated. However, after the row vector corresponding to this pattern has been added to  $\mathcal{S}$ , the rank of  $[\mathcal{S}]$  is still 10. Therefore this row is also removed from  $\mathcal{S}$ , and the new response pattern  $\{3, 4\}$  is evaluated. Once the row vector corresponding to this pattern is added to  $\mathcal{S}$ , the rank of  $[\mathcal{S}]$  becomes 11, and since this number equals the total number of parameters, the algorithm terminates.

### The BLIMIT function

A computerized procedure for testing the local identifiability of the BLIM, which includes the various features examined in the previous sections, has been implemented as a MATLAB function.<sup>1</sup> This function is named BLIMIT, which stands for

<sup>1</sup> The MATLAB code of the BLIMIT function can be requested by sending an e-mail to the first author (luca.stefanutti@unipd.it).



“basic local independence model identification test.” In its basic usage, the function takes as its input argument a binary matrix representation of the knowledge structure for which the identification test has to be performed, and it returns a detailed identification report, which is saved into an external text file.

In this section, the main features of the BLIMIT function are illustrated by means of a simple example. Suppose that one would like to test whether the BLIM is locally identifiable for the knowledge structure

$$\mathcal{K} = \{\emptyset, \{1\}, \{2\}, \{1, 2, 3\}, \{1, 2, 4\}, Q\}$$

on the set  $Q = \{1, 2, 3, 4\}$ . The BLIMIT function accepts as an input argument a *model*, which is a MATLAB-structured object in which at least one field named *states* is defined. This field is a binary matrix, consisting of  $|\mathcal{K}|$  rows and  $|Q|$  columns, that represents the knowledge structure to be tested. Each row of the matrix represents a distinct knowledge state in  $\mathcal{K}$ , and each column a different item in  $Q$ . A 1 at the intersection between row  $i$  and column  $j$  of this matrix means that knowledge state  $i$  contains item  $j$ , whereas a 0 means the opposite. Suppose that the model is called *mymodel*; the knowledge structure  $\mathcal{K}$  is then specified by the following syntax:

```
mymodel.states = [
    0 0 0 0
    1 0 0 0
    0 1 0 0
    1 1 1 0
    1 1 0 1
    1 1 1 1
];
```

The function is then called by the syntax `blimit (mymodel)`. Optionally, values for the model parameters can also be specified as input arguments (there is one field for each vector of parameters, labeled, respectively, *beta* for the careless error probabilities, *eta* for the guessing probabilities, and *pi* for the knowledge state probabilities). If these values are not specified, the BLIMIT function generates them at random, and this corresponds to selecting a point  $\theta_0$  from the parameter space.

The function does not compute the whole Jacobian matrix of the BLIM but rather, as explained in [Constructing the Jacobian matrix in practice](#), a maximal independent subset of row vectors from the Jacobian is constructed, and any subsequent computation is performed on this subset of vectors—or more precisely, on a matrix  $\mathbf{M}$  whose rows are some permutation of the vectors in this subset.

The output from the function is a text file that, by default, is printed on the MATLAB command window. Basically, this file consists of three main parts: a general

information section and two other sections that appear only when the model is not identifiable. The first of them is a table, called the *submatrix rank analysis table*, in which the matrix  $\mathbf{M}$  is partitioned in different ways; the last part contains specific diagnostic information about the item parameters.

The output from the BLIMIT function is now examined for the example at hand. Thus, suppose that the function has been called for the model *mymodel* defined above. The first part of the output file looks like the following one

---

```
B L I M I T
BASIC LOCAL INDEPENDENCE MODEL IDENTIFICATION TEST
Number of items:           4
Number of knowledge states: 6
Total number of parameters: 13
Jacobian matrix rank:      10
Null space dimension (NSD): 3
Identification problems detected:
Jacobian matrix is not full rank.
```

---

Besides the numbers of items and states, the report identifies the total number of parameters in the model (which in our example is  $2|Q| + |\mathcal{K}| - 1 = 2 \times 4 + 6 - 1 = 13$ ), the Jacobian rank, and the null space dimension. Notice that the rank of the Jacobian is only 10. Since this number is less than the total number of parameters, the Jacobian is not full rank, and thus the model is not locally identifiable. The difference between the total number of parameters and the Jacobian rank is called the *null space dimension* (NSD). This is the overall number of dimensions along which tradeoffs among the model parameters are allowed. Since  $NSD = 3$ , in our example there are, in total, three distinct tradeoff dimensions.

