Learning, Forgetting, and the Correlation of Knowledge in Knowledge Space Theory

Jeffrey Matayoshi and Hasan Uzun

McGraw Hill ALEKS, Irvine, CA, United States

Abstract

In this work we introduce and study multiple properties and conditions related to the modeling of student knowledge in knowledge space theory (KST). We begin by looking at a property called forgetting consistency, which enforces a systematic way of forgetting within a knowledge structure. Next, we analyze in detail a concept we call positive knowledge correlation. This concept postulates that knowing more should not make it less likely that a student knows a particular concept. Among other things, we find that satisfying positive knowledge correlation implies the knowledge structure is closed under both union and intersection, and we also perform an empirical evaluation to assess the validity of the property. Finally, in the context of an adaptive assessment, we conclude with an analysis of the related concept of a positively correlated updating rule.

Keywords: knowledge space theory, positive knowledge correlation, forgetting consistency, positively correlated updating rule

1. Introduction

Knowledge space theory (KST) is a mathematical model of knowledge introduced by Doignon and Falmagne (1985). Since the time of its introduction, it has been successfully used in many applications involving the learning and assessment of knowledge (Cosyn et al., 2021; de Chiusole et al., 2020; Doble et al., 2019; Falmagne et al., 2013; Falmagne and Doignon, 2011; Hockemeyer et al., 1997; Lynch and Howlin, 2014). In this work we introduce and examine the implications of multiple properties pertaining to knowledge spaces and the modeling of student knowledge. We begin by looking at the significance of a property we call *forgetting consistency*. This condition enforces a notion of forgetting within the knowledge space, and we show that—in combination with one other condition—it implies the knowledge space is closed under intersection. Next, we look at a more general concept we refer to as *positive knowledge correlation*. The basic idea of positive knowledge correlation is that knowing more should not make it less likely for a student to know something else. We show that this property has fairly strong implications, as it implies that the knowledge structure is closed under both union and intersection. Additionally, we give the results of an empirical study that seemingly supports the concept of positive knowledge correlation. Finally, in the context of an adaptive assessment, we conclude with a discussion of the related concept of a *positively correlated updating rule*.

1.1. Background on Knowledge Space Theory

In this section we briefly introduce a few KST concepts that are necessary for our subsequent work. Much of this follows the exposition in Falmagne and Doignon (2011); thus, for a more thorough introduction to KST we refer the reader there. We begin with the related notions of a *knowledge structure* and a *knowledge space*.

Definition 1.1. A *knowledge structure* is a pair $(Q, \mathcal{K})$ in which Q is a nonempty set, and $\mathcal{K}$ is a family of subsets of Q, containing at least Q and the empty set $\varnothing$. The set Q is called the *domain* of the knowledge structure. Its elements are referred to as *questions* or *items* and the subsets in the family $\mathcal{K}$ are labeled *(knowledge) states*. Since $\cup\mathcal{K} = Q$, we shall sometimes simply say that $\mathcal{K}$ is the knowledge structure when reference to the underlying domain is not necessary. If a knowledge structure $\mathcal{K}$ is closed under union, we say that $\mathcal{K}$ is a *knowledge space*.

In this work Q is always assumed to be a finite set—thus, as a consequence all the knowledge structures we consider are also finite. Motivated by pedagogical assumptions, Cosyn and Uzun (2009) introduced two axioms that define a *learning space*, a specific type of knowledge structure.

Definition 1.2. A knowledge structure $(Q, \mathcal{K})$ is called a learning space if it satisfies the following conditions.

[LS] *Learning smoothness.* For any two states K, L such that $K \subset L$, there exists a finite chain of states

$$K = K_0 \subset K_1 \subset \cdots \subset K_p = L$$

such that $|K_i \setminus K_{i-1}| = 1$ for $1 \leq i \leq p$ and so $|L \setminus K| = p$.

[LC] *Learning consistency.* If K, L are two states satisfying $K \subset L$ and q is an item such that $K \cup \{q\} \in \mathcal{K}$, then $L \cup \{q\} \in \mathcal{K}$.

A useful concept associated with knowledge structures is well-gradedness, which we define as in Doignon and Falmagne (1997).

Definition 1.3. Let Δ denote the standard symmetric difference between sets. Then, a family of sets $\mathcal{F}$ is *well-graded* if for any $A, B \in \mathcal{F}$ with $|A \Delta B| = n$, there exists a finite sequence of sets $A = K_0, K_1, \ldots, K_n = B$ in $\mathcal{F}$ such that $|K_{i-1} \Delta K_i| = 1, i = 1, \ldots, n$. The sequence of sets $A = K_0, K_1, \ldots, K_n = B$ satisfying these conditions is called a *tight path* between A and B.

A notable result from Cosyn and Uzun (2009) showed that a learning space is equivalent to a well-graded union-closed family.

Theorem 1.4 (Cosyn and Uzun)**.** Let $\mathcal{F}$ be a family of sets containing the empty set. Then, $\mathcal{F}$ is well-graded and union-closed if and only if [LS] and [LC] are satisfied. In other words, well-graded union-closed families of sets are characterized by the axioms [LS] and [LC].

2. Forgetting in KST

As the historic focus of KST has been on the learning process, relatively less attention has been paid to the concept of forgetting. In particular, the Ebbinghaus forgetting curve (Averell and Heathcote, 2011; Ebbinghaus, 1913) is a model that represents the decay of knowledge over time. Recent empirical research on the ALEKS system has shown numerous examples of forgetting curves in the context of an actual implementation of KST (Cosyn et al., 2021; Matayoshi et al., 2018, 2019, 2020). Motivated by these results, we introduce the following condition for forgetting within a knowledge space.

Definition 2.1. [FC] *Forgetting consistency.* If K, L are two states satisfying $K \subset L$ and q is an item such that $L \setminus \{q\} \in \mathcal{K}$, then $K \setminus \{q\} \in \mathcal{K}$.

Note that [FC] is, in a sense, analogous to [LC], the learning consistency condition. However, while learning consistency ensures that a student who knows more is always able to learn the same as a student who knows less, forgetting consistency ensures that a student who knows less is able to forget the same knowledge as a student who knows more. In other words, if a student who knows more is able to “forget” item q, it seems plausible that a student with less knowledge is able to forget q as well. Our next result shows that a knowledge structure satisfying [LS] and [FC] is well-graded and closed under intersection.

Theorem 2.2. Let $\mathcal{K}$ be a knowledge structure satisfying [LS] and [FC]. Then, $\mathcal{K}$ is well-graded and closed under intersection.

Proof. Let $K, L \in \mathcal{K}$. We first show that $K \cap L \in \mathcal{K}$. By [LS], there exists a tight path from K to $Q = \cup\mathcal{K}$ given by

$$\begin{aligned} K = K_0 \subset K_1 = K \cup \{k_1\} \subset \cdots \\ \subset K_{n-1} = K \cup \{k_1, \ldots, k_{n-1}\} \subset Q = K_n = K \cup \{k_1, \ldots, k_n\}, \end{aligned} \tag{2.1}$$

where $n = |Q \setminus K|$. Consider $k_n \in K_n \setminus K_{n-1}$. By [FC] we have that $L \setminus \{k_n\} \in \mathcal{K}$, as $L \subseteq K_n = Q$ and $K_n \setminus \{k_n\} = K_{n-1} \in \mathcal{K}$. Then, applying a similar procedure with $k_{n-1} \in K_{n-1} \setminus K_{n-2}$, it follows that $L \setminus \{k_{n-1}, k_n\} \in \mathcal{K}$. Iteratively applying this procedure a total of n times, we end with $M = L \setminus \{k_1, k_2, \ldots, k_{n-1}, k_n\} \in \mathcal{K}$, where each $k_i \in K_i \setminus K_{i-1}$ for $i = 1, \ldots, n$. Note that since $\{k_1, k_2, \ldots, k_{n-1}, k_n\} = Q \setminus K$, we have that

$$\begin{aligned} M &= L \setminus \{k_1, k_2, \ldots, k_{n-1}, k_n\} \\ &= L \setminus (Q \setminus K) \\ &= L \setminus (L \setminus K) \\ &= K \cap L. \end{aligned}$$

Thus, $K \cap L \in \mathcal{K}$, and it follows that $\mathcal{K}$ is closed under intersection.

