Chapter 1

Probabilistic Knowledge Space Theory

In 1985, Jean-Paul Doignon and Jean-Claude Falmagne introduced *Knowledge Space Theory* (briefly, KST; Doignon & Falmagne, 1985). For a detailed treatise on KST refer to the monograph Knowledge Spaces by Doignon & Falmagne (1999). Reviews of KST are provided by Falmagne (1989b) and Falmagne et al.[1] (1990). Further references on KST can be obtained at `http://wundt.uni-graz.at/`; see also the references listed in Remarks, **4** at the end of Section 1.2.

This chapter reviews some of the basic concepts of KST. The main concern is to give an introduction to general (!) probabilistic knowledge structures. Special emphasis is put on the modeling process as formal methods are introduced to model (**unobservable**) psychological assumptions in the light of (**observable**) heuristics, i.e.[2] data. In Section 1.1, we review some of the basic **deterministic** concepts of KST relevant for the discussion of general probabilistic knowledge structures. This includes *knowledge structures*, *(quasi ordinal) knowledge spaces*, and *surmise relations*. Their interpretations within the educational context are sketched briefly. Section 1.1 ends with the *Birkhoff-Theorem* providing a linkage between quasi ordinal knowledge spaces and surmise relations on a fixed item set. Section 1.2 covers some of the basic **probabilistic** concepts of KST. *Probabilistic knowledge structures* are introduced, and, based on this, we review two general probabilistic models for response data, the *basic probabilistic model*, and the *basic local independence model*. The class of so-called *stochastic learning paths models* is mentioned only marginally. Brief remarks on the interpretation and dimensionality of the probabilistic models reviewed supplement Section 1.2.

[1] The abbreviation "et al." stands for "and others (et alii)".
[2] The abbreviation "i.e." stands for "that is (id est)".

1.1 Knowledge Structures and Spaces, and Surmise Relations

Section 1.1 covers some of the basic **deterministic** concepts of KST. We only focus on those concepts that are relevant for the discussion of general probabilistic knowledge structures.

Definition 1.1. A *knowledge structure* is a pair $(Q, \mathcal{K})$ with Q a non-empty set, and $\mathcal{K}$ a family of subsets of Q containing at least the empty set $\emptyset$ and Q. The set Q is called the *domain* of the knowledge structure. The elements $q \in Q$, resp.[3] $K \in \mathcal{K}$, are referred to as *(test) items*, resp. *knowledge states*. With abuse of terminology, we also say *"$\mathcal{K}$ is a knowledge structure on Q"*.

The set Q is supposed to be a set of *dichotomous* (also, *bi-valued*) items, i.e., items with two possible response categories. For instance, Q could be a set of questions/problems in an examination on Euclidean geometry each of which a subject can either solve or fail to solve; or, Q could represent a set of statements in a questionnaire with each of which a subject can either agree or disagree. Throughout, in this dissertation, we use the first interpretation of Q as a set of bi-valued questions/problems that can either be *solved* (coded as 1) or *not solved* (coded as 0). Here, "solved" and "not solved" stand for the **observed** responses of a subject (the *manifest level*). This has to be distinguished from a subject's actual **unobservable** knowledge of the solution to an item (the *latent level*). In the latter case, if a subject actually (latently) knows resp. does not know the solution to the item, we say that the subject *masters* (also coded as 1) resp. *does not master* (also coded as 0) the item.

For a set X, let 2^X denote its *power-set*, i.e., the set of all subsets of X. Let further $|X|$ stand for the *cardinality* of X. The actually observed responses of a subject to the items in Q can be represented by the subset $R \in 2^Q$ of Q, containing exactly the items in Q that are correctly solved by the subject, i.e., he/she solves the items in R, and fails to solve the items in the *complement* $Q \setminus R := \{q \in Q : q \notin R\}$ of R. This subset, R, is called the *response pattern* of the subject. In particular, the power-set 2^Q constitutes the set of all possible response patterns. Thus, there are $|2^Q| = 2^{|Q|}$ possible response patterns.

