On Query Procedures to Build Knowledge Structures

Ivo Düntsch[1]
School of Information and Software Engineering
University of Ulster
Newtownabbey, BT 37 0QB, N.Ireland
I.Duentsch@ulst.ac.uk

Günther Gediga[1]
FB Psychologie / Methodenlehre
Universität Osnabrück
49069 Osnabrück, Germany
gg@Luce.Psycho.Uni-Osnabrueck.DE

January 10, 1996

[1]The ordering of authors is alphabetical, and equal authorship is implied.

Abstract

Let Q be a finite set of problems. A *knowledge state* is the set of problems a person is capable of solving. A family K of knowledge states is called a *knowledge structure*, if the empty set and Q are elements of K. When K is closed under union, the structure is called a *knowledge space*. There exist effective algorithms to generate a knowledge space by querying an expert using an entail relation.

In this note we show that more general knowledge structures can be generated using modified query procedures. The general theory of the Galois connections between entail relations and knowledge structures is explored. Finally, we present query procedures which can be applied to generate any knowledge structure.

Key words: knowledge structure, entail relation, Galois connection, ability testing, dichotomous data

1 Introduction

The aim of modern scaling theory of knowledge representation is a structural description of knowledge states of subjects, given a test procedure which consists of a finite set Q of posed problems. For each subject, the result of such a procedure is the set of problems which the subject has solved. The theoretical counterpart to such a set of solved problems is called a *knowledge state.*

A *knowledge structure* a collection of knowledge states, i.e. an element of 2^{2^Q}. A knowledge structure serves as a model of all possible states which subjects may have, if we test their knowledge with a procedure based on the problem set Q.

A knowledge structure which is closed under set–theoretic union is called a *knowledge space* (Falmagne et al. (1990)).

There exist effective algorithms to generate a knowledge space by querying an expert using an entail relation as shown in Koppen & Doignon (1990). In this note we show that more general knowledge structures can be generated using modified query procedures. The general theory of the Galois connections between entail relations and knowledge structures is explored. Finally, we present query procedures which can be applied to generate any knowledge structure

We shall suppose that knowledge structures contain $\emptyset$, (i.e. no problem can be solved), and Q (i.e. all problems can be solved), and denote the set of all knowledge structures on Q by $\mathcal{K}$.

With inclusion as natural order, $\mathcal{K}$ is a complete and atomic sublattice of $\langle 2^{2^Q}, \cap, \cup \rangle$ (in fact, a Boolean algebra) with smallest element $\{\emptyset, Q\}$ and largest element 2^Q. If R is a binary relation on X, and $x \in X$, then $ran_R x = \{y \in X : xRy\}$. If $\sigma : X \to Y$ is a mapping and $A \in X$, we sometimes denote the image of A under σ by A^σ.

In order to uncover the possible knowledge states of a group of subjects one puts to an expert questions relating one group of problems to another. Among the questions that can be asked are

1. Is problem p a prerequisite for problem q?
2. If no problem of A can be solved by some subject, can we assume that no problem of B can be solved by the same subject?
3. If all problems of A can be solved by some subject, can we assume that all problems of B can also be solved by the same subject?

In this way, binary relations on the powerset 2^Q of Q can be defined which will lead us to an approximation of the set of potential knowledge states: The first component A contains the input of a query, the second component B the outcome, and $\langle A, B \rangle$ is in the relation iff the expert answers *Yes*. In some cases it will be possible to completely determine the knowledge structure. Depending on what questions are asked of the expert, the resulting relation will have some built in properties: Suppose that $\langle A, B\rangle \in R$ iff solving all problems of A entails solving all problems of B, in other words, that the answer to question 3 above is *Yes*. Then, R clearly contains the converse of set inclusion, i.e. if $A \supseteq B$, then $\langle A, B\rangle \in R$. A moment's reflection tells us that R is transitive as well. Similarly, if $\langle A, B\rangle \in T$ iff solving no problem of A entails not being able to solve any problem in B, then T includes $\supseteq$ and T is transitive. Thus, R and T are quasi orders on 2^Q. It also seems sensible to present the expert only with non empty sets of problems, and we shall assume that the second components are non empty as well. This does not restrict the generality of our outlook. We regard each query procedure as a binary relation on the set $(2^Q)^+ = 2^Q \setminus \{\emptyset\}$, and we denote the set of all these relations by $\mathcal{P}$. Observe that $\mathcal{P}$ is a complete and atomic Boolean set algebra.

If an empirical situation lets us observe a certain set K of knowledge states, free from noise such as careless errors or lucky guesses, and if we furthermore assume that the questioning of the expert has led to a set K' of possible states, then it seems safe to assume that $K \subseteq K'$. In other words, refining the query procedure – or extending the resulting relation on 2^Q –, will produce some K'' such that $K \subseteq K'' \subseteq K'$. Thus, in associating a query procedure with a set of states by a function $\varphi : \mathcal{P} \to \mathcal{K}$ we shall require that φ is antitone, i.e.

$$\text{For all } R,\ S \in \mathcal{P},\ R \subseteq S \text{ implies } S^{\varphi} \subseteq R^{\varphi}.$$

Conversely, in associating a set of states K with a query relation R by a function $\psi : \mathcal{K} \to \mathcal{P}$ we shall require that ψ is antitone as well. Finally, to relate the two connections we shall suppose that

$$R \subseteq R^{\varphi\psi} \text{ and } K \subseteq K^{\psi\varphi},$$

in other words, that the pair $\langle \varphi, \psi\rangle$ forms a Galois connection between the ordered sets $\langle \mathcal{K}, \subseteq\rangle$ and $\langle \mathcal{P}, \subseteq\rangle$: If $\langle X, \leq\rangle$ and $\langle Y, \leq\rangle$ are partially ordered sets, a pair $\langle \psi, \varphi\rangle$ is called a *Galois connection* between X and Y, if $\psi : X \to Y$ and $\varphi : Y \to X$ are antitone (i.e. dually order preserving) mappings, and $x \leq x^{\psi\varphi}$, $y \leq y^{\varphi\psi}$ for all $x \in X$, $y \in Y$. $x \in X$ is called *Galois closed* with respect to $\langle \psi, \varphi\rangle$ if $x = x^{\psi\varphi}$. If the choice of $\langle \psi, \varphi\rangle$ is clear from the context, we just speak of closed sets.