We next show that a tight path exists from K to L. To start, consider the sequence from K to $K \cap L$

$$\begin{aligned} K = A_0 \supset A_1 = L \setminus \{a_1\} \supset \cdots \\ \supset A_{m-1} = L \setminus \{a_1, \ldots, a_{m-1}\} \supset L \setminus \{a_1, \ldots, a_{m-1}, a_m\} = K \cap L, \end{aligned} \tag{2.2}$$

where $m = |K \setminus L|$; note this sequence exists since [LS] holds for $\mathcal{K}$. Next, we can define an analogous sequence from L to $K \cap L$

$$\begin{aligned} L = B_0 \supset B_1 = L \setminus \{b_1\} \supset \cdots \\ \supset B_{n-1} = L \setminus \{b_1, \ldots, b_{n-1}\} \supset L \setminus \{b_1, \ldots, b_{n-1}, b_n\} = K \cap L, \end{aligned} \tag{2.3}$$

where $n = |L \setminus K|$; as before, this sequence exists since $\mathcal{K}$ satisfies [LS]. Concatenating these two sequences, we now have a sequence from K to L of length $m + n = |K \Delta L|$, and the well-gradedness of $\mathcal{K}$ then follows. $\square$

Combining Theorem 2.2 with Theorem 1.4, we immediately get the following.

Corollary 2.3. Let $\mathcal{K}$ be a knowledge structure satisfying [LS], [LC], and [FC]. Then, $\mathcal{K}$ is a well-graded knowledge structure that is closed under union and intersection.

3. Positive Knowledge Correlation

In Section 2 we introduced forgetting consistency, which is analogous to the learning consistency condition that is required of a learning space. Our next goal is to develop a set of conditions that we refer to as *positive knowledge correlation*. In what follows, we assume that, for a set of items Q, we have a probability distribution on $\mathcal{P}(Q)$, the power set of Q. This probability distribution might represent the distribution of the states in a knowledge structure, with this distribution being derived from some reference population of students. Or, as another example, the probability distribution could represent the uncertainty around a particular student's knowledge state during a KST-based assessment.

The main idea behind the concept of positive knowledge correlation is that knowing more should make it more likely—or, at the very least, it should not make it less likely—that a student knows a particular item q. Conversely, knowing less should make it less likely—or, at the very least, it should not make it more likely—that a student knows an item q. As we discuss in more detail in the next subsection, in some sense these conditions can be viewed as probabilistic analogues of learning and forgetting consistency.

3.1. Conditions for Positive Knowledge Correlation

We begin with the following definition.

Definition 3.1. For a nonempty set of items Q, let P be a probability distribution on $\mathcal{P}(Q)$, the power set of Q. Define the set family

$$\mathcal{K}_P = \{K \subseteq Q \mid P(K) > 0\}. \tag{3.1}$$

If $P(\varnothing) > 0$ and $P(Q) > 0$, it follows that $(Q, \mathcal{K}_P)$ is a knowledge structure; in such a case, we say it is the knowledge structure induced by P.

In what follows, we also need to make use of the following standard definitions from order theory.

Definition 3.2. Let Q be a set of items. For a set $A \subseteq Q$, the *upper closure* of A is defined as

$$\{B \subseteq Q \mid A \subseteq B\}, \tag{3.2}$$

while the *lower closure* of A is given by

$$\{B \subseteq Q \mid B \subseteq A\}. \tag{3.3}$$

The next definition gives us a convenient way of representing sets of items that are known or not known.

Definition 3.3. Let Q be a set of items. For a set $A \subseteq Q$ let I_A^+ be the upper closure of A and I_A^- be the lower closure of A^{c}; that is,

$$I_A^+ = \{B \subseteq Q \mid A \subseteq B\} \tag{3.4}$$

and

$$\begin{aligned} I_A^- &= \{B \subseteq Q \mid B \subseteq A^{\mathsf{c}}\} \\ &= \{B \subseteq Q \mid A \cap B = \varnothing\}. \end{aligned} \tag{3.5}$$

Plainly speaking, I_A^+ consists of all the sets that contain all the items from A; intuitively, these are the sets where all the items in A are known. Then, I_A^- consists of all the sets containing no items from A; in this case, these are the sets where none of the items in A are known.

Given a family $\mathcal{F} \subseteq \mathcal{P}(Q)$, we next define

$$P(\mathcal{F}) := \begin{cases} 0, & \text{if } \mathcal{F} \text{ is empty} \\ \displaystyle\sum_{K \in \mathcal{F}} P(K), & \text{otherwise.} \end{cases} \tag{3.6}$$

Using the above definition, for $A, B, C \subseteq Q$, if $\sum_{K \in I_B^+ \cap I_C^-} P(K) > 0$ we can compute

$$\begin{aligned} P\left(I_A^+ \,\middle|\, I_B^+, I_C^-\right) &= \frac{P\left(I_A^+ \cap I_B^+ \cap I_C^-\right)}{P\left(I_B^+ \cap I_C^-\right)} \\ &= \frac{\sum_{K \in I_A^+ \cap I_B^+ \cap I_C^-} P(K)}{\sum_{K \in I_B^+ \cap I_C^-} P(K)}. \end{aligned} \tag{3.7}$$

That is, $P\left(I_A^+ \,\middle|\, I_B^+, I_C^-\right)$ is the conditional probability of knowing the items in A, given that all the items in B are known and none of the items in C are known.

With these definitions in hand, we are now ready to introduce the concept of *positive knowledge correlation*.

Definition 3.4. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$. Let $q, r \in Q$ and $B, C \subseteq Q$. Suppose that

$$P\left(I_{\{q\}}^+ \,\middle|\, I_{B\cup\{r\}}^+, I_C^-\right) \geq P\left(I_{\{q\}}^+ \,\middle|\, I_B^+, I_C^-\right) \tag{3.8}$$

whenever $P\left(I_{B\cup\{r\}}^+ \cap I_C^-\right) > 0$. In such a case, we say that P satisfies the property of *positive knowledge correlation*.

In words, (3.8) says that, compared to knowing only the items in B, knowing all the items in $B \cup \{r\}$ should not decrease the probability of knowing q. Also, notice that by requiring $P\left(I_{B\cup\{r\}}^+ \cap I_C^-\right) > 0$ it implicitly follows that $P\left(I_B^+ \cap I_C^-\right) > 0$ as well. To see this, observe that $I_{B\cup\{r\}}^+ \cap I_C^- \subseteq I_B^+ \cap I_C^-$, which implies that

$$P\left(I_B^+ \cap I_C^-\right) \geq P\left(I_{B\cup\{r\}}^+ \cap I_C^-\right) > 0.$$

As such, both conditional probabilities in (3.8) are well-defined. Note that we use requirements of a similar form for many of our subsequent results.

Our next result shows that, if we assume the relevant conditional probabilities are all well-defined, the property of positive knowledge correlation can be formulated in several equivalent ways.

Theorem 3.5. Let Q, P, q, r, B, and C be as in Definition 3.4. Assume that $P\left(I_{B\cup\{r\}}^+ \cap I_C^-\right) > 0$ and $P\left(I_B^+ \cap I_{C\cup\{r\}}^-\right) > 0$. Then, the following inequalities are all equivalent.

(i) $P\left(I_{\{q\}}^+ \,\middle|\, I_{B\cup\{r\}}^+, I_C^-\right) \geq P\left(I_{\{q\}}^+ \,\middle|\, I_B^+, I_C^-\right)$

(ii) $P\left(I_{\{q\}}^+ \,\middle|\, I_B^+, I_{C\cup\{r\}}^-\right) \leq P\left(I_{\{q\}}^+ \,\middle|\, I_B^+, I_C^-\right)$

(iii) $P\left(I^-_{\{q\}} \,\middle|\, I^+_B, I^-_{C\cup\{r\}}\right) \geq P\left(I^-_{\{q\}} \,\middle|\, I^+_B, I^-_C\right)$

(iv) $P\left(I^-_{\{q\}} \,\middle|\, I^+_{B\cup\{r\}}, I^-_C\right) \leq P\left(I^-_{\{q\}} \,\middle|\, I^+_B, I^-_C\right)$