Similarly, the actual latent state of knowledge of a subject with respect to the items in Q is represented by the subset $K \subset Q$, containing exactly the items in Q the subject is capable of mastering. This subset, K, is called the *knowledge state* of the subject. Given a knowledge structure $(Q, \mathcal{K})$, it is assumed that the only states of knowledge possible are the ones in $\mathcal{K}$. In this sense, $\mathcal{K}$ captures the organization of the knowledge ($\mathcal{K}$, a *knowledge structure*). In particular, in

[3]The abbreviation "resp." stands for "respectively".

contrast to the response patterns with 2^Q,[4] not all subsets are assumed to be states of knowledge, and the set of all possible states of knowledge is given by $\mathcal{K}$. Thus, there are $|\mathcal{K}|$ possible knowledge states in the population of reference.

In its practical application in education, a knowledge structure is interpreted as follows. Given a certain field of knowledge, say, Euclidean geometry, the set Q represents a (finite) collection of problems or questions appropriate enough to give a fine-grained and representative coverage of Euclidean geometry. Then, given a population of reference, $\mathcal{K}$ represents the collection of all those and exactly those states of knowledge that are actually (latently) inherent in units of the population with respect to the items in Q. In other words, idealized, these and only these knowledge states $K \in \mathcal{K}$ could be observed as the response patterns if no response errors (i.e., careless error and lucky guess) would occur.

Knowledge spaces are special knowledge structures.

Definition 1.2. A knowledge structure $(Q, \mathcal{K})$ is called a *knowledge space* iff[5] $\mathcal{K}$ is closed under arbitrary unions—i.e., for all $\mathcal{F} \subset \mathcal{K}$, $\bigcup \mathcal{F} \in \mathcal{K}$. If a knowledge space $(Q, \mathcal{K})$ is (additionally) closed under arbitrary intersections—i.e., for all $\mathcal{F} \subset \mathcal{K}$, $\bigcap \mathcal{F} \in \mathcal{K}$—, then it is called a *quasi ordinal* knowledge space.

Definition 1.1 and 1.2 are *"persons-related"*, i.e., they concern with subjects' latent states of knowledge ($K \in \mathcal{K}$). There is another formulation, the *surmise relation*,[6] which is *"items-related"*, concerning with dependencies between the individual items ($I \in Q$). Quasi ordinal knowledge spaces (*persons-level*) and surmise relations (*items-level*) are linked by the *Birkhoff-Theorem* (Theorem 1.4).

In the sequel, these notions are provided.

Definition 1.3. Let Q be a non-empty set of items. A *surmise relation* on Q is any *quasi order* on Q, i.e., any reflexive and transitive binary relation on Q.

In its practical application in education, a surmise relation (on Q) may model latent hierarchies between the items (in Q) based on unobservable dependencies of the type: *A subject, (latently) capable of mastering test item $J \in Q$, is also (latently) capable of mastering test item $I \in Q$.* This is modeled by a surmise relation $\leq$ on Q with $I \leq J$, and any pair in $\leq$ is interpreted in this way.

Now, we come to the *Birkhoff-Theorem* (in its application in KST).

Theorem 1.4 (Birkhoff, 1937). *Let Q be a non-empty set of items. There is a one-to-one correspondence between the family of all quasi ordinal knowledge spaces $\mathcal{K}$ on Q, and the family of all quasi orders $\leq$ on Q. This correspondence*

[4] In general, any subset $R \in 2^Q$ may be "generated" as a response pattern by the knowledge states in $\mathcal{K}$ if *response errors* (i.e., *careless error* and *lucky guess*) are committed.

[5] The abbreviation "iff" stands for "if and only if".

[6] The notion of a *surmise relation* is introduced in Definition 1.3.

is defined through the equivalences ($p, q \in Q$, resp. $K \subset Q$)