The remaining part of the report provides detailed diagnostic information that can be used for fixing the identification problem. The submatrix rank analysis table is a table in which the Jacobian matrix is partitioned in different ways (see [Tradeoff dimensions](#)). The complete Jacobian matrix can be regarded as the joining of three different submatrices corresponding to the three different types of parameters in the BLIM model (careless error, lucky guess, and state probabilities). For the example at hand, the submatrix rank analysis table looks like the listing below. Each of the three submatrices listed in the upper part of the table has as a number of columns that corresponds to the number of parameters of a certain type (N.PAR). In terms of the notation used in [Tradeoff dimension](#), [BETA] stands for the submatrix  $\mathbf{B}$ , [ETA] stands for  $\mathbf{E}$ , an [PI] for  $\mathbf{P}$ . The submatrix  $\mathbf{B}$  is the part of the Jacobian whose columns are the  $n = 4$  different careless error probabilities of the items; the submatrix  $\mathbf{E}$  is the part of the Jacobian whose columns are the  $n = 4$  different lucky guess probabilities; the submatrix  $\mathbf{P}$  is the part whose columns are the knowledge state

probabilities (for  $m$  knowledge states, there are on the whole  $m - 1$  such columns).

---

Submatrix rank analysis table  
 [BETA] = submatrix of the careless error parameters  
 [ETA] = submatrix of the lucky guess parameters  
 [PI] = submatrix of the state probabilities

---

SUBMATRIX	N. PAR	RANK	NSD	TRADEOFF DIM
[BETA]	4	4	0	0
[ETA]	4	4	0	0
[PI]	5	5	0	0
[BETA ETA]	8	8	0	0
[BETA PI]	9	8	1	1
[ETA PI]	9	8	1	1
[BETA ETA PI]	13	10	3	1

---

For each of the three submatrices the number of parameters (columns) and the rank are indicated. For instance, for the **B** submatrix there are four parameters, and the rank is also 4. The null space dimension is also indicated, which is 0 for all three submatrices, in our example. The last column of the table indicates the trade-off dimension (TRADEOFF DIM) corresponding to each submatrix. For individual submatrices this number always corresponds to the null space dimension.

In the whole, this part of the table is aimed at highlighting tradeoffs within single submatrices. In our example we see that there are no such kind of trade-offs.

In the second part of the table, all possible pairs of the individual submatrices are considered. Like in the first part of the table, also here the total number of parameters and submatrix rank are indicated along with the NSD and TD. Tradeoffs occur in those submatrix pairs having a positive TRADEOFF DIM. In our example, this is the case of submatrix pairs **[B | P]** and **[E | P]**. In each of these two pairs the residual is 1, meaning that in each of them a tradeoff is allowed along exactly one dimension. The fact that the trade-off dimension of **[B | P]** is 1 means that there is a tradeoff between some careless error parameter and some knowledge state probability. Similar conclusions can be drawn concerning the submatrix pair **[E | P]**.

It is important to bear in mind that this part of the table is concerned with tradeoffs that occur *across* submatrices, and all of them involve exactly two types of parameters (e.g., careless error and knowledge state probability). These types of tradeoffs are called *second-order tradeoffs*, and should be distinguished from the *first-order tradeoffs* occurring *within* individual submatrices (i.e., involving parameters of the same type).

In the third part of the table, *third-order tradeoffs*, that is tradeoffs involving all three types of parameters along the same dimension are considered. In our example the number of third-order tradeoff dimensions is exactly 1. This means that there is some tradeoff involving, along the same dimension, a careless error probability, a lucky guess probability and also some knowledge state probability.

At this point, if we sum up all of the values displayed in the TRADEOFF DIM column of the table, we obtain 3, namely the overall null space dimension of the Jacobian matrix. Therefore, we know that in the whole there are:

- no first-order tradeoff dimensions,
- two second-order tradeoff dimensions, and
- one third-order tradeoff dimension.