We shall later need the following properties of Galois connections which can be found in Erné (1982):

Lemma 1.1. *Let $\langle X, \subseteq\rangle$ and $\langle Y, \subseteq\rangle$ be complete sublattices of the powerset of some set, and $\varphi : X \to Y$ be a mapping. Then, the following are equivalent:*

1. *There is a (unique) mapping $\psi : Y \to X$ such that $\langle \varphi, \psi\rangle$ is a Galois connection.*

2. φ is antitone and for every $y \in Y$ there is a largest $x \in X$ with $y \subseteq \varphi(x)$.

3. For every $Z \subseteq X$, $\varphi(\bigcup Z) = \bigcap\{\varphi(x) : x \in Z\}$. □

The approach to knowledge structures via Galois connections was taken by Monjardet (1970) and by Koppen & Doignon (1990). They call a relation $R \in \mathcal{P}$ an *entail relation* on Q if

1. R is a quasiorder containing $\supseteq$,

2. For each non empty $A \subseteq Q$, $ran_R A$ has a greatest element.

It may be interesting to note the entail relations are the dependency relations which are used in theory of rough sets, and also correspond to the keys for a relational system in Garey & Johnson (1979). In particular, there is an intimate connection between entail relations on Q and congruences on the semilattice $\langle 2^Q, \cup\rangle$, see Novotný (1997) and the references therein for details.

In order to capture knowledge spaces by a suitable query procedure, Koppen & Doignon (1990) define functions $\varphi : \mathcal{P} \to \mathcal{K}$ by

$$R \overset{\varphi}{\mapsto} \{C \subseteq Q : (\forall \langle A, B\rangle \in R)(A \cap C = \emptyset \Rightarrow B \cap C = \emptyset)\}$$

and $\psi : \mathcal{K} \to \mathcal{P}$ by

$$K \overset{\psi}{\mapsto} \{\langle A, B\rangle \in (2^Q)^+ \times (2^Q)^+ : (\forall C \in K)(A \cap C = \emptyset \Rightarrow B \cap C = \emptyset)\},$$

(cf. the relation T on page 1) and show [1] that

Proposition 1.2. *Koppen & Doignon (1990)*

1. The pair $\langle\varphi, \psi\rangle$ forms a Galois connection between the ordered sets $\langle\mathcal{P}, \subseteq\rangle$ and $\langle\mathcal{K}, \subseteq\rangle$.

2. With respect to $\langle\varphi, \psi\rangle$, the closed elements of $\langle\mathcal{K}, \subseteq\rangle$ are the knowledge spaces, and the closed elements of $\langle\mathcal{P}, \subseteq\rangle$ are the entail relations. □

For unexplained notation and definitions in lattice theory the reader is invited to consult Birkhoff (1967) or Grätzer (1978).

[1] We have slightly modified the construction in Koppen & Doignon (1990) in order to better handle the presence (or absence) of $\emptyset$.

2 Galois connections which produce entail relations

In this section we shall characterize those Galois connections between $\langle \mathcal{P}, \subseteq \rangle$ and $\langle \mathcal{K}, \subseteq \rangle$ in which the Galois closed structures of $\langle \mathcal{P}, \subseteq \rangle$ are the entail relations. From this description we obtain an internal characterization of the Galois closure of knowledge structures.

We denote by $\mathcal{E}$ the set of all entail relations on Q, and set $\mathbf{1} = (2^Q)^+ \times (2^Q)^+$, $\mathbf{0} = \supseteq \cap \mathbf{1}$. A relation $R \in \mathcal{E}$ is called *proper*, if $R \neq \mathbf{1}$. Also, for $D \in (2^Q)^+$, set

$$R_D = \{\langle A, B \rangle \in \mathbf{1} : A \subseteq D \Rightarrow B \subseteq D\}.$$

Note that $R_Q = \mathbf{1}$.

Proposition 2.1. *1. $\langle \mathcal{E}, \subseteq \rangle$ is a complete dually atomistic lattice in which the meet operation is intersection.*

2. The dual atoms of $\mathcal{E}$ are the relations of the form R_D, $D \in (2^Q)^+$, $D \neq Q$.

Proof. 1. It follows from 1.2 that $\langle \mathcal{E}, \subseteq \rangle$ is dually isomorphic to the complete lattice of closure systems on Q, i.e. $\cap$–closed subsets of 2^Q, (see Birkhoff (1967), p 111), whence 1. follows.

2. It is straightforward to check that R_D is an entail relation. Note that for $A \in (2^Q)^+$

$$\max ran_{R_D} A = \begin{cases} D, & if\ A \subseteq D \\ Q, & otherwise, \end{cases}$$

a fact we shall use later on. Now, suppose that $S \in \mathcal{E}$ and $R_D \subseteq S$, $R_D \neq S$. Then, there is some $\langle A, B \rangle \in S$ such that $A \subseteq D$ and $B \not\subseteq D$. We shall show that $S = \mathbf{1}$: Let $\langle G, H \rangle \in \mathbf{1}$. Then, $\langle G, A \rangle \in R_D$, since $A \subseteq D$, and $\langle B, H \rangle \in R_D$, since $B \not\subseteq D$. $R_D \subseteq S$ and the transitivity of S now imply that $\langle G, H \rangle \in S$.