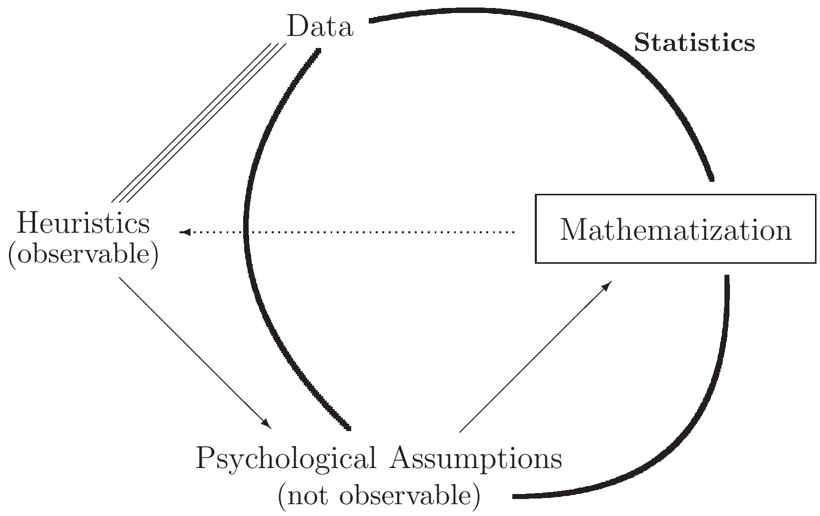
$$\begin{aligned} p \leq q \quad &:\Longleftrightarrow \quad \forall K \in \mathcal{K} : [q \in K \Longrightarrow p \in K], \\ K \in \mathcal{K} \quad &:\Longleftrightarrow \quad \forall (p \leq q) : [q \in K \Longrightarrow p \in K]. \end{aligned}$$

Proof. See Doignon & Falmagne (1999), pp.[7] 39-40. □

1.2 Probabilistic Knowledge Structures

Section 1.2 reviews basic *probabilistic* concepts of KST. This is done especially emphasizing the process of modeling psychological assumptions by mathematical methods. For this purpose, the scheme in Figure 1 is pursued.[8]

Figure 1
Heuristics & Data, Psychological Assumptions, and Mathematization



1.2.1 Why Probabilities ?—Standard Example

As an example, we consider the knowledge structure

$$\mathcal{H} := \Big\{\emptyset, \{a\}, \{b\}, \{a,b\}, \{a,b,c\}, \{a,b,d\}, \{a,b,c,d\}, \{a,b,c,e\}, Q\Big\}$$

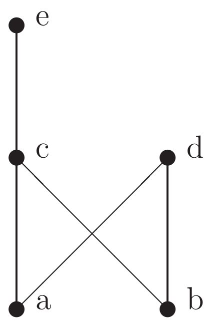
[7]The abbreviation "pp." stands for "pages".
[8]It is a *rough* scheme and should be understood for *intuitive illustration* only.

on the domain $Q := \{a, b, c, d, e\}$ (briefly, the *standard example*).

This is a quasi ordinal knowledge space, i.e., it is closed under arbitrary unions and intersections (cp.[9] Definition 1.2). The surmise relation, $\leq_{\mathcal{H}}$, derived from this space according to the Birkhoff-Theorem (Theorem 1.4) is even a *partial order*, i.e., an *antisymmetric* quasi order. Its *Hasse diagram* is given in Figure 2.

Figure 2

Hasse Diagram of $(Q, \leq_{\mathcal{H}})$



Let P be a population of reference.

Heuristics

We observe the following *heuristics*.

H1 A subject (randomly) sampled from P can have a response pattern $R \subset Q$ with $R \notin \mathcal{H}$.

H2 A subject sampled from P can solve an item $J \in Q$ but fail to solve an item $I \in Q$ with $I \leq_{\mathcal{H}} J$.

H3 In series of trials, there are varying relative frequencies of response patterns.

Psychological Assumptions

The following *psychological assumption* is at the heart of probabilistic KST.

A1 Every subject of P is always in one of the latent states of knowledge $H \in \mathcal{H}$. These states occur with different proportions in P.

Assuming **A1**, however, is in contradiction with the heuristics in **H1-H3**. Thus, under the assumption **A1**, we are forced to explain these defects. An explanation is given by the (unobservable) psychological assumption **A2**.

[9]The abbreviation "cp." stands for "compare".

A2 *Response errors*, i.e., *careless errors* (CEs) and *lucky guesses* (LGs), occur while subjects respond to items. These effects are random, making all kinds of response patterns possible a-priori.

According to the assumption **A2** and the heuristics **H3**, it is plausible to make the further two assumptions **A3** and **A4**.

A3 Response patterns occur randomly with certain likelihoods.