It remains to discover which specific model parameters are in a tradeoff relation. Since we learned that there is a second-order tradeoff dimension involving  $\beta$  and  $\pi$  parameters, the next table in the report contains detailed diagnostic information at the item level, concerning the  $\beta$  parameters:

---

```
Item diagnostics for [BETA PI] submatrix
Second-order tradeoff dimensions:      1
-----
```

PARAMS	VALUES	DIM #1
BETA1	0.3085	
BETA2	0.1326	
BETA3	0.4122	-1.4826
BETA4	0.4913	1.0000

---

The text above the table serves as a reminder that for the submatrix pair at issue there is exactly 1 tradeoff dimension. The table contains one row for each item and three columns. The first column is a list of the item parameter labels, whereas the second column contains the parameter values. The third column, labeled DIM #1, is the most important one. In [Detecting parameters that are involved in tradeoffs](#), it has been shown that the reduced row echelon form of the Jacobian matrix might contain nonbasic columns, and that those columns correspond to parameters that are involved in tradeoffs. More precisely, each of the non basic columns is a linear combination of the basic columns that are to the left of it in the RREF of the Jacobian matrix. The values displayed in the DIM #1 column of the table are the negative coefficients of this linear combination, with the convention that, if a coefficient is zero then the corresponding row is empty and that the coefficient of the parameter that can be expressed as a linear combination of the others is 1.

Looking at the table we discover that the  $\beta$  parameters of Items 3 and 4 are not linearly independent. These two parameters are involved in a tradeoff with some state probability along a single dimension. Since there are two parameters and only one tradeoff dimension, it suffices to set just one of them to a constant value to fix the tradeoff problem.

If one is interested in retrieving linear coefficients for all parameters (including knowledge state probabilities) involved in the tradeoff, it is possible to call BLIMIT with the syntax `info=blimit(myModel)`. The output argument `info` contains a number of fields, among which there is the array `info.DiagBetaPi` in which the

mentioned coefficients are stored. The array has as many rows as the number of parameters in the  $[B | P]$  submatrix, and as many columns as the number of tradeoff dimensions for this submatrix.

By examining the submatrix rank analysis table we also noticed that there is a second-order tradeoff dimension involving the  $\eta$  and  $\pi$  parameters. This is why in the next table of the report there are detailed diagnostic information at item level concerning the  $[E | P]$  submatrix pair:

---

```
Item diagnostics for [ETA PI] submatrix
Second-order tradeoff dimensions:      1
-----
```

PARAMS	VALUES	DIM #1
ETA1	0.3651	-0.5048
ETA2	0.1719	1.0000
ETA3	0.2920	
ETA4	0.0539	

---

By examining the third column of this table we discover that there is a tradeoff involving the  $\eta$  parameters of Items 1 and 2 and some knowledge state probability. As in the previous case, it suffices to set at a constant value one of the two lucky guess probabilities to fix the tradeoff problem. If one is interested in retrieving the whole set of linear coefficients, they are stored in the array `info.DiagEtaPi`.

Finally we observed that a third-order tradeoff dimension also exists. Detailed item-level diagnostic information concerning this type of tradeoff is provided in the next table of the report:

---

```
Item diagnostics for [BETA ETA PI] submatrix
Third-order tradeoff dimensions:      1
-----
```

PARAMS	VALUES	DIM #1
BETA1	0.3085	
BETA2	0.1326	
BETA3	0.4122	-2.8428
ETA1	0.3651	1.0000
ETA3	0.2920	
ETA4	0.0539	

---

Notice that parameters  $\beta_4$  and  $\eta_2$  are missing in this table. They are both involved in some second-order trade-off dimension and, for this reason, they are removed from the analysis of third-order trade-off dimensions.

From this last table we observe that the third-order tradeoff dimension involves parameters  $\beta_3$  and  $\eta_1$ , plus some knowledge state probability. The trade-off problem is solved by setting exactly one of these parameters to some constant value. The whole set of linear coefficients for third-order tradeoffs is stored in the array `info.DiagBetaEtaPi`.

We now have complete information for fixing the identification problem. Exactly three parameters in the model have to be set to some constant value, and this corresponds to excluding these parameters from the model. To be more precise, these parameters are still in the model, but they are not free to vary. As a consequence, the first partial derivative of the prediction function with respect to these parameters will be zero. As suggested by the procedure, the parameters that should be excluded are  $\eta_1$ ,  $\eta_2$ , and  $\beta_4$ .