Proof.
(i) $\Longrightarrow$ (ii):
We have

$$\begin{aligned}
P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_{C\cup\{r\}}\right) &= \frac{P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)}{P\left(I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)} \\
&= \frac{P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_C\right) - P\left(I^+_{\{q\}} \cap I^+_{\{r\}} \cap I^+_B \cap I^-_C\right)}{P\left(I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)} \\
&= \frac{P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_C\right) - P\left(I^+_{\{q\}} \,\middle|\, I^+_{B\cup\{r\}}, I^-_C\right) \cdot P\left(I^+_{\{r\}} \cap I^+_B \cap I^-_C\right)}{P\left(I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)} \\
&\leq \frac{P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_C\right) - P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \cdot P\left(I^+_{\{r\}} \cap I^+_B \cap I^-_C\right)}{P\left(I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)},
\end{aligned}$$

where the inequality in the last line follows from (i). Next, we have

$$= \frac{P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_C\right) - P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \cdot P\left(I^+_B \cap I^-_C\right) \cdot P\left(I^+_{\{r\}} \,\middle|\, I^+_B, I^-_C\right)}{P\left(I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)}$$

$$= \frac{P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_C\right) - P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_C\right) \cdot P\left(I^+_{\{r\}} \,\middle|\, I^+_B, I^-_C\right)}{P\left(I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)}$$

$$= P\left(I^+_{\{q\}} \cap I^+_B \cap I^-_C\right) \frac{1 - P\left(I^+_{\{r\}} \,\middle|\, I^+_B, I^-_C\right)}{P\left(I^+_B \cap I^-_{\{r\}} \cap I^-_C\right)}$$

$$= P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \frac{1 - P\left(I^+_{\{r\}} \,\middle|\, I^+_B, I^-_C\right)}{P\left(I^-_{\{r\}} \,\middle|\, I^+_B, I^-_C\right)}$$

$$= P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right).$$

(ii) $\Longrightarrow$ (iii):
Starting from (ii) we have

$$\begin{aligned} & P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_{C\cup\{r\}}\right) \le P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \\ \iff \quad & 1 - P\left(I^-_{\{q\}} \,\middle|\, I^+_B, I^-_{C\cup\{r\}}\right) \le 1 - P\left(I^-_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \\ \iff \quad & P\left(I^-_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \le P\left(I^-_{\{q\}} \,\middle|\, I^+_B, I^-_{C\cup\{r\}}\right), \end{aligned}$$

as claimed.
(iii) $\Longrightarrow$ (iv):

We have

$$
\begin{aligned}
P\left(I_{\{q\}}^{-} \,\middle|\, I_{B\cup\{r\}}^{+}, I_{C}^{-}\right) &= \frac{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{C}^{-}\right)} \\
&= \frac{P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right) - P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{\{r\}}^{-} \cap I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{C}^{-}\right)} \\
&= \frac{P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right) - P\left(I_{\{q\}}^{-} \,\middle|\, I_{B}^{+}, I_{C\cup\{r\}}^{-}\right) \cdot P\left(I_{B}^{+} \cap I_{\{r\}}^{-} \cap I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{C}^{-}\right)} \\
&\leq \frac{P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right) - P\left(I_{\{q\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right) \cdot P\left(I_{B}^{+} \cap I_{\{r\}}^{-} \cap I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{C}^{-}\right)},
\end{aligned}
$$

where the inequality in the last line follows from (iii). Next, we have

$$
\begin{aligned}
&= \frac{P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right) - P\left(I_{\{q\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right) \cdot P\left(I_{B}^{+} \cap I_{C}^{-}\right) \cdot P\left(I_{\{r\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{C}^{-}\right)} \\
&= \frac{P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right) - P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right) \cdot P\left(I_{\{r\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{C}^{-}\right)} \\
&= P\left(I_{B}^{+} \cap I_{\{q\}}^{-} \cap I_{C}^{-}\right) \frac{1 - P\left(I_{\{r\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \cap I_{B}^{+} \cap I_{C}^{-}\right)} \\
&= P\left(I_{\{q\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right) \frac{1 - P\left(I_{\{r\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right)}{P\left(I_{\{r\}}^{+} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right)} \\
&= P\left(I_{\{q\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right).
\end{aligned}
$$

(iv) $\Longrightarrow$ (i):

Starting from (iv) we have

$$\begin{aligned} & P\left(I_{\{q\}}^{-} \,\middle|\, I_{B\cup\{r\}}^{+}, I_{C}^{-}\right) \leq P\left(I_{\{q\}}^{-} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right) \\ \iff \quad & 1 - P\left(I_{\{q\}}^{+} \,\middle|\, I_{B\cup\{r\}}^{+}, I_{C}^{-}\right) \leq 1 - P\left(I_{\{q\}}^{+} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right) \\ \iff \quad & P\left(I_{\{q\}}^{+} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right) \leq P\left(I_{\{q\}}^{+} \,\middle|\, I_{B\cup\{r\}}^{+}, I_{C}^{-}\right), \end{aligned}$$

as claimed. □

Theorem 3.5 shows there are several equivalent ways to view the concept of positive knowledge correlation. That is, while the formulation in (i) is in terms of knowing more—analogous to the learning consistency condition—the concept of positive knowledge correlation can also be viewed as a condition on knowing less, such as in (ii). Furthermore, analogous to forgetting consistency, (iii) implies that knowing less should not lower the probability of an item q being unknown.

Notice that the inequalities listed in Definition 3.4 and Theorem 3.5 are all stated in terms of the probability of knowing—or, not knowing—a single item. Our next result shows that, perhaps surprisingly, the more general case of knowing or not knowing groups of items is implied by these simpler inequalities.

Theorem 3.6. Let Q, P, q, and r be as in Definition 3.4. Furthermore, let A, B_1, B_2, C_1 and C_2 be subsets of Q such that $B_1 \subseteq B_2$, $C_1 \subseteq C_2$, $P\left(I_{B_2}^{+} \cap I_{C_1}^{-}\right) > 0$, and $P\left(I_{B_1}^{+} \cap I_{C_2}^{-}\right) > 0$. Assume that Definition 3.4 holds. Then, the following hold as well.

(a) $P\left(I_{A}^{+} \,\middle|\, I_{B_2}^{+}, I_{C_1}^{-}\right) \geq P\left(I_{A}^{+} \,\middle|\, I_{B_1}^{+}, I_{C_2}^{-}\right)$

(b) $P\left(I_{A}^{-} \,\middle|\, I_{B_2}^{+}, I_{C_1}^{-}\right) \leq P\left(I_{A}^{-} \,\middle|\, I_{B_1}^{+}, I_{C_2}^{-}\right)$

Proof.

(a) $P\left(I_{A}^{+} \,\middle|\, I_{B_2}^{+}, I_{C_1}^{-}\right) \geq P\left(I_{A}^{+} \,\middle|\, I_{B_1}^{+}, I_{C_2}^{-}\right)$:

To start, suppose $P\left(I_{A}^{+} \cap I_{B_1}^{+} \cap I_{C_1}^{-}\right) = 0$. Observe that

$$I_{A}^{+} \cap I_{B_2}^{+} \cap I_{C_1}^{-} \subseteq I_{A}^{+} \cap I_{B_1}^{+} \cap I_{C_1}^{-}$$

and

$$I_A^+ \cap I_{B_1}^+ \cap I_{C_2}^- \subseteq I_A^+ \cap I_{B_1}^+ \cap I_{C_1}^-,$$

from which it follows that

$$P\left(I_A^+ \,\middle|\, I_{B_2}^+, I_{C_1}^-\right) = P\left(I_A^+ \,\middle|\, I_{B_1}^+, I_{C_2}^-\right) = 0.$$

Next, assume $P\left(I_A^+ \cap I_{B_1}^+ \cap I_{C_1}^-\right) > 0$. Let $B_2 \setminus B_1 = \{b_1, b_2, \ldots, b_n\}$ and $A = \{a_1, a_2, \ldots, a_k\}$. Note that for any $A' \subseteq A$ we have

$$I_A^+ \cap I_{B_1}^+ \cap I_{C_1}^- \subseteq I_{A'}^+ \cap I_{B_1}^+ \cap I_{C_1}^-,$$

which implies that

$$P\left(I_{A'}^+ \cap I_{B_1}^+ \cap I_{C_1}^-\right) \geq P\left(I_A^+ \cap I_{B_1}^+ \cap I_{C_1}^-\right) > 0. \tag{3.9}$$