To show that every dual atom of $\mathcal{E}$ is of this form, it suffices to show that for each proper $R \in \mathcal{E}$ there is some $D \in (2^Q)^+$ with $D \neq Q$ and $R \subseteq R_D$. Suppose that $R \in \mathcal{E}$ such that for every $D \in (2^Q)^+, D \neq Q$ there is some $\langle A, B \rangle \in R$ with $A \subseteq D$, $B \not\subseteq D$. Let $C \in (2^Q)^+$, and set $M = max\ ran_R C$. If $M \neq Q$, there is some $\langle A, B \rangle \in R$ such that $A \subseteq M$ and $B \not\subseteq M$. $A \subseteq M$ implies that $\langle C, A \rangle \in R$, and from the transitivity of R we obtain $\langle C, B \rangle \in R$. But then $B \subseteq M$, a contradiction. It follows that $M = Q$, and hence $R = \mathbf{1}$. □

We now turn to Galois connections. Let

$$\alpha : \langle \mathcal{K}, \subseteq \rangle \to \langle \mathcal{P}, \subseteq \rangle \text{ and } \beta : \langle \mathcal{P}, \subseteq \rangle \to \langle \mathcal{K}, \subseteq \rangle$$

be antitone mappings such that $\langle \alpha, \beta \rangle$ is a Galois connection and $ran\ \alpha = \mathcal{E}$; such a Galois connection will be henceforth called an *entail connection*. Denote the set of all Galois closed elements of $\mathcal{K}$ by $\mathcal{K}^{\alpha\beta}$. For $C \subseteq Q$, $C \neq \emptyset$, let $K_C = \{\emptyset, C, Q\}$. Note that for $C \subseteq Q$ and $C \neq Q$, K_C is an atom of $\mathcal{K}$, and that each atom of $\mathcal{K}$ is of this form. If $K \in \mathcal{K}$, we set $K^+ = K \setminus \{\emptyset\}$.

Proposition 2.2. *1. Each atom of $\langle \mathcal{K}, \subseteq \rangle$ is Galois closed with respect to $\langle \alpha, \beta \rangle$.*

2. The mapping α is completely determined by its values on the atoms of $\mathcal{K}$.

Proof. 1. Since $\mathcal{K}^{\alpha\beta}$ is dually order isomorphic to $\mathcal{E}$, it is atomic with $2^{|Q|} - 2$ atoms. By Lemma 1.1 and the fact that both $\mathcal{P}$ and $\mathcal{K}$ are complete set lattices, we obtain that $\mathcal{K}^{\alpha\beta}$ is closed under intersections. Thus, if K and K' are atoms of $\mathcal{K}^{\alpha\beta}$, then $K \cap K' = \{\emptyset, Q\}$. Since $\mathcal{K}^{\alpha\beta}$ has $2^{|Q|} - 2$ atoms, it follows that the atoms of $\mathcal{K}^{\alpha\beta}$ are exactly the atoms of $\mathcal{K}$.

2. Let $K \in \mathcal{K}$. If $K = \{\emptyset, Q\}$ or $K = \bigcup \mathcal{K}$, then $K^\alpha = \mathbf{1}$ or $K^\alpha = \mathbf{0}$, irrespective of α. Otherwise, $K = \bigcup\{K_C : C \in K^+\}$, and

$$K^\alpha = \left(\bigcup\{K_C : C \in K^+\}\right)^\alpha = \bigcap\{K_C^\alpha : C \in K^+\}$$

by Lemma 1.1. □

Remark. Proposition 2.2.1 does not hold in general, even if both structures involved are complete atomistic lattices. A counterexample can be found in Andréka et al. (1995).

We can now describe all Galois connections $\langle \mathcal{P}, \subseteq \rangle$ and $\langle \mathcal{K}, \subseteq \rangle$ which produce entail relations. This is an instance of the more general construction of polarities induced by a binary relation as described in Birkhoff (1967).

Proposition 2.3. *Let f be a permutation of 2^Q with $f(\emptyset) = \emptyset$ and $f(Q) = Q$, and define $\alpha : \langle \mathcal{K}, \subseteq \rangle \to \langle \mathcal{P}, \subseteq \rangle$ by*

$$K \overset{\alpha}{\mapsto} \bigcap\{R_{f(C)} : C \in K^+\}.$$

Then, α can be uniquely extended to a Galois connection $\langle \alpha, \beta \rangle$ such that ran $\alpha = \mathcal{E}$, and

$$R^\beta = \{C \subseteq Q : (\forall \langle A, B \rangle \in R)(A \subseteq f(C) \Rightarrow B \subseteq f(C)\} = \{C \subseteq Q : R \subseteq R_{f(C)}\} \cup \{\emptyset\}.$$

Furthermore, each entail connection between $\langle \mathcal{P}, \subseteq \rangle$ and $\langle \mathcal{K}, \subseteq \rangle$ can be obtained this way.

Proof. If K_C is an atom of $\mathcal{K}$ then $K_C^\alpha = R_{f(C)}$. The images of the atoms of $\mathcal{K}$ are exactly the dual atoms of $\mathcal{E}$ by Proposition 2.1. Furthermore, 2.1. implies that *ran* $\alpha = \mathcal{E}$, and Lemma 1.1 tells us that α can be extended to a Galois connection $\langle \alpha, \beta \rangle$.

Suppose that $R \in \mathcal{P}$, and let

$$T = \{C \subseteq Q : (\forall \langle A, B \rangle \in R)(A \subseteq f(C) \Rightarrow B \subseteq f(C)\}.$$

Since $f(\emptyset) = \emptyset$ and $f(Q) = Q$, we have $\emptyset, Q \in T$, and thus $T \in \mathcal{K}$. By Lemma 1.1 we need to show that T is the largest $K \in \mathcal{K}$ for which $R \subseteq K^\alpha$. We first prove $R \subseteq T^\alpha$: If $T^\alpha = \mathbf{1}$, there is nothing to show. Otherwise, $T^\alpha = \bigcap\{R_{f(C)} : C \in T^+\}$. Let $C \in T^+$; then, for all $\langle A, B \rangle \in R$, $A \subseteq f(C)$ implies $B \subseteq f(C)$ by definition of T. Thus, $\langle A, B \rangle \in R_{f(C)}$ and it follows that $R \subseteq T^\alpha$.