A4 Even when conditioning on a knowledge state $H \in \mathcal{H}$ (i.e., given $H \in \mathcal{H}$), response patterns occur randomly with certain likelihoods.

Mathematization

The assumptions **A1** and **A4** are mathematized in **M1** resp. **M4**.

M1 $p : \mathcal{H} \longrightarrow [0,1]$, $H \mapsto p(H)$ with

$$\begin{aligned} p(H) &\geq 0 \quad \text{(proportion of } H \text{ in } P\text{)}, \\ \sum_{H \in \mathcal{H}} p(H) &= 1 \quad \text{(every } P_k \in P \text{ in one } H \in \mathcal{H}\text{)}. \end{aligned}$$

M4 $r : 2^Q \times \mathcal{H} \longrightarrow [0,1]$, $(R,H) \mapsto r(R,H)$ with

$$\begin{aligned} r(R,H) &\geq 0 \quad \text{(likelihood for } R \text{ given } H\text{)}, \\ \sum_{R \in 2^Q} r(R,H) &= 1 \quad \text{for every } H \in \mathcal{H} \quad \text{(sure event)}. \end{aligned}$$

The mathematization, **M3**, of **A3** is a corollary of **M1** and **M4**.

M3 Let $\rho : 2^Q \longrightarrow \mathbb{R}$, $R \mapsto \rho(R) := \sum_{H \in \mathcal{H}} r(R,H)p(H)$. Then:[10]

$$\begin{aligned} \rho : 2^Q &\longrightarrow [0,1], \\ \sum_{R \in 2^Q} \rho(R) &= 1. \end{aligned}$$

Thus, $\rho(R) \in [0,1]$ measures the (unconditional) likelihood for $R \in 2^Q$.

Before we proceed with mathematizing **A2**, let us consider an example.

Example

Let $(Q, \mathcal{H})$ be the standard example. What is the probability of response pattern $R := \{c, d\}$, i.e., $\rho(\{c, d\})$, given the fact that subjects never respond correctly to questions not in their knowledge states ?

[10]For a proof, see Corollary 1.7.

(Hint: In this case, $r(\{c,d\},H) = 0$ for all $H \in \mathcal{H}$ with $\{c,d\} \not\subset H$.)
It holds:

$$\begin{aligned}
\rho(\{c,d\}) &= \sum_{H \in \mathcal{H}} r(\{c,d\},H)p(H) \\
&\stackrel{(i)}{=} \sum_{H \in \mathcal{H}, \{c,d\} \subset H} r(\{c,d\},H)p(H) \\
&= \sum_{H \in \{\{a,b,c,d\},Q\}} r(\{c,d\},H)p(H) \\
&= r(\{c,d\},\{a,b,c,d\})p(\{a,b,c,d\}) + r(\{c,d\},Q)p(Q).
\end{aligned}$$

Ad[11] *(i).* For all $H \in \mathcal{H}$ with $\{c,d\} \not\subset H$, $r(\{c,d\},H) = 0$. □

Let us return to the problem of mathematizing assumption **A2**. In order to mathematize **A2**, we have to concretize the assumption **A2** more clearly. This is done by separating the vague assumption **A2** into three more concrete psychological assumptions **A2.1**, **A2.2**, and **A2.3**.

A2.1 Fixing the latent state of knowledge of a subject **while she is responding to the items**, her responses to the items are independent.[12]

A2.2 The conditional likelihood for not solving an item contained in a subject's latent state of knowledge, i.e., committing a *careless error*, is the same for all knowledge states containing the item, but it varies from item to item. In other words, the likelihoods for a careless error are attached to the items, and they do not vary with the knowledge states containing the items.

A2.3 The conditional likelihood for solving an item not contained in a subject's latent state of knowledge, i.e., committing a *lucky guess*, is the same for all knowledge states not containing the item, but it varies from item to item. In other words, the likelihoods for a lucky guess are attached to the items, and they do not vary with the knowledge states not containing the items.

LI (i.e., **A2.1**) can be mathematized as follows.