From (3.8) it then follows that

$$P\left(I_{\{b_1\}}^+ \,\middle|\, I_{B_1}^+, I_{C_1}^-\right) \leq P\left(I_{\{b_1\}}^+ \,\middle|\, I_{B_1 \cup \{a_1\}}^+, I_{C_1}^-\right),$$

where both conditional probabilities are well-defined by 3.9. Repeatedly applying (3.8) to $a_1, a_2, \ldots, a_k$ we get

$$\begin{aligned}
P\left(I_{\{b_1\}}^+ \,\middle|\, I_{B_1}^+, I_{C_1}^-\right) &\leq P\left(I_{\{b_1\}}^+ \,\middle|\, I_{B_1 \cup \{a_1\}}^+, I_{C_1}^-\right) \\
&\leq P\left(I_{\{b_1\}}^+ \,\middle|\, I_{B_1 \cup \{a_1, a_2\}}^+, I_{C_1}^-\right) \\
&\;\;\vdots \\
&\leq P\left(I_{\{b_1\}}^+ \,\middle|\, I_{B_1 \cup \{a_1, \ldots, a_k\}}^+, I_{C_1}^-\right) \\
&= P\left(I_{\{b_1\}}^+ \,\middle|\, I_A^+, I_{B_1}^+, I_{C_1}^-\right),
\end{aligned}$$

where each of the conditional probabilities is well-defined by 3.9. We then have

$$
\begin{aligned}
& P\left(I^+_{\{b_1\}} \,\middle|\, I^+_A, I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \\
\iff \quad & \frac{P\left(I^+_{\{b_1\}} \cap I^+_A \cap I^+_{B_1} \cap I^-_{C_1}\right)}{P\left(I^+_A \cap I^+_{B_1} \cap I^-_{C_1}\right)} \geq \frac{P\left(I^+_{\{b_1\}} \cap I^+_{B_1} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \\
\iff \quad & \frac{P\left(I^+_{\{b_1\}} \cap I^+_A \cap I^+_{B_1} \cap I^-_{C_1}\right)}{P\left(I^+_{\{b_1\}} \cap I^+_{B_1} \cap I^-_{C_1}\right)} \geq \frac{P\left(I^+_A \cap I^+_{B_1} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \\
\iff \quad & P\left(I^+_A \,\middle|\, I^+_{B_1 \cup \{b_1\}}, I^-_{C_1}\right) \geq P\left(I^+_A \,\middle|\, I^+_{B_1}, I^-_{C_1}\right).
\end{aligned}
$$

Applying a similar procedure to $b_2, \ldots, b_n$ it follows that

$$
P\left(I^+_A \,\middle|\, (I^+_{B_2}, I^-_{C_1}\right) \geq P\left(I^+_A \,\middle|\, I^+_{B_1}, I^-_{C_1}\right). \tag{3.10}
$$

Next, let $C_2 \setminus C_1 = \{c_1, c_2, \ldots, c_m\}$. Repeatedly applying (3.8) gives

$$
\begin{aligned}
P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) &= 1 - P\left(I^+_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \\
&\geq 1 - P\left(I^+_{\{c_1\}} \,\middle|\, I^+_{B_1 \cup \{a_1\}}, I^-_{C_1}\right) \\
&\geq 1 - P\left(I^+_{\{c_1\}} \,\middle|\, I^+_{B_1 \cup \{a_1, a_2\}}, I^-_{C_1}\right) \\
&\;\;\vdots \\
&\geq 1 - P\left(I^+_{\{c_1\}} \,\middle|\, I^+_{B_1 \cup \{a_1, \ldots, a_k\}}, I^-_{C_1}\right) \\
&= 1 - P\left(I^+_{\{c_1\}} \,\middle|\, I^+_A, I^+_{B_1}, I^-_{C_1}\right) \\
&= P\left(I^-_{\{c_1\}} \,\middle|\, I^+_A, I^+_{B_1}, I^-_{C_1}\right).
\end{aligned}
$$

where each of the conditional probabilities is well-defined by 3.9. We then have

$$
\begin{aligned}
& P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^-_{\{c_1\}} \,\middle|\, I^+_A, I^+_{B_1}, I^-_{C_1}\right) \\
\iff \quad & \frac{P\left(I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \geq \frac{P\left(I^+_A \cap I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{C_1}\right)}{P\left(I^+_A \cap I^+_{B_1} \cap I^-_{C_1}\right)} \\
\iff \quad & \frac{P\left(I^+_A \cap I^+_{B_1} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \geq \frac{P\left(I^+_A \cap I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{C_1}\right)} \\
\iff \quad & P\left(I^+_A \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^+_A \,\middle|\, I^+_{B_1}, I^-_{C_1 \cup \{c_1\}}\right).
\end{aligned}
$$

Applying a similar procedure to $c_2, \ldots, c_m$ it follows that

$$
P\left(I^+_A \,\middle|\, (I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^+_A \,\middle|\, I^+_{B_1}, I^-_{C_2}\right). \tag{3.11}
$$

Combining (3.10) and (3.11) we arrive at the claimed inequality (a):

$$
P\left(I^+_A \,\middle|\, I^+_{B_2}, I^-_{C_1}\right) \geq P\left(I^+_A \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^+_A \,\middle|\, I^+_{B_1}, I^-_{C_2}\right). \tag{3.12}
$$

(b) $P\left(I^-_A \,\middle|\, I^+_{B_2}, I^-_{C_1}\right) \leq P\left(I^-_A \,\middle|\, I^+_{B_1}, I^-_{C_2}\right)$:

To start, suppose $P\left(I^-_A \cap I^+_{B_1} \cap I^-_{C_1}\right) = 0$. Then,

$$
I^-_A \cap I^+_{B_2} \cap I^-_{C_1} \subseteq I^-_A \cap I^+_{B_1} \cap I^-_{C_1}
$$

and

$$
I^-_A \cap I^+_{B_1} \cap I^-_{C_2} \subseteq I^-_A \cap I^+_{B_1} \cap I^-_{C_1},
$$

from which it follows that

$$
P\left(I^-_A \,\middle|\, I^+_{B_2}, I^-_{C_1}\right) = P\left(I^-_A \,\middle|\, I^+_{B_1}, I^-_{C_2}\right) = 0.
$$

Next, assume $P\left(I^-_A \cap I^+_{B_1} \cap I^-_{C_1}\right) > 0$. Let $B_2 \setminus B_1 = \{b_1, b_2, \ldots, b_n\}$ and $A = \{a_1, a_2, \ldots, a_k\}$. Note that for any $A' \subseteq A$ we have

$$
I^-_A \cap I^+_{B_1} \cap I^-_{C_1} \subseteq I^-_{A'} \cap I^+_{B_1} \cap I^-_{C_1},
$$

which implies that

$$P\left(I^-_{A'} \cap I^+_{B_1} \cap I^-_{C_1}\right) \geq P\left(I^-_{A} \cap I^+_{B_1} \cap I^-_{C_1}\right) > 0. \tag{3.13}$$

From (ii) in Theorem 3.5 it then follows that

$$P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1\cup\{a_1\}}\right),$$

where both conditional probabilities are well-defined by 3.13. Repeatedly applying (ii) in Theorem 3.5 to $a_1, a_2, \ldots, a_k$ we get

$$\begin{aligned}
P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) &\geq P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1\cup\{a_1\}}\right) \\
&\geq P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1\cup\{a_1,a_2\}}\right) \\
&\;\;\vdots \\
&\geq P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1\cup\{a_1,\ldots,a_k\}}\right) \\
&= P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{A}, I^-_{C_1}\right),
\end{aligned}$$

where each of the conditional probabilities is well-defined by 3.13. We then have

$$\begin{aligned}
& P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^+_{\{b_1\}} \,\middle|\, I^+_{B_1}, I^-_{A}, I^-_{C_1}\right) \\
\iff \quad & \frac{P\left(I^+_{\{b_1\}} \cap I^+_{B_1} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \geq \frac{P\left(I^+_{\{b_1\}} \cap I^+_{B_1} \cap I^-_{A} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{A} \cap I^-_{C_1}\right)} \\
\iff \quad & \frac{P\left(I^+_{B_1} \cap I^-_{A} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \geq \frac{P\left(I^+_{\{b_1\}} \cap I^+_{B_1} \cap I^-_{A} \cap I^-_{C_1}\right)}{P\left(I^+_{\{b_1\}} \cap I^+_{B_1} \cap I^-_{C_1}\right)} \\
\iff \quad & P\left(I^-_{A} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^-_{A} \,\middle|\, I^+_{B_1\cup\{b_1\}}, I^-_{C_1}\right).
\end{aligned}$$

Applying a similar procedure to $b_2, \ldots, b_n$ it follows that

$$P\left(I^-_{A} \,\middle|\, (I^+_{B_1}, I^-_{C_1}\right) \geq P\left(I^-_{A} \,\middle|\, I^+_{B_2}, I^-_{C_1}\right). \tag{3.14}$$

Next, let $C_2 \setminus C_1 = \{c_1, c_2, \dots, c_m\}$. Repeatedly applying (iii) from Theorem 3.5 gives

$$\begin{aligned} P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) &\leq P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1 \cup \{a_1\}}\right) \\ &\leq P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1 \cup \{a_1, a_2\}}\right) \\ &\ \ \vdots \\ &\leq P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1 \cup \{a_1, \dots, a_k\}}\right) \\ &= P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{A}, I^-_{C_1}\right). \end{aligned}$$

where each of the conditional probabilities is well-defined by 3.13. We then have

$$\begin{aligned} & P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \leq P\left(I^-_{\{c_1\}} \,\middle|\, I^+_{B_1}, I^-_{A}, I^-_{C_1}\right) \\ \iff \quad & \frac{P\left(I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \leq \frac{P\left(I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{A} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{A} \cap I^-_{C_1}\right)} \\ \iff \quad & \frac{P\left(I^+_{B_1} \cap I^-_{A} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{C_1}\right)} \leq \frac{P\left(I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{A} \cap I^-_{C_1}\right)}{P\left(I^+_{B_1} \cap I^-_{\{c_1\}} \cap I^-_{C_1}\right)} \\ \iff \quad & P\left(I^-_{A} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \leq P\left(I^-_{A} \,\middle|\, I^+_{B_1}, I^-_{C_1 \cup \{c_1\}}\right). \end{aligned}$$