Now, let $R \subseteq K^\alpha$. If $K^\alpha = \mathbf{1}$, then $K = \{\emptyset, Q\} \subseteq T$. Otherwise,

$$K = \bigcap\{R_{f(C)} : C \in K^+\}.$$

Suppose that $C \in K^+$. Then, $R \subseteq K^\alpha \subseteq R_{f(C)}$, and the definition of $R_{f(C)}$ imply $C \in T$. The other equation is just the definition of $R_{f(C)}$.

Conversely, suppose that $\langle \alpha, \beta \rangle$ is a Galois connection between $\langle \mathcal{K}, \subseteq \rangle$ and $\langle \mathcal{P}, \subseteq \rangle$ for which *ran* $\alpha = \mathcal{E}$. For each atom K_C of $\mathcal{K}$ there is a unique dual atom R_D of $\mathcal{K}$ such that $K_C^\alpha = R_D$ and vice versa. Setting $f(C) = D$, $f(\emptyset) = \emptyset$ and $f(Q) = Q$ gives us the desired result. □

Thus, with each entail connection $\langle \alpha, \beta \rangle$ we can associate a unique permutation f_α of 2^Q which leaves $\emptyset$ and Q fixed, and vice versa. We have the following generalization of 4.4. of Koppen & Doignon (1990):

Proposition 2.4. *Let $\langle \alpha, \beta \rangle$ be an entail connection with associated permutation f, and $R \in \mathcal{E}$. Then,*

$$D \in R^\beta \textit{ iff } (\forall C \in (2^Q)^+)(\langle f(D), C \rangle \in R \Rightarrow C \subseteq f(D)).$$

Proof. "⇒": Suppose that $D \in R^\beta$ and $\langle f(D), C \rangle \in R$. Since $D \in R^\beta$, $A \subseteq f(D)$ implies $B \subseteq f(D)$ for all $\langle A, B \rangle \in R$, in particular, $\langle f(D), C \rangle \in R$ implies $C \subseteq f(D)$.

"⇐": Suppose D satisfies the condition, and let $\langle A, B \rangle \in R$. If $A \subseteq f(D)$, then $\langle f(D), A \rangle \in R$, and thus $\langle f(D), B \rangle \in R$ by the transitivity of R. The condition on D now implies $B \subseteq f(D)$. □

To give an internal characterization of $K^{\alpha\beta}$ we first prove

Lemma 2.5. *Let $\mathcal{D} = \{D_i : i \leq k\}$ be a family of non empty proper subsets of Q, $R = \bigcap\{R_D : D \in \mathcal{D}\}$, and $C \subseteq Q, C \neq \emptyset$. Then, $R \subseteq R_C$ iff*

1. *$C = Q$ or*
2. *There are $D_0, ..., D_n \in \mathcal{D}$ such that $\emptyset \neq D_0 \cap ... \cap D_n = C$.*

Proof. "⇒": Suppose that $C \neq Q$, and let $J = \{i \leq k : C \not\subseteq D_i\}$, $I = \{i \leq k : C \subseteq D_i\}$. If $I = \emptyset$, then $\langle C, D \rangle \in R$ for all $D \in (2^Q)^+$; in particular, we have $\langle C, -C \rangle \in R$. On the other hand, $\langle C, -C \rangle \notin R_C$, contradicting that $R \subseteq R_C$. It follows that $\emptyset \neq C \subseteq \bigcap\{D_i : i \in I\}$. Assume that $q \in \bigcap\{D_i : i \in I\} \setminus C$, and set $D = C \cup \{q\}$. Then, $D \subseteq \bigcap\{D_i : i \in I\}$, and $\langle C, D \rangle \in R - R_C$, a contradiction.

"⇐": Suppose that $C \neq Q$ and that C fulfills the condition. Let $\langle A, B \rangle \in R$. Then, in particular, *max* $ran_R A \subseteq \bigcap\{R_{D_i} : i \leq n\} = C$, and hence $\langle A, B \rangle \in R_C$. □

Proposition 2.6. *Let $K \in \mathcal{K}$ be a proper knowledge structure and $\langle \alpha, \beta \rangle$ be an entail connection with associated permutation f. Then,*

$$K^{\alpha\beta} = \{C \subseteq Q : C = \emptyset \text{ or } f(C) = f(D_0) \cap f(D_1) \cap ... \cap f(D_n) \text{ for some } D_0, ..., D_n \in K^+\}.$$

Proof. Let $R = K^\alpha = \bigcap \{R_{f(C)} : C \in K^+\}$. Then, $R = \{C \subseteq Q : R \subseteq R_{f(C)}\} \cup \{\emptyset\}$ by Proposition 2.3. Suppose that $C \neq \emptyset, Q$. Then, by the previous Lemma,

$$C \in R^\beta \text{ iff there are } D_0, ..., D_n \in K^+ \text{ such that } f(C) = \bigcap \{f(D_i) : i \leq n\},$$

which proves the claim. □

Observe that this generalizes the result of Koppen & Doignon (1990) that their Galois closed knowledge structures are exactly the knowledge spaces: There, the associated permutation is defined by $f(C) = -C$, and, for $C \neq \emptyset$,

$$\begin{aligned} C \in K^{\psi\varphi} &\iff f(C) = f(D_0) \cap f(D_1) \cap ... \cap f(D_n) \\ &\iff -C = -D_0 \cap -D_1 \cap ... \cap -D_n \\ &\iff C = D_0 \cup D_1 \cup ... \cup D_n \end{aligned}$$

for some $D_0, ..., D_n \in K^+$.