Exclusion of specific parameters can be done in BLIMIT by passing to the function two additional logical vectors (one for each of the two error parameter types) by means of which one can specify which item parameters should be removed from the model. To this aim we add two new vectors to the object `mymodel`, labeled, respectively `beta0` and `eta0`. Since we want to exclude  $\beta_4$  the first of the two vectors is:

```
mymodel.beta0 = [
true
true
true
false
];
```

Since we want to exclude both  $\eta_1$  and  $\eta_2$ , the second vector is:

```
mymodel.eta0 = [
false
false
true
true
];
```

After these changes, as expected, a new call to `blimit(mymodel)` produces the following result:

---

```
B L I M I T
BASIC LOCAL INDEPENDENCE MODEL IDENTIFICATION TEST
Number of items:                4
Number of knowledge states:     6
Total number of parameters:    10
Jacobian matrix rank:          10
Null space dimension (NSD):     0
No identification problems detected.
```

---

Therefore, by removing exactly three parameters ( $\eta_1$ ,  $\eta_2$ , and  $\beta_4$ ) from the model, we were able to solve the identification problem.

At this point it might be argued that the parameter values remain ambiguous anyway: one sets some of them to be constant but, since other choices are possible, the values of the constants might affect how one interprets the parameters. However it should be observed that local identification assesses in a model construction stage, that is *before* any data is observed. It is at that stage that one decides which parameters are free to vary and which are not. In this respect local identification assessment can be regarded as a type of model construction tool, a way to establish if a model, with all of its parameters, is falsifiable or not. If not, then one knows that some parameters of the model cannot be isolated or separated from others.

In the example given above, parameters  $\eta_1$ ,  $\eta_2$ , and  $\beta_4$  were set to zero with the aim of providing a simple and immediate illustration of how the BLIMIT procedure works. In concrete applications however one might consider the alternative of redefining the knowledge structure by removing critical knowledge states and/or adding new ones. This option is discussed in the next two empirical examples.

## Two empirical examples

This section exemplifies the application of the BLIMIT function to real data sets from the domain of elementary probability theory.

The first data set consists of responses that 300 students provided to a collection of 19 problems. An analysis of the content of the problems resulted in the hypothesis that five skills (computation of the probability of an event, probability of the complement of an event, stochastic independence, union of mutually exclusive events, deck of cards) were required in order to solve them. Via the conjunctive model

(Doignon & Falmagne, 1999), each problem was associated with the skills necessary and sufficient to solve it according to the skill map represented in Table 1. Consider Problem 11 as an example: “Throw a dice. What is the probability of obtaining 1 or 4?” The problem was assumed to require the skills concerning computation of the probability of an event and union of events. The knowledge structure  $\mathcal{K}_1$  delineated by the given skill map contains 24 knowledge states.

$$\mathcal{K}_1 = \{\emptyset, \{1\}, \{2\}, \{6\}, \{14\}, \{1, 4\}, \{2, 4\}, \{1, 6, 10\}, \{1, 14, 17\}, \{2, 6, 8\}, \{1, 2, 11, 19\}, \{1, 4, 6, 10\}, \{1, 4, 14, 17\}, \{1, 2, 4, 11, 19\}, \{3, 5, 6, 12, 14\}, \{2, 3, 5, 6, 8, 12, 14\}, \{1, 2, 7, 11, 13, 14, 17, 19\}, \{1, 2, 4, 7, 11, 13, 14, 17, 19\}, \{1, 2, 6, 8, 9, 10, 11, 18, 19\}, \{1, 3, 5, 6, 10, 12, 14, 16, 17\}, \{1, 2, 4, 6, 8, 9, 10, 11, 18, 19\}, \{1, 3, 4, 5, 6, 10, 12, 14, 16, 17\}, \{1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19\}, Q\}$$

Model parameters were estimated by maximum-likelihood. The fit of the BLIM for  $\mathcal{K}_1$  was tested using Pearson’s chi-square statistic and a parametric bootstrap (217 response patterns had expected frequency  $<1$ ), and it turned out to be good ( $p = .16$ ).

Local identifiability of the BLIM for  $\mathcal{K}_1$  was tested through the BLIMIT function. The values of the model parameters (careless error probabilities, guessing probabilities, knowledge state probabilities) estimated on the empirical data were passed to the function, together with the knowledge structure  $\mathcal{K}_1$ .

The model is not locally identifiable, and there is one second-order tradeoff dimension involving the  $\beta$  parameter of Problem 4 and some state probability.