Applying a similar procedure to $c_2, \dots, c_m$ it follows that

$$P\left(I^-_{A} \,\middle|\, (I^+_{B_1}, I^-_{C_1}\right) \leq P\left(I^-_{A} \,\middle|\, I^+_{B_1}, I^-_{C_2}\right). \tag{3.15}$$

Combining (3.14) and (3.15) we arrive at the claimed inequality (b):

$$P\left(I^-_{A} \,\middle|\, I^+_{B_2}, I^-_{C_1}\right) \leq P\left(I^-_{A} \,\middle|\, I^+_{B_1}, I^-_{C_1}\right) \leq P\left(I^-_{A} \,\middle|\, I^+_{B_1}, I^-_{C_2}\right). \tag{3.16}$$

□

In a sense, positive knowledge correlation can be interpreted as a probabilistic analogue of [LC], the learning consistency condition, and [FC], the forgetting consistency condition. Recall that the learning consistency condition postulates that knowing more does not prevent the learning of something new. In comparison, positive knowledge correlation can be interpreted as saying that knowing more does not make knowing something else less likely. Then, as the forgetting consistency condition postulates that knowing less does not prevent the forgetting of something already learned, positive knowledge correlation says that knowing less makes it less likely that something else is known. To help us formalize these ideas, we next introduce the following conditions, which can be thought of as slightly weaker versions of the property of positive knowledge correlation.

Definition 3.7. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$. We define the following two conditions.

3(a) For any $q \in Q$ and $B, C \subseteq Q$, where $C \neq \varnothing$, we have

$$P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \geq P\left(I^+_{\{q\}} \,\middle|\, I^-_C\right), \tag{3.17}$$

whenever the conditional probabilities are well-defined.

3(b) For any $q \in Q$ and $B, C \subseteq Q$, where $B \neq \varnothing$, we have

$$P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right) \leq P\left(I^+_{\{q\}} \,\middle|\, I^+_B\right), \tag{3.18}$$

whenever the conditional probabilities are well-defined.

The formulation in Theorem 3.6 makes it clear that both 3(a) and 3(b) follow directly from the property of positive knowledge correlation—however, as these new conditions are similar in form to the inequalities in Definition 3.4, it's not completely clear that they are weaker properties. To show this is in fact the case, we next give examples where 3(a) and 3(b) are satisfied, but positive knowledge correlation fails to hold.

Example 3.8. For $Q = \{x, y, z\}$ and $0 < \alpha < \frac{1}{8}$, consider the following probability distribution P on $\mathcal{P}(Q)$.

$$\begin{aligned} P(\{x\}) = P(\{y\}) &= \frac{1}{8} + \alpha, & P(\{z\}) &= \alpha, \\ P(\{x,z\}) = P(\{y,z\}) &= \frac{1}{8} - \alpha, & P(\{x,y\}) &= \frac{1}{4} - \alpha, \\ P(\{x,y,z\}) &= \alpha, & P(\varnothing) &= \frac{1}{4} - \alpha \end{aligned}$$

Note that, as the states correspond to the power set of $\{x, y, z\}$, $\mathcal{K} = \mathcal{K}_P$ is closed under both union and intersection.

- 3(a) $\not\Longrightarrow$ Definition 3.4:

Starting with 3(a), note that the following inequalities all hold when $\alpha \leq \frac{1}{16}$.

$$
\begin{aligned}
P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{y\}}, I^-_{\{z\}}\right) &= \frac{2}{3} - \frac{8}{3}\alpha \geq \frac{1}{2} = P\left(I^+_{\{x\}} \,\middle|\, I^-_{\{z\}}\right) \\
P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{x\}}, I^-_{\{z\}}\right) &= \frac{2}{3} - \frac{8}{3}\alpha \geq \frac{1}{2} = P\left(I^+_{\{y\}} \,\middle|\, I^-_{\{z\}}\right) \\
P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{z\}}, I^-_{\{y\}}\right) &= 1 - 8\alpha \geq \frac{1}{2} = P\left(I^+_{\{x\}} \,\middle|\, I^-_{\{y\}}\right) \\
P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{z\}}, I^-_{\{x\}}\right) &= 1 - 8\alpha \geq \frac{1}{2} = P\left(I^+_{\{y\}} \,\middle|\, I^-_{\{x\}}\right) \\
P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x\}}, I^-_{\{y\}}\right) &= \frac{1}{2} - 4\alpha \geq \frac{1}{4} = P\left(I^+_{\{z\}} \,\middle|\, I^-_{\{y\}}\right) \\
P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{y\}}, I^-_{\{x\}}\right) &= \frac{1}{2} - 4\alpha \geq \frac{1}{4} = P\left(I^+_{\{z\}} \,\middle|\, I^-_{\{x\}}\right)
\end{aligned}
\tag{3.19}
$$

Thus, while we've now shown that 3(a) holds for $\alpha \leq \frac{1}{16}$, notice that for $\alpha < \frac{1}{16}$ we have

$$
P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x,y\}}\right) = 4\alpha < \frac{1}{4} = P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x\}}\right), \tag{3.20}
$$

contradicting (3.8). Combining this result with the fact that 3(a) is implied by the property of positive knowledge correlation, we can now see that 3(a) is in fact the weaker condition.

- 3(b) $\not\Longrightarrow$ Definition 3.4:

We next turn to 3(b). Note that the following inequalities all hold when

$\frac{1}{16} \leq \alpha < \frac{1}{8}$.

$$\begin{aligned}
P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{y\}}, I^-_{\{z\}}\right) &= \frac{2}{3} - \frac{8}{3}\alpha \leq \frac{1}{2} = P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{y\}}\right) \\
P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{x\}}, I^-_{\{z\}}\right) &= \frac{2}{3} - \frac{8}{3}\alpha \leq \frac{1}{2} = P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{x\}}\right) \\
P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{z\}}, I^-_{\{y\}}\right) &= 1 - 8\alpha \leq \frac{1}{2} = P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{z\}}\right) \\
P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{z\}}, I^-_{\{x\}}\right) &= 1 - 8\alpha \leq \frac{1}{2} = P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{z\}}\right) \\
P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x\}}, I^-_{\{y\}}\right) &= \frac{1}{2} - 4\alpha \leq \frac{1}{4} = P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x\}}\right) \\
P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{y\}}, I^-_{\{x\}}\right) &= \frac{1}{2} - 4\alpha \leq \frac{1}{4} = P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{y\}}\right)
\end{aligned} \tag{3.21}$$

Thus, we have now shown that 3(b) holds for $\frac{1}{16} \leq \alpha < \frac{1}{8}$. However, for $\frac{1}{16} < \alpha < \frac{1}{8}$ we get

$$P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x\}}, I^-_{\{y\}}\right) = \frac{1}{2} - 4\alpha < \frac{1}{4} = P\left(I^+_{\{z\}} \,\middle|\, I^-_{\{y\}}\right), \tag{3.22}$$

contradicting (3.8); thus, the property of positive knowledge correlation does not follow from 3(b).

The previous example shows that each condition in Definition 3.7 is weaker than the property of positive knowledge correlation. Additionally, Example 3.8 is important for another reason. While we will eventually show that the conditions in Definition 3.7 are useful from a technical perspective, the example suggests they are not as compelling for modeling student knowledge. Specifically, while 3(a) holds when $\alpha < \frac{1}{16}$, from (3.20) we can see that, in comparison to knowing only x, knowing both x and y makes it *less* likely that z is known, which runs counter to the motivation for the property of positive knowledge correlation. Similarly, while 3(b) holds when $\alpha > \frac{1}{16}$, (3.22) shows that, in comparison to not knowing y, adding the additional assumption of knowing x makes it *less* likely that z is known, which again goes against the intuition of positive knowledge correlation. Thus, while we will see shortly that the conditions in Definition 3.7 have important implications, for the above reasons we choose to formulate the property of positive knowledge correlation as done in Definition 3.4.

Before moving on to the main consequences of the conditions in Definition 3.7—and, hence, the property of positive knowledge correlation—we need to derive the following equivalent forms of these conditions.

Lemma 3.9. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$.