3 A double entail query procedure

It was argued convincingly by Falmagne et al. (1990) that requiring a set K of states to be closed under intersection is too rigid, since a problem may be solved in different ways and different prerequisites could have been be used to arrive at a solution for this problem. A similar argument can be employed to reason that requiring a set of states to be closed under union is likewise unsatisfactory, though possibly less so: Given that the solution of a problem depends on certain skills, a combination of skills minimally required to solve, say, two problems p and q may enable the subject to solve other problems as well. Even worse, using different strategies (i.e. combinations of skills) may produce different minimal upper bounds for the states containing both p and q, so that the set of knowledge states need not even be a lattice. Indeed, we can have the following situations:

1. K is a sublattice of $\langle 2^Q, \cup, \cap, \emptyset, Q \rangle$.
2. $\langle K, \subseteq \rangle$ induces a sub-semilattice of $\langle 2^Q, \cup \rangle$.
3. $\langle K, \subseteq \rangle$ induces a sub-semilattice of $\langle 2^Q, \cap \rangle$.
4. $\langle K, \subseteq \rangle$ has a lattice structure.

5. K is arbitrary.

It might be worthy of mention that every finite semilattice with smallest and largest elements is in fact a lattice, so that we have the implications 2. $\Rightarrow$ 4. and 3. $\Rightarrow$ 4. above.

The question arises, which of these situations can be adequately captured by query procedures. The result of Koppen & Doignon (1990) takes care of 2. above, and it was noted in Falmagne et al. (1990) that asking the expert to determine a quasi order on Q by $p \leq q$ iff the capability of solving of q enables the subject to solve p as well gives rise to knowledge structures which are sublattices of $\langle 2^Q, \cup, \cap, \emptyset, Q \rangle$.

To arrive at a query procedure for 3. we let $\langle \sigma, \rho \rangle$ be the entail connection associated with the identity function on 2^Q, so that

$$\begin{aligned} K^\sigma &= \{\langle A, B\rangle : (\forall C \in K)(A \subseteq C \Rightarrow B \subseteq C)\}, \\ T^\rho &= \{C \subseteq Q : (\forall \langle A, B\rangle \in T)(A \subseteq C \Rightarrow B \subseteq C)\}. \end{aligned}$$

The next result follows immediately from 2.3:

Proposition 3.1. *For each $K \in \mathcal{K}$, $K^{\sigma\rho}$ is the closure of K under $\cap$.* □

It may be argued that this construction does not really help, since Falmagne's arguments against closure of K under intersection still apply. However, if we combine this query procedure with the one by Koppen & Doignon (1990), we can capture knowledge structures which are lattices with respect to set inclusion:

Proposition 3.2. *Let $\langle \psi, \varphi \rangle$ be the entail connection of Proposition 1.2, and $\langle \sigma, \rho \rangle$ the entail connection above. If $K \in \mathcal{K}$, and $\langle K, \vee \rangle$ is a join semilattice whose natural ordering is $\subseteq$, then, $K = K^{\sigma\rho} \cap K^{\psi\varphi}$.*

Proof. "$\subseteq$" is clear, since $K \subseteq K^{\psi\varphi}$ and $K \subseteq K^{\sigma\rho}$. For the converse, let $C \in K^{\psi\varphi} \cap K^{\sigma\rho}$. Then, there are $A_i, B_j \in K, i \leq r, j \leq s$ such that $\bigcup_{i \leq r} A_i = C = \bigcap_{j \leq s} B_j$. Let $D = sup_K\{A_i : i \leq r\}$. Since $\vee$ is compatible with $\subseteq$, we have $A \cup B \subseteq A \vee B$ for all $A, B \in K$, and therefore $C \subseteq D$. Since $\langle K, \vee \rangle$ is a semilattice, we also have $D \in K$. Now, each B_j is an upper bound for all A_j, and thus $D \subseteq B_j$ for all $j \leq s$. Hence, $C \subseteq D \subseteq \bigcap_{j \leq s} B_j = C$ which implies $C = D \in K$. □

The assignment $K \mapsto K^{\sigma\rho} \cap K^{\psi\varphi}$ - regarded as a combination of query procedures - thus captures knowledge structures which are lattices whose natural order is set inclusion.

The converse does not hold in general, as the example $Q = \{1, 2, 3, 4, 5\}$ and

$$K = \{\emptyset, \{1\}, \{2\}, \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, Q\}$$

shows. Such an example does not seem unreasonable: A subject may employ different strategies to solve problems 1 and 2 involving different skills, which, in one instance, allow her to solve 3 and 4 as well, while the skills used in another strategy enable her to solve 3 and 5.

A simple example shall demonstrate the double entail procedure. The true knowledge structure K is the union of a Guttman scale and its reverse:[2]

$$K = \left\{ \begin{array}{ccccc} \emptyset, & \{1\}, & \{1,2\}, & \{1,2,3\}, & Q, \\ & \{4\}, & \{3,4\}, & \{2,3,4\} & \end{array} \right\}$$

Using the Koppen-Doignon procedure, querying a perfect expert will result in: $\{1,3\}R\{2\}$, $\{1,4\}R\{2\}$, $\{1,4\}R\{3\}$, $\{2,4\}R\{3\}$, and the corresponding knowledge space looks like this:

$$K^{\psi\varphi} = \left\{ \begin{array}{ccccc} \emptyset, & \{1\}, & \{1,2\}, & \{1,2,3\}, & Q, \\ & & \{1,4\}, & \{1,2,4\}, & \\ & & & \{1,3,4\}, & \\ & \{4\}, & \{3,4\}, & \{2,3,4\} & \end{array} \right\}$$

Let T the relation based on the question type: "If a subject is able to solve all problems in A, the subject will solve all problems in B." Given the structure K, the query procedure will produce the following result: $\{2,4\}T\{3\}$, $\{1,3\}T\{2\}$, $\{1,4\}T\{2\}$, $\{1,4\}T\{3\}$. Therefore,

$$K^{\sigma\rho} = \left\{ \begin{array}{ccccc} \emptyset, & \{1\}, & \{1,2\}, & \{1,2,3\}, & Q, \\ & \{2\}, & \{2,3\}, & & \\ & \{3\}, & & & \\ & \{4\}, & \{3,4\}, & \{2,3,4\} & \end{array} \right\}$$

The structure K is captured, because $K = K^{\psi\varphi} \cap K^{\sigma\rho}$.