At this point, there are two possibilities for solving the identification problem. On the one hand, we can modify the knowledge structure that is tested. An indication in this direction is to modify the skill assignment concerning Problem 4 because this last was involved in the tradeoff dimension. Problem 4 says: “Assuming a deck of cards contains 52 different playing cards, what is the probability of extracting a black 4?” The problem was assumed to require the skills concerning computation of the probability of an event and the knowledge of what cards are in a 52-card deck. When the last skill is deleted from Problem 4, a new knowledge structure is obtained that contains 16 knowledge states. The BLIM is locally identifiable for the new structure, and its fit is good ( $p = .13$ ). However, it is worth noticing that this solution is not plausible from a theoretical point of view, because it assumes that knowing the 52-card deck is not necessary for solving Problem 4.

On the other hand, when we do not want to modify the knowledge structure, we can solve the identification problem by removing the parameter  $\beta_4$  from the model through the BLIMIT function.

The second data set consists of the responses that 67 students provide to a collection of 13 problems. An analysis of the content of the problems reveals that four skills (conditional probability, law of total probability, probability of the complement of an event, stochastic independence) were required to solve them. Via the competency model (Doignon & Falmagne, 1999), each problem was associated with its competencies according to the skill multimap represented in Table 2. All problems were associated with one competency, except for problem 13 which was associated with two competencies. Problem 13 reads: “Given two independent events  $A$  and  $B$  in a sample space  $S$ , the following probability is known:  $P(A | B) = .2$ . Find  $P(\bar{A})$ .” The first competency specifies that the problem requires the skills concerning probability of the complement of an event, conditional probability and stochastic independence to be solved. The second competency constitutes an alternative solution strategy specifying that Problem 13 could also be solved by simply subtracting the single probability value given in the text from 1. The knowledge structure  $\mathcal{K}_2$  delineated by the skill multimap contains 30 knowledge states.

$$\mathcal{K}_2 = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{13\}, \{1, 13\}, \{2, 13\}, \{3, 13\}, \{2, 13\}, \{1, 2, 5\}, \{1, 3, 6\}, \{1, 4, 7\}, \{2, 3, 8\}, \{2, 4, 9\}, \{3, 4, 10\}, \{1, 2, 5, 13\}, \{1, 3, 6, 13\}, \{1, 4, 7, 13\}, \{2, 3, 8, 13\}, \{2, 4, 9, 13\}, \{3, 4, 10, 13\}, \{2, 3, 4, 8, 9, 10\}, \{1, 2, 3, 5, 6, 8, 11\}, \{1, 2, 4, 5, 7, 9, 12\}, \{1, 3, 4, 6, 7, 10, 13\}, \{2, 3, 4, 8, 9, 10, 13\}, \{1, 2, 3, 5, 6, 8, 11, 13\}, \{1, 2, 4, 5, 7, 9, 12, 13\}, Q\}$$

The fit of the BLIM for  $\mathcal{K}_2$  turned out to be good ( $p = .15$ ). Maximum-likelihood estimates of model parameters were passed to the BLIMIT function, along with the knowledge structure  $\mathcal{K}_2$ . The model is not locally identifiable, and there is one second-order tradeoff dimension involving the  $\eta$  parameters of Problems 1, 2, 3, and 13 and some knowledge state probability.

As in the previous example, we can solve the identification problem by modifying the knowledge structure or by excluding a parameter from the model. With respect to the first solution, we can modify the skill assignment of Problem 13 which was involved in the tradeoff to a greater extent than Problems 1, 2, and 3. When the alternative solution strategy is deleted from Problem 13, a new knowledge structure is obtained that contains 16 knowledge states.

**Table 1** Skill map from example 1

Problem	Skills	Problem	Skills
1	{pb}	11	{pb, un}
2	{un}	12	{cp, id}
3	{cp, id}	13	{pb, cp, un}
4	{pb, cd}	14	{cp}
5	{cp, id}	15	{pb, cp, un, id}
6	{id}	16	{pb, cp, id}
7	{pb, cp, un}	17	{pb, cp}
8	{un, id}	18	{pb, un, id}
9	{pb, un, id}	19	{pb, un}
10	{pb, id}		

cd, card deck; cp, complement of an event; id, stochastic independence; pb, probability of an event; un, union of events

The BLIM is locally identifiable for the new structure, and its fit is good ( $p = .14$ ).

With respect to the second solution, we can remove the parameter  $\eta_{13}$  from the model by using the BLIMIT function.