(a) Condition 3(a) can be equivalently written as follows:

For any $q \in Q$ and $B, C \subseteq Q$, where $C \neq \varnothing$, we have

$$P\left(I_B^+ \,\middle|\, I_{\{q\}}^+, I_C^-\right) \geq P\left(I_B^+ \mid I_C^-\right) \tag{3.23}$$

whenever the conditional probabilities are well-defined.

(b) Condition 3(b) can be equivalently written as follows:

For any $q \in Q$ and $B, C \subseteq Q$, where $B \neq \varnothing$, we have

$$P\left(I_C^- \middle| I_B^+\right) \leq P\left(I_C^- \middle|\, I_B^+, I_{\{q\}}^-\right), \tag{3.24}$$

whenever the conditional probabilities are well-defined.

Proof.

(a) From the definition of conditional probability, 3(a) can be written as

$$\begin{aligned}
& \frac{P\left(I_{\{q\}}^+ \cap I_B^+ \cap I_C^-\right)}{P\left(I_B^+ \cap I_C^-\right)} \geq \frac{P\left(I_{\{q\}}^+ \cap I_C^-\right)}{P\left(I_C^-\right)} \\
\iff \quad & \frac{P\left(I_{\{q\}}^+ \cap I_B^+ \cap I_C^-\right)}{P\left(I_{\{q\}}^+ \cap I_C^-\right)} \geq \frac{P\left(I_B^+ \cap I_C^-\right)}{P\left(I_C^-\right)} \\
\iff \quad & P\left(I_B^+ \,\middle|\, I_{\{q\}}^+, I_C^-\right) \geq P\left(I_B^+ \mid I_C^-\right).
\end{aligned}$$

(b) 3(b) can be equivalently written as

$$\begin{aligned}
& 1 - P\left(I_{\{q\}}^- \,\middle|\, I_B^+, I_C^-\right) \leq 1 - P\left(I_{\{q\}}^- \,\middle|\, I_B^+\right) \\
\iff \quad & P\left(I_{\{q\}}^- \,\middle|\, I_B^+\right) \leq P\left(I_{\{q\}}^- \,\middle|\, I_B^+, I_C^-\right).
\end{aligned}$$

From the definition of conditional probability, the last inequality can be written as follows.

$$
\begin{aligned}
& \frac{P\left(I_B^+ \cap I_{\{q\}}^-\right)}{P\left(I_B^+\right)} \leq \frac{P\left(I_B^+ \cap I_{\{q\}}^- \cap I_C^-\right)}{P\left(I_B^+ \cap I_C^-\right)}. \\
\iff \qquad & \frac{P\left(I_B^+ \cap I_C^-\right)}{P\left(I_B^+\right)} \leq \frac{P\left(I_B^+ \cap I_{\{q\}}^- \cap I_C^-\right)}{P\left(I_B^+ \cap I_{\{q\}}^-\right)}. \\
\iff \qquad & P\left(I_C^- \middle| I_B^+\right) \leq P\left(I_C^- \middle| I_B^+, I_{\{q\}}^-\right)
\end{aligned}
$$

□

Lemma 3.10. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$ and $P(Q) > 0$. Suppose also that 3(a) holds. Then, the knowledge structure $\mathcal{K} = \mathcal{K}_P$ satisfies the learning consistency condition [LC].

Proof. Let $K, L \in \mathcal{K}$ with $K \subset L$ and let q be an item in Q such that $K \cup \{q\} \in \mathcal{K}$. From (3.1) we know that $P(K \cup \{q\}) > 0$. In order for learning consistency to hold, we need to show that $P(L \cup \{q\}) > 0$. We have

$$
\begin{aligned}
P(L \cup \{q\}) &= P\left(I_L^+ \cap I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) \\
&= P\left(I_L^+ \middle| I_{\{q\}}^+, I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) \cdot P\left(I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right), \qquad (3.25)
\end{aligned}
$$

where the second equality uses the definition of conditional probability. Note that $P\left(I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) > 0$ since $K \cup \{q\} \in I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-$, and so the above conditional probability is well-defined. Next, notice that if $L \cup \{q\} = Q$ we are done, as $P(Q) > 0$ by assumption. Thus, assuming $L \cup \{q\} \neq Q$, which implies that $(L \cup \{q\})^{\mathsf{c}} \neq \varnothing$, we can then apply 3.23 to get

$$
\begin{aligned}
&\geq P\left(I_L^+ \,\middle|\, I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) \cdot P\left(I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) \qquad \text{by 3.23} \\
&= \frac{P\left(I_L^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right)}{P\left(I_{(L\cup\{q\})^{\mathsf{c}}}^-\right)} \cdot P\left(I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) \\
&\geq \frac{P(L)}{P\left(I_{(L\cup\{q\})^{\mathsf{c}}}^-\right)} \cdot P\left(I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) \qquad \text{since } L \in I_L^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^- \\
&\geq P(L) \cdot P\left(I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^-\right) \\
&\geq P(L) \cdot P\left(K \cup \{q\}\right) \qquad \text{since } K \cup \{q\} \in I_{\{q\}}^+ \cap I_{(L\cup\{q\})^{\mathsf{c}}}^- \\
&> 0,
\end{aligned}
$$

where the last inequality follows from the fact that $L \in \mathcal{K}$ and $K \cup \{q\} \in \mathcal{K}$. $\square$

Lemma 3.11. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$ and $P(Q) > 0$. Suppose also that 3(b) holds. Then, the knowledge structure $\mathcal{K} = \mathcal{K}_P$ satisfies the forgetting consistency condition [FC].

Proof. Let $K, L \in \mathcal{K}$ with $K \subset L$ and let q be an item in K such that $L \setminus \{q\} \in \mathcal{K}$. From (3.1) we know that $P(L \setminus \{q\}) > 0$. In order for forgetting consistency to hold, we need to show that $P\left(K \setminus \{q\}\right) > 0$. We have

$$
\begin{aligned}
P\left(K \setminus \{q\}\right) &= P\left(I_{K\setminus\{q\}}^+ \cap I_{\{q\}}^- \cap I_{K^{\mathsf{c}}}^-\right) \\
&= P\left(I_{K^{\mathsf{c}}}^- \,\middle|\, I_{K\setminus\{q\}}^+, I_{\{q\}}^-\right) \cdot P\left(I_{K\setminus\{q\}}^+ \cap I_{\{q\}}^-\right), \qquad (3.26)
\end{aligned}
$$

where the second equality uses the definition of conditional probability. Note that $P\left(I_{K\setminus\{q\}}^+ \cap I_{\{q\}}^-\right) > 0$ since $L \setminus \{q\} \in I_{K\setminus\{q\}}^+ \cap I_{\{q\}}^-$, and so the above conditional probability is well-defined. Next, notice that if $K \setminus \{q\} = \varnothing$ we are done, as $P(\varnothing) > 0$ by assumption. Thus, assuming $K \setminus \{q\} \neq \varnothing$, we can apply 3.24 to get

$$
\begin{aligned}
&\geq P\left(I^-_{K^c} \,\middle|\, I^+_{K\setminus\{q\}}\right) \cdot P\left(I^+_{K\setminus\{q\}} \cap I^-_{\{q\}}\right) \qquad \text{by 3.24}\\
&= \frac{P\left(I^+_{K\setminus\{q\}} \cap I^-_{K^c}\right)}{P\left(I^+_{K\setminus\{q\}}\right)} \cdot P\left(I^+_{K\setminus\{q\}} \cap I^-_{\{q\}}\right)\\
&\geq \frac{P(K)}{P\left(I^+_{K\setminus\{q\}}\right)} \cdot P\left(I^+_{K\setminus\{q\}} \cap I^-_{\{q\}}\right) \qquad \text{since } K \in I^+_{K\setminus\{q\}} \cap I^-_{K^c}\\
&\geq P(K) \cdot P\left(I^+_{K\setminus\{q\}} \cap I^-_{\{q\}}\right)\\
&\geq P(K) \cdot P\left(L \setminus \{q\}\right) \qquad \text{since } L \setminus \{q\} \in I^+_{K\setminus\{q\}} \cap I^-_{\{q\}}\\
&> 0,
\end{aligned}
$$

where the last inequality follows from the fact that $K \in \mathcal{K}$ and $L \setminus \{q\} \in \mathcal{K}$. $\square$

Combining Lemmas 3.10 and 3.11 leads immediately to the following theorem.

Theorem 3.12. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$ and $P(Q) > 0$. Suppose also that P satisfies the property of positive knowledge correlation. Then, $\mathcal{K} = \mathcal{K}_P$ satisfies [LC] and [FC].

Our next two results show that 3(a) and 3(b) imply closure under union and closure under intersection, respectively.