It is straightforward to show that not all knowledge structures can be captured by the proposed procedure: Because the resulting structure is an intersection of a $\cap$-stable and a $\cup$-stable structure, we may always find non capturable structures with $|Q| \geq 5$, in which all subsets having k elements are totally missing, while all subsets having $k-1$ or $k+1$ elements are present.

The question remains which knowledge structures can be captured by combining entail connections. The answer is given by

Proposition 3.3. *There is a family $\langle \alpha_i, \beta_i \rangle_{i \in I}$ of entail correspondences such that $K = \bigcap_{i \in I} K^{\alpha_i \beta_i}$ for all knowledge structures K.*

[2] The composed Guttman scale may look somewhat obscure. Nevertheless, the structure can be regarded as a prototype for the situation where the student population is divided into two groups, one being taught by a lecturer using a top–down approach, and the other group by a lecturer using a bottom–up approach on the same material. The problems are chosen from a very specialized context (problem 1) to a very general context (problem 4). If we assume that the teaching style is directly connected with the skills the students learn, we can suppose that the resulting knowledge structure is a composed Guttman scale.

Proof. For each proper non empty subset A of Q, choose a dual atom B of 2^Q above A, and let f_A be the permutation of 2^Q which exchanges A and B, and leaves everything else fixed; let $\mathcal{F}$ be the family of all chosen permutations.

For each $f \in \mathcal{F}$, let $\langle \alpha_f, \beta_f \rangle$ be its associated entail connection. Let $K \in \mathcal{K}$ and assume that $A \in \bigcap_{f \in \mathcal{F}} K^{\alpha_f \beta_f} \setminus K$. Then, in particular, $A \neq Q$, and thus we can choose a dual atom B of 2^Q such that $A \subseteq B$, and some $f \in \mathcal{F}$ which exchanges A and B and leaves everything else fixed. By our assumption, $A \in K^{\alpha_f \beta_f}$, so that

$$B = f(A) = f(D_0) \cap f(D_1) \cap \dots \cap f(D_n)$$

for some $D_0, \dots, D_n \in K^+$ by 2.6. Since B is a dual atom of 2^Q, we must have $B = f(D_i)$ for some $i \leq n$. But then, $B = f(A) = f(D_i)$ implies $A = D_i$, contradicting $A \notin K$. □

This shows that, in principle, every knowledge structure can be captured by combining query procedures. However, the procedure given above is of no practical value, and we should like to know whether fewer and more intuitive procedures will suffice to obtain the same result.

4 Strategy based query procedures

Building knowledge structures using a query procedure based on entail relations faces two problems:

1. The double entail procedure cannot capture all possible knowledge structures.
2. The combination of two structural very similar query procedures may be too demanding on an expert.

In the sequel we shall offer alternative query procedures based on strategy examination. Before we proceed, we recall a few definitions and results from Düntsch & Gediga (1995)

4.1 Skills and strategies

Let S be a fixed finite non empty sets of skills, and let $\delta : 2^S \to 2^Q$ be a function. Intuitively, δ associates with each set of skills X the set of problems which can be solved with the skills in X. It is reasonable to assume that by acquiring more skills an individual is able to solve at least the same problems as before. Thus, δ should be monotone, i.e. $X \subseteq Y \Rightarrow \delta(X) \subseteq \delta(Y)$. Thus, we call a function $\delta : 2^S \to 2^Q$ a *problem function* if

1. $\delta(\emptyset) = \emptyset$, $\delta(S) = Q$
2. δ is monotone with respect to $\subseteq$.

A function $\gamma : Q \to 2^{2^S}$ with the property that for each $q \in Q$

$$\gamma(q) \text{ is a non empty set of non empty, pairwise incomparable subsets of } S$$

is called a *skill function*. Intuitively, each element of $\gamma(q)$ is a set of skills minimally sufficient to solve problem q. Following Falmagne et al. (1990), we call a subset A of Q a γ - *knowledge state*, if there is some $X \subseteq S$ such that A is exactly the set of problems q for which a member of $\gamma(q)$ is a subset of X. The collection of all such knowledge states is denoted by K_γ.

An element of $\gamma(q)$ is called a *strategy* for q. More generally, given a non empty set A of problems, let $C(A)$ be the set of ranges of all choice functions $f : A \to \bigcup\{\gamma(q) : q \in A\}$. Each element of $C(A)$ is a subset of 2^S and intersects each $\gamma(q)$ in at least one element.

For $Z \in C(A)$, we let $M_Z = \bigcup Z \subseteq S$, and we call M_Z a *strategy* for (solving all problems of) A.

Each skill function implicitly describes a query procedure: For each problem $q \in Q$ we ask the expert which sets of skills are minimally sufficient to solve q.

In Düntsch & Gediga (1995) we have shown that the concepts of problem function and skill function are equivalent, and we have described how this query procedure translates into a set K of knowledge states: If γ is a given skill function with associated collection K_γ of knowledge states, and $\delta : 2^S \to 2^Q$ is defined by

$$\delta(X) = \{q \in Q : (\exists Y \in \gamma(q))Y \subseteq X\}$$

then $ran\ \delta = K_\gamma$.

Furthermore, every family of subsets of Q which contains $\emptyset$ and Q can be realized as a set of states for an appropriate skill function. A similar result was independently established in Doignon (1994).

Objections may be raised against a skill theoretic foundation of knowledge structures. Even though this note is not the appropriate place to answer these objections in detail, we will briefly point out some arguments for and against this approach. For further information we refer the reader to Doignon (1994) and Düntsch & Gediga (1995).

The "any state combination is possible" argument: The major hypothesis of the knowledge structure project is that not all sets of problems constitute knowledge states by assuming stability under intersection ("closure systems") or unions ("knowledge spaces"). It was shown in Doignon (1994) that – from a skill theoretical point of view – these systems use the following assumptions: knowledge spaces can be obtained from skill knowledge structures in which every strategy consists of only one skill, whereas closure systems can be obtained from skill knowledge structures in which any $\gamma(q)$ consists of exactly one set. On the other hand, in the knowledge structure project, the formulation of a skill function serves as a frame to specify the restrictions of the theory by the user himself without using technical restrictions. Statistical techniques such as cross-validation are necessary in either approach to validate the chosen theoretic assumptions.