M2.1 Let $R \in 2^Q$ and $H \in \mathcal{H}$ (standard example). Let $\mathbf{x} := (x_a, x_b, \ldots, x_e)$ with $x_q \in \{0,1\}$ ($q \in Q$) be the vector of item scores corresponding to R. Further, let $r(R,H)$ be the conditional probability for R given H (cp. **M4**), and let $r_{qx_q|H}$ denote the conditional probability for answering x_q to item $q \in Q$ given H. Then, LI means that

$$r(R,H) = \prod_{q \in Q} r_{qx_q|H}.$$

[11] The Latin word "ad" stands for "in regards to".

[12] This is the assumption of *local* (or, *conditional*) *independence*, LI, central to latent class and latent trait models. Note: The assumption of local independence requires the latent "ability/ies" of a person to remain the same during testing ! It is a strong assumption !

From LI, we obtain (for $X, Y \subset Q$, $X \setminus Y := \{q \in Q : q \in X, q \notin Y\}$)

$$
\begin{aligned}
r(R,H) &= \prod_{q \in Q} r_{qx_q|H} \\
&= \left[\prod_{q \in H\setminus R} r_{q0|H}\right] \cdot \left[\prod_{q \in H\cap R} r_{q1|H}\right] \cdot \left[\prod_{q \in R\setminus H} r_{q1|H}\right] \cdot \left[\prod_{q \in Q\setminus(R\cup H)} r_{q0|H}\right].
\end{aligned}
$$

The assumptions **A2.2** and **A2.3** are mathematized as follows.

M2.2 $\beta : Q \longrightarrow [0,1[,\ q \mapsto \beta(q) =: \beta_q$ with

$$
\begin{aligned}
\beta_q &\in [0,1[\ \text{ (careless error probability at } q\text{)}, \\
r_{q0|H} &= \beta_q \ \text{ for all } H \in \mathcal{H} \text{ with } q \in H.
\end{aligned}
$$

M2.3 $\eta : Q \longrightarrow [0,1[,\ q \mapsto \eta(q) =: \eta_q$ with

$$
\begin{aligned}
\eta_q &\in [0,1[\ \text{ (lucky guess probability at } q\text{)}, \\
r_{q1|H} &= \eta_q \ \text{ for all } H \in \mathcal{H} \text{ with } q \notin H.
\end{aligned}
$$

Since $r_{q0|H} + r_{q1|H} = 1$ for all $q \in Q$ and $H \in \mathcal{H}$, it holds

$$
\begin{aligned}
r_{q1|H} &= 1 - \beta_q \ \text{ for all } H \in \mathcal{H} \text{ with } q \in H, \\
r_{q0|H} &= 1 - \eta_q \ \text{ for all } H \in \mathcal{H} \text{ with } q \notin H.
\end{aligned}
$$

Taking all the previous considerations into account, we arrive at the following expression for $r(R,H)$ (as a consequence of **A2.1-A2.3**):

$$
\begin{aligned}
r(R,H) &\overset{2.1}{=} \prod_{q \in Q} r_{qx_q|H} \\
&= \left[\prod_{q \in H\setminus R} r_{q0|H}\right] \cdot \left[\prod_{q \in H\cap R} r_{q1|H}\right] \cdot \left[\prod_{q \in R\setminus H} r_{q1|H}\right] \cdot \left[\prod_{q \in Q\setminus(R\cup H)} r_{q0|H}\right] \\
&\overset{2.2/3}{=} \left[\prod_{q \in H\setminus R} \beta_q\right] \cdot \left[\prod_{q \in H\cap R} (1-\beta_q)\right] \cdot \left[\prod_{q \in R\setminus H} \eta_q\right] \cdot \left[\prod_{q \in Q\setminus(R\cup H)} (1-\eta_q)\right].
\end{aligned}
$$

Finally, the manifest distribution $\rho(R) = \sum_{H \in \mathcal{H}} r(R,H)p(H)$ fulfills ($R \in 2^Q$)

$$
\begin{aligned}
\rho(R) = \sum_{H \in \mathcal{H}} \Bigg\{ &\left[\prod_{q \in H\setminus R} \beta_q\right] \cdot \left[\prod_{q \in H\cap R} (1-\beta_q)\right] \\
&\cdot \left[\prod_{q \in R\setminus H} \eta_q\right] \cdot \left[\prod_{q \in Q\setminus(R\cup H)} (1-\eta_q)\right] \Bigg\} p(H).
\end{aligned}
$$