Lemma 3.13. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$ and $P(Q) > 0$. Assume that 3(a) holds. Then, the knowledge structure $\mathcal{K} = \mathcal{K}_P$ is closed under union.

Proof. Let $K, L \subseteq Q$ with $P(K) > 0$ and $P(L) > 0$. Note that, by assumption, the inequality holds if $K \cup L = Q$. Furthermore, it also holds if $K \subseteq L$, as we have $P(K \cup L) = P(L) > 0$, or if $L \subseteq K$, since $P(K \cup L) = P(K) > 0$. Assuming that $K \cup L \neq Q$, $K \not\subseteq L$, and $L \not\subseteq K$, let $K \setminus L = \{r_1, r_2, ..., r_m\}$. We have

$$
\begin{aligned}
P(K \cup L) &= P\left(I^+_{\{r_1\}} \cap I^+_{K\cup L\setminus\{r_1\}} \cap I^-_{(K\cup L)^c}\right)\\
&= P\left(I^+_{K\cup L\setminus\{r_1\}} \,\middle|\, I^+_{\{r_1\}}, I^-_{(K\cup L)^c}\right) \cdot P\left(I^+_{\{r_1\}} \cap I^-_{(K\cup L)^c}\right).
\end{aligned}
$$

Note that $P\left(I^+_{\{r\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) > 0$ since $K \in I^+_{\{r\}} \cap I^-_{(K\cup L)^\mathsf{c}}$; as such, the above conditional probability is well-defined. Next, since we're assuming $K\cup L \neq Q$, we have $(K \cup L)^\mathsf{c} \neq \varnothing$; thus, we can apply (3.23) to get

$$\begin{aligned}
&\geq P\left(I^+_{K\cup L\setminus\{r_1\}} \,\middle|\, I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P\left(I^+_{\{r_1\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \\
&= \frac{P\left(I^+_{K\cup L\setminus\{r_1\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right)}{P\left(I^-_{(K\cup L)^\mathsf{c}}\right)} \cdot P\left(I^+_{\{r_1\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \\
&\geq \frac{P\left(I^+_{K\cup L\setminus\{r_1\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right)}{P\left(I^-_{(K\cup L)^\mathsf{c}}\right)} \cdot P(K) \qquad \text{since } K \in I^+_{\{r_1\}} \cap I^-_{(K\cup L)^\mathsf{c}} \\
&\geq P\left(I^+_{K\cup L\setminus\{r_1\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P(K).
\end{aligned}$$

Proceeding next with $r_2 \in K \setminus L$, we have

$$\begin{aligned}
&= P\left(I^+_{\{r_2\}} \cap I^+_{K\cup L\setminus\{r_1,r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P(K) \\
&= P\left(I^+_{K\cup L\setminus\{r_1,r_2\}} \,\middle|\, I^+_{\{r_2\}}, I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P\left(I^+_{\{r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P(K).
\end{aligned}$$

Note that $K \in I^+_{\{r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}$, which means $P\left(I^+_{\{r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) > 0$. Applying (3.23) once again gives

$$\begin{aligned}
&\geq P\left(I^+_{K\cup L\setminus\{r_1,r_2\}} \,\middle|\, I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P\left(I^+_{\{r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P(K) \\
&= \frac{P\left(I^+_{K\cup L\setminus\{r_1,r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right)}{P\left(I^-_{(K\cup L)^\mathsf{c}}\right)} \cdot P\left(I^+_{\{r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P(K) \\
&\geq \frac{P\left(I^+_{K\cup L\setminus\{r_1,r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right)}{P\left(I^-_{(K\cup L)^\mathsf{c}}\right)} \cdot P(K)^2 \qquad \text{since } K \in I^+_{\{r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}} \\
&\geq P\left(I^+_{K\cup L\setminus\{r_1,r_2\}} \cap I^-_{(K\cup L)^\mathsf{c}}\right) \cdot P(K)^2.
\end{aligned}$$

Proceeding similarly one at a time for $r_3, ..., r_m$, we have

$$
\begin{aligned}
P(K \cup L) &\geq P\left(I^+_{K \cup L \setminus \{r_1, r_2, .., r_m\}} \cap I^-_{(K \cup L)^\complement}\right) \cdot P(K)^m \\
&= P\left(I^+_L \cap I^-_{(K \cup L)^\complement}\right) \cdot P(K)^m \\
&\geq P(L) \cdot P(K)^m \qquad \text{since } L \in I^+_L \cap I^-_{(K \cup L)^\complement} \\
&> 0.
\end{aligned}
$$

Therefore, $\mathcal{K}$ is closed under union. $\square$

Lemma 3.14. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$ and $P(Q) > 0$. Assume that 3(b) holds. Then, the knowledge structure $\mathcal{K} = \mathcal{K}_P$ is closed under intersection.

Proof. Let $K, L \subseteq Q$ with $P(K) > 0$ and $P(L) > 0$. We will show that $P(K \cap L) > 0$, which implies that $\mathcal{K}$ is intersection-closed. Note that, by assumption, the inequality holds if $K \cap L = \varnothing$. Furthermore, it also holds if $K \subseteq L$, as we have $P(K \cap L) = P(K) > 0$, or if $L \subseteq K$, since $P(K \cap L) = P(L) > 0$. Assuming $K \cap L \neq \varnothing$, $K \nsubseteq L$, and $L \nsubseteq K$, let $K \setminus L = \{q_1, q_2, \cdots, q_n\}$. We have

$$
\begin{aligned}
P(K \cap L) &= P\left(I^+_{K \cap L} \cap I^-_{\{q_1\}} \cap I^-_{((K \cap L) \cup \{q_1\})^\complement}\right) \\
&= P\left(I^-_{((K \cap L) \cup \{q_1\})^\complement} \,\middle|\, I^+_{K \cap L}, I^-_{\{q_1\}}\right) \cdot P\left(I^+_{K \cap L} \cap I^-_{\{q_1\}}\right).
\end{aligned}
$$

Note that $P\left(I^+_{K \cap L} \cap I^-_{\{q_1\}}\right) > 0$ since $L \in I^+_{K \cap L} \cap I^-_{\{q_1\}}$; as such, the above conditional probability is well-defined. Since we're assuming $K \cap L \neq \varnothing$, we can apply (3.24) to get

$$
\begin{aligned}
&\geq P\left(I^-_{((K \cap L) \cup \{q_1\})^\complement} \,\middle|\, I^+_{K \cap L}\right) \cdot P\left(I^+_{K \cap L} \cap I^-_{\{q_1\}}\right) \\
&= \frac{P\left(I^+_{K \cap L} \cap I^-_{((K \cap L) \cup \{q_1\})^\complement}\right)}{P\left(I^+_{K \cap L}\right)} \cdot P\left(I^+_{K \cap L} \cap I^-_{\{q_1\}}\right) \\
&\geq \frac{P\left(I^+_{K \cap L} \cap I^-_{((K \cap L) \cup \{q_1\})^\complement}\right)}{P\left(I^+_{K \cap L}\right)} \cdot P(L) \qquad \text{since } L \in I^+_{K \cap L} \cap I^-_{\{q_1\}} \\
&\geq P\left(I^+_{K \cap L} \cap I^-_{((K \cap L) \cup \{q_1\})^\complement}\right) \cdot P(L).
\end{aligned}
$$

Note that $L \in I^+_{K\cap L} \cap I^-_{\{q_2\}}$, which means $P\left(I^+_{K\cap L} \cap I^-_{\{q_2\}}\right) > 0$. Thus, proceeding similarly for $q_2 \in K \setminus L$, by applying (3.24), we have

$$
\begin{aligned}
&= P\left(I^+_{K\cap L} \cap I^-_{\{q_2\}} \cap I^-_{((K\cap L)\cup\{q_1,q_2\})^{\mathsf{c}}}\right) \cdot P(L) \\
&= P\left(I^-_{((K\cap L)\cup\{q_1,q_2\})^{\mathsf{c}}} \middle| I^+_{K\cap L}, I^-_{\{q_2\}}\right) \cdot P\left(I^+_{K\cap L} \cap I^-_{\{q_2\}}\right) \cdot P(L) \\
&\geq P\left(I^-_{((K\cap L)\cup\{q_1,q_2\})^{\mathsf{c}}} \middle| I^+_{K\cap L}\right) \cdot P\left(I^+_{K\cap L} \cap I^-_{\{q_2\}}\right) \cdot P(L) \qquad \text{by (3.24)} \\
&= \frac{P\left(I^+_{K\cap L} \cap I^-_{((K\cap L)\cup\{q_1,q_2\})^{\mathsf{c}}}\right)}{P\left(I^+_{K\cap L}\right)} \cdot P\left(I^+_{K\cap L} \cap I^-_{\{q_2\}}\right) \cdot P(L) \\
&\geq \frac{P\left(I^+_{K\cap L} \cap I^-_{((K\cap L)\cup\{q_1,q_2\})^{\mathsf{c}}}\right)}{P\left(I^+_{K\cap L}\right)} \cdot P(L)^2,
\end{aligned}
$$

where the last line follows from the fact that $L \in I^+_{K\cap L} \cap I^-_{\{q_2\}}$. Proceeding similarly one at a time for $q_3, q_4, \cdots, q_n$, we get

$$
\begin{aligned}
P(K \cap L) &\geq P\left(I^+_{K\cap L} \cap I^-_{((K\cap L)\cup\{q_1,q_2,..,q_n\})^{\mathsf{c}}}\right) \cdot P(L)^n \\
&= P\left(I^+_{K\cap L} \cap I^-_{K^{\mathsf{c}}}\right) \cdot P(L)^n \\
&\geq P(K) \cdot P(L)^n \qquad \text{since } K \in I^+_{K\cap L} \cap I^-_{K^{\mathsf{c}}} \\
&> 0.
\end{aligned}
$$