The "any skill combination is possible" argument: Skill knowledge structures do not assume a priori that there is a specified dependency in the set of skills; this is not a new phenomenon. Besides, posing restrictions on the set of skills (e.g. in terms of a learning process, like the theory of Falmagne et al. (1990)) is better done in probabilistic terms and not in algebraic all-or-none restrictions. A skill knowledge structure will set up the basic events for a probabilistic skill knowledge structure theory.

It may be illuminating to translate the conditions for the query relations above into a scenario which, in addition to problems, also considers the skills needed to solve these problems. Thus, suppose that S is a set of skills, $\gamma : Q \to 2^{2^S}$ a skill function and $\delta : 2^S \to 2^Q$ its associated problem function.

1. Being able to solve q entails being able to solve p: If this is true, then every strategy for q produces a strategy for p a fact already observed in Falmagne et al. (1990), which led to the rejection of this query procedure as an adequate means to construct a knowledge structure. Thus, we have

$$(\forall X \in \gamma(q))(\exists Y \in \gamma(p))(Y \subseteq X).$$

2. Failing all problems in A entails failing all problems in B: We interpret this as saying that if a set X of skills is not sufficient to solve any problem in A, then X is not sufficient to solve any problem in B:

$$(\forall X \subseteq S)[(\forall p \in A)(\forall Y \in \gamma(p))Y \not\subseteq X \Rightarrow (\forall p \in B)(\forall Y \in \gamma(p))Y \not\subseteq X].$$

3. Being able to solve all problems in A entails being able to solve all problems in B, i.e. any strategy for A contains a strategy for B:

$$(\forall Y \in C(A))(\exists Z \in C(B))M_Z \subseteq M_Y.$$

Observe that in all of these statements the strategies for A are universally quantified, i.e. they do not distinguish among different strategies which could be used to solve A.

4.2 The strategy list query procedure

We now describe a query procedure, which is based on a list of all different strategies to solve a set of problems:

1. Choose a nonempty $A \subsetneq Q$ and fix it.

2. Given A, the expert is asked to list all strategies $C_i(A), i \leq k$, which can be used to solve all problems in A.

3. For each strategy $C_i(A)$, the expert names the set of all problems B_i, which can be solved using $C_i(A)$. Note that $A \subseteq B_i$ for all $i \leq k$.

This way, a mapping $\Gamma : 2^Q \to 2^{2^Q}$ is defined with $\Gamma(A) = \{B_i : i \leq k\}$. Note that each element of $\Gamma(A)$ is a knowledge state, since it is the image under δ of $C_i(A) \subseteq S$, and that A itself is a state if and only if $B_i = A$ for some $i \leq k$.

To facilitate the exclusion of non states, let us define

$$\Lambda(A) = \{C \subseteq Q : A \subseteq C, \text{ and } C \text{ is a proper non empty subset of a minimal element of } \{B_i : i \leq k\}\}.$$

We now have

Proposition 4.1. *Let K be a knowledge structure and Γ, Λ be defined as above. Then,*

1. $\bigcup_{A \subseteq Q}\{X \subseteq Q : X \in \Gamma(A)\} \cup \{\emptyset, Q\} = K$.
2. $B \notin K$ *for all* $B \in \Lambda(A)$.

Proof. 1. "$\subseteq$" follows immediately from the definition of Γ.

"$\supseteq$": Let $A \in K, A \neq \emptyset, Q$. Then, there is a set X of skills such that $\delta(X) = A$. X contains the union $C_i(A)$ of a strategy for A, and $\delta(X) = \delta(C_i(A)) = A$.

2. Let $B \in \Lambda(A)$ and assume that $B \in K$. Then, there is a set of skills X such that $\delta(X) = B$, and, since $A \subseteq B$, X contains some $C_i(A)$. Now, $C_i(A) \subseteq X$ implies that $\delta(C_i) \subseteq \delta(X) = B$. Hence, B contains some B_i, contrary to our definition of B. □

The result of Proposition 4.1 can be applied to obtain a more efficient query procedure. Because known states and non states are of no interest for the query, a book keeping algorithm excludes any already classified subset of Q. To illustrate the procedure, we show in Table 1 how this query procedure works with a skill function γ for the 2-Guttman-scale example, which provides different strategies for any problem q and any set $A \in (2^Q)^+$:

$$\begin{aligned}
\gamma(1) &= \{\{p_1\}, \{q_1, q_2, q_3, q_4\}, \{p_1, q_1\}\} \\
\gamma(2) &= \{\{p_1, p_2\}, \{q_1, q_2, q_3\}, \{p_1, q_1\}\} \\
\gamma(3) &= \{\{p_1, p_2, p_3\}, \{q_1, q_2\}, \{p_1, q_1\}\} \\
\gamma(4) &= \{\{p_1, p_2, p_3, p_4\}, \{q_1\}, \{p_1, q_1\}\}
\end{aligned}$$

The query procedure terminates, since all elements of 2^Q are classified into states or non states.

Table 1: Example of the strategy list procedure

Query sets $A \subseteq Q$	$\Gamma(A)$	Additional states	Additional non–states
Start:		$\emptyset, Q$	-
$\{1\}$	$\{\{1\}, Q\}$	$\{1\}$	-
$\{2\}$	$\{\{1,2\},\{2,3,4\}\}$	$\{1,2\},\{2,3,4\}$	$\{2\}$, $\{2,4\}$, $\{2,3\}$
$\{3\}$	$\{\{1,2,3\},\{3,4\}\}$	$\{1,2,3\},\{3,4\}$	$\{3\}$, $\{1,3\}$
$\{4\}$	$\{\{4\}, Q\}$	$\{4\}$	-
$\{1,2\}$ (state)			
$\{1,3\}$ (non state)			
$\{1,4\}$	$\{Q, Q\}$	-	$\{1,4\}$, $\{1,2,4\}$, $\{1,3,4\}$
Stop!			