## Conclusions

Identifiability is a general and fundamental issue in parametric modeling of the data. A parametric model is said to be identifiable if the parameter set giving rise to a certain prediction is uniquely determined. When more than a single parameter set gives rise to the same model predictions, then the model is said to be nonidentifiable. In this case the model parameters cannot be interpreted even when the model displays a good or even excellent fit to the data. As a consequence, any explanatory conclusions drawn from an analysis of the model parameters may be misleading.

The focus of this article was on identifiability of probabilistic knowledge structures, that is knowledge structures equipped with a probability distribution on the knowledge

**Table 2** Skill map from example 2

Problem	Competencies	Problem	Competencies
1	{cp}	8	{tt, cd}
2	{tt}	9	{tt, id}
3	{cd}	10	{cd, id}
4	{id}	11	{cp, tt, cd}
5	{cp, tt}	12	{cp, tt, id}
6	{cp, cd}	13	{cp, cd, id} {as}
7	{cp, id}		

as, alternative strategy; cd, conditional probability; cp, complement of an event; id, stochastic independence; tt, total probability

states. The BLIM is a special case of a probabilistic knowledge structure where the responses to the items are locally independent given the knowledge state of an individual. In the BLIM the conditional probabilities of the response patterns, given the knowledge states depend on two types of parameters of the items: a lucky guess probability  $\eta$  and a careless error probability  $\beta$ .

Identifiability of the BLIM depends on the particular knowledge structure on which it is defined. To the best of our knowledge, the set theoretical properties of those knowledge structures for which the BLIM turns out to be identifiable are far from clear. Indeed, a theoretical characterization of these kinds of knowledge structures is still missing. For this reason any method or procedure that aims at assessing identifiability of the BLIM which is general enough to be applied to any finite knowledge structure, although of moderate size, would be welcome. This was precisely the purpose of the present article.

Based on the work of Bamber and van Santen (1985, 2000), a method and a corresponding computerized procedure for assessing local identifiability of the BLIM have been proposed and described. The method operates on the Jacobian matrix of the model's prediction function. By this method it is possible to systematically detect those parameters of a nonidentifiable model that are involved in tradeoffs. The theory behind the proposed method sheds light on the existence of three different types of tradeoff dimensions involving, respectively, parameters of the same type, of two or three different types. This paves the way to future investigations aimed at systematic classifications of the types of nonidentifiability that arise with probabilistic knowledge structures in general and with the BLIM in particular.

A MATLAB procedure, called BLIMIT, which implements the proposed identification assessment method was described in a number of example applications. The input to the procedure is a binary matrix representing the knowledge structure for which the BLIM identification has to be tested, and the output is a text report with a main response saying whether the model is identifiable, followed by detailed diagnostic information at item level. In particular, information at item level allows detection of those specific item parameters that are involved in tradeoff dimensions. This information is particularly useful for modifying or correcting the knowledge structure and turning it into an identifiable one.

It is our hope that the proposed method and procedure might be useful to researchers that are interested in a systematic study of the identifiability of probabilistic knowledge structures, as well as to practitioners that, in applying such types of models would like to ascertain identifiability of the particular knowledge structure they intend to use.

## References

- Bamber, D., & van Santen, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, *29*, 443–473.
- Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, *44*, 20–40.
- Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man–Machine Studies*, *23*, 175–196.
- Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. New York, NY: Springer.
- Falmagne, J.-C., & Doignon, J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, *41*, 1–23.
- Falmagne, J.-C., & Doignon, J.-P. (1988b). A Markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology*, *32*, 232–258.
- Falmagne, J.-C., & Doignon, J.-P. (2011). *Learning spaces*. New York, NY: Springer.
- Heller, J., & Repitsch, C. (2011). Exploiting prior information in stochastic knowledge assessment. *Methodology* Advance online publication. doi:10.1027/1614-2241/a000035
- Heller, J., & Wickelmaier, F. (2011). *Parameter estimation in probabilistic knowledge structures*. Manuscript in preparation.
- Meyer, C. D. (2000). *Matrix analysis and applied linear algebra*. Philadelphia, PA: SIAM.
- Schrepp, M. (2005). About the connection between knowledge structures and latent class models. *Methodology*, *1*, 93–103.
- Smith, P. L. (1998). Attention and luminance detection: A quantitative analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 105–133. doi:10.1037/0096-1523.24.1.105
- Stefanutti, L., & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74*, 83–96.
- Ünlü, A. (2011). A note on the connection between knowledge structures and latent class models. *Methodology*, *7*, 63–67.