5 Discussion

We have presented three different methods for the generation of knowledge structures: The skill function approach, which is based on the mapping $\gamma : Q \to 2^{2^S}$, the strategy list procedure, which results in a mapping $\Gamma : 2^Q \to 2^{2^Q}$, and the double entail relation approach, which has the deficit that there are knowledge structures that cannot be captured by this procedure. If we ask an ideal expert, the first two methods can be used to generate the true knowledge structure, while the double entail relation approach need only give an approximation.

Because all these procedures are theoretical valid, we should consider some criteria to choose the best query procedure:

Flexibility: The strategy list procedure seems a very flexible tool: It can be combined with any other approach at any point of time, since we only need to know the actual states and non states to start the strategy list procedure.

Psychological complexity: Psychologically, the skill function approach is the most complex procedure. The expert needs a sound theory which serves as background to formulate the skill function. Even if there were such a theory, it seems to be hard to formulate more than one strategy per problem. Raven (1965), for example, formulates a skill theory for his Coloured Progressive Matrices which consists only of one strategy for each problem. Although on first glance this procedure does not seem to have any advantage over presenting the expert with sets of problems and asking her/him to answer for each one of them one of the questions mentioned at the beginning, we feel that the approach based on skills – i.e. asking the expert to produce a skill function – is less demanding than it seems: In setting the problems in the first place, the expert should already have in mind the mastering of which sets of skills will be tested by each problem, so that most of the work can already be done at this stage. Furthermore, the aim of constructing tests is that the results of the test procedure offer some insight about the skills of the tested individuals. Therefore, there ideally should be a skill theory that produces the knowledge states, a process, which we describe by the mapping γ.

The strategy list procedure uses aggregations of skills, which we have called strategies. We feel that the term "strategy" is more intuitive than "skill combination" and that the labeling of a strategy is by far easier than the combination of many skill atoms.

The double entail relation approach uses comparisons of subsets A, B of Q. These comparisons do not take into account different strategies available for solving A and B. There is no guarantee that the expert really decides this way; there is not even a hint that she/he should. So, we are doubtful whether the outcome of an entail relation approach really captures the true structure.

Number of interactions: The last criterion concerns the complexity of one question. A further criterion is the number of necessary interactions with the expert to get the true knowledge structure. If the knowledge structure builds the Boolean algebra 2^Q, the skill function approach dominates the other procedures, because it is easy to construct an independent skill base, but there is no relation or strategy based procedure that could shorten the query procedure. Further investigation should be done in comparing the proposed procedures in case of highly structured knowledge structures.

Our aim in this note was to investigate the Galois connection between knowledge structures and entail relations, and the generation of query procedures for such weakly structured domains as knowledge structures. A final evaluation of which method is the most suitable query procedure is outside the scope of the present paper. This has to be done by performing psychological experiments, which we are currently undertaking.

References

ANDRÉKA, H., DÜNTSCH, I. & NÉMETI, I. (1995), Binary relations and permutation groups. *Mathematical Logic Quarterly*, **41**, 197–216.

BIRKHOFF, G. (1967), *Lattice Theory*, vol. 25 of *Am. Math. Soc. Colloquium Publications*. Amer. Math. Soc., Providence, 3rd edn.

DOIGNON, J.-P. (1994), Knowledge spaces and skill assignments. In: G. Fischer & D. Laming (eds.), *Contributions to Mathematical Psychology, Psychometrics, and Methodology*, Springer, Berlin.

DÜNTSCH, I. & GEDIGA, G. (1995), Skills and knowledge structures. *British Journal of. Mathematical and Statistical Psychology*, **48**, 9–27.

ERNÉ, M. (1982), *Einführung in die Ordnungstheorie*. Bibliographisches Institut, Zürich.

FALMAGNE, J.-C., KOPPEN, M., VILLANO, M., DOIGNON, J.-P. & JOHANNESEN, J. (1990), Introduction to knowledge spaces: How to build, test and search them. *Psychological Review*, **97**, 202–234.

GAREY, M. R. & JOHNSON, D. S. (1979), *Computers and intractibility: A guide to the theory of NP–completeness*. Freeman & Co, San Francisco.

GRÄTZER, G. (1978), *General Lattice Theory*. Birkhäuser, Basel.

KOPPEN, M. & DOIGNON, J.-P. (1990), How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology*, **34**, 311–331.

MONJARDET, D. (1970), Tresses, fuseaux, préordres et topologies. *Mathématique et Sciences Humaines*, **30**, 11–22.

NOVOTNÝ, M. (1997), Dependence spaces of information systems. In: E. Orłowska (ed.), *Incomplete Information – Rough Set Analysis*, pp. 193–246, Physica – Verlag, Heidelberg.

RAVEN, J. C. (1965), *Guide to using the couloured progressive matrices sets A, Ab, B (Revised version)*. Gieve, London.

Abbreviations and Symbols

2^X: Power set of X.

$(2^X)^+ := 2^X \setminus \{\emptyset\}$

$ran_R x := \{y \in X : xRy\}$.

A^σ: The image of $A \in X$ under the mapping $\sigma : X \to Y$.

Q: Denotes a finite set of problems.

K: Throughout we use K, K', etc. to denote knowledge structures.

K^+: $K \setminus \{\emptyset\}$.

$\mathcal{K}$: The Boolean algebra of all knowledge structures on Q.

$\mathcal{P}$: The set of all binary relations on $(2^Q)^+$.

$\mathcal{E}$: The set of all entail relations on $(2^Q)^+$.

S: A fixed finite non empty set of skills.

δ: The problem function.

γ: The skill function.

M_Z: The union of all skill sets in $Z \subseteq 2^{2^S}$ (a strategy).

Γ: The resulting mapping of the strategy list procedure.