Therefore, $\mathcal{K}$ is closed under intersection. □

Combining Lemmas 3.13 and 3.14, it follows immediately that a knowledge structure satisfying the property of positive knowledge correlation is necessarily closed under both union and intersection.

Theorem 3.15. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$ and $P(Q) > 0$. Suppose also that P satisfies the property of positive knowledge correlation. Then, $\mathcal{K} = \mathcal{K}_P$ is closed under both union and intersection.

Now that we have shown 3(a) and 3(b) imply a knowledge structure is closed under union and intersection, respectively, a natural question is whether these conditions are satisfied for *any* knowledge structure that is closed under both union and intersection. Continuing with Example 3.8, we next show that this is not the case.

Example 3.8 (continuing from p. 17)**.** Notice that the knowledge structure is closed under both union and intersection for any value of α in $(0, \frac{1}{8})$. However, consider that the inequalities in (3.19) are satisfied for $0 < \alpha \leq \frac{1}{16}$, but they do not hold for $\frac{1}{16} < \alpha < \frac{1}{8}$; thus, for the latter set of values condition 3(a) does not hold, while the knowledge structure is closed under both union and intersection. Similarly, the inequalities in (3.21) are satisfied for $\frac{1}{16} \leq \alpha < \frac{1}{8}$, but they do not hold for $0 < \alpha < \frac{1}{16}$; as such, in the latter case condition 3(b) does not hold, even though the knowledge structure is once again closed under both union and intersection.

Another question is whether the property of positive knowledge correlation can be reformulated without the inclusion of I_C^-; that is, is it possible to get the same results by formulating the property in terms of only knowing items, without mentioning the items that are not known? Our next example shows that this is not possible.

Example 3.16. For $Q = \{x, y, z\}$ consider the following probability distribution P on $\mathcal{P}(Q)$.

$$
\begin{aligned}
P(\{x\}) = P(\{y\}) &= \frac{1}{10}, & P(\{z\}) &= 0,\\
P(\{x, z\}) &= \frac{1}{10}, & P(\{x, y\}) = P(\{y, z\}) &= 0,\\
P(\{x, y, z\}) &= \frac{3}{5}, & P(\varnothing) &= \frac{1}{10}
\end{aligned}
$$

Note that this knowledge structure is closed under intersection, but it is not closed under union. Based on the probability distribution above, we can compute the following probabilities.

$$\begin{aligned} P\left(I^+_{\{y\}}\right) = P\left(I^+_{\{z\}}\right) &= \frac{7}{10} \\ P\left(I^+_{\{x\}}\right) &= \frac{8}{10} \\ P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{y\}}\right) = P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{z\}}\right) = P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{y\}}\right) &= \frac{6}{7} \\ P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{x\}}\right) &= \frac{3}{4} \\ P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{z\}}\right) &= 1 \\ P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x\}}\right) &= \frac{7}{8} \\ P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{x,z\}}\right) &= \frac{6}{7} \\ P\left(I^+_{\{x\}} \,\middle|\, I^+_{\{y,z\}}\right) = P\left(I^+_{\{z\}} \,\middle|\, I^+_{\{x,y\}}\right) &= 1 \end{aligned}$$

Observe that for any permutation σ of the items in $\{x, y, z\}$ we have

$$P\left(I^+_{\{\sigma(1)\}} \,\middle|\, I^+_{\{\sigma(2)\}}\right) \geq P\left(I^+_{\{\sigma(1)\}}\right)$$

and

$$P\left(I^+_{\{\sigma(1)\}} \,\middle|\, I^+_{\{\sigma(2),\sigma(3)\}}\right) \geq P\left(I^+_{\{\sigma(1)\}} \,\middle|\, I^+_{\{\sigma(2)\}}\right).$$

However, note that from the state probabilities we can compute

$$P\left(I^+_{\{y\}} \,\middle|\, I^+_{\{x\}}, I^-_{\{z\}}\right) = 0 \not\geq \frac{1}{3} = P\left(I^+_{\{y\}} \,\middle|\, I^-_{\{z\}}\right),$$

in violation of (3.8). Thus, since $\mathcal{K}$ is not union-closed, it follows that formulating (3.8) solely in terms of what is known is not enough to guarantee the resulting knowledge structure is closed under both union and intersection.

3.2. The FKG Inequality and Positive Knowledge Correlation

In the Section 3.1 we proved several results that followed when a knowledge structure and its associated probability distribution satisfy the property of positive knowledge correlation. In this next part we work in the other direction—that is, assuming we have a probability distribution on a knowledge structure, when can we conclude the property of positive knowledge correlation is satisfied? To answer this question, we need to make use of the Fortuin-Kasteleyn-Ginibre (FKG) inequality (Fortuin et al., 1971).

Theorem 3.17 (FKG inequality)**.** Let Γ be a finite distributive lattice. Let m be a positive measure on Γ satisfying the following condition:

(A) For all x and y in Γ,

$$m(x \vee y)m(x \wedge y) \geq m(x)m(y), \tag{3.27}$$

where $\vee$ and $\wedge$ represent the *join* and *meet* operations, respectively. Let f and g be monotonically increasing (or decreasing) functions on Γ. The following positive correlation inequality then holds:

$$\left(\sum_{x\in\Gamma} f(x)g(x)m(x)\right)\left(\sum_{x\in\Gamma} m(x)\right) \geq \left(\sum_{x\in\Gamma} f(x)m(x)\right)\left(\sum_{x\in\Gamma} g(x)m(x)\right). \tag{3.28}$$

The inequality becomes negatively correlated—i.e., the inequality sign is flipped—if one of f and g is monotonically increasing and the other is monotonically decreasing.

In order to apply the FKG inequality, we let $\Gamma = \mathcal{K}$, where $\mathcal{K}$ is a knowledge structure defined on a set of items Q. Furthermore, we assume that $\mathcal{K}$ is closed under both union and intersection—in this case, the join and meet operations are then represented by set union and set intersection, respectively. Finally, we assume that $m = P$ is a probability distribution on $\mathcal{P}(Q)$, the power set of the items in Q. Using these assumptions, our next result shows a connection between (3.27) and positive knowledge correlation.

Theorem 3.18. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$, $P(Q) > 0$, and $\mathcal{K} = \mathcal{K}_P$ is closed under both union and intersection. Suppose also that for every $K, L \in \mathcal{K}$ we have

$$P(K \cup L)P(K \cap L) \geq P(K)P(L). \tag{3.29}$$

Then, P satisfies the property of positive knowledge correlation.

Before proving Theorem 3.18, we first prove the result for a more specific and technical set of conditions.

Lemma 3.19. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$, $P(Q) > 0$, and $\mathcal{K} = \mathcal{K}_P$ is closed under both union and intersection. Suppose also that for every $B, C \subseteq Q$ and $K, L \in \mathcal{K}$ we have

$$P\left(\{K \cup L\} \cap I_B^+ \cap I_C^-\right) \cdot P\left(\{K \cap L\} \cap I_B^+ \cap I_C^-\right) \geq \\ P\left(\{K\} \cap I_B^+ \cap I_C^-\right) \cdot P\left(\{L\} \cap I_B^+ \cap I_C^-\right). \quad (3.30)$$

Then, P satisfies the property of positive knowledge correlation.

Proof. Let $q, r \in Q$, with B and C defined as above. For any $K \in \mathcal{K}$ define

$$m(K) := \begin{cases} P(K), & \text{if } K \in I_B^+ \cap I_C^- \\ 0, & \text{if } K \notin I_B^+ \cap I_C^-. \end{cases} \quad (3.31)$$

For a set family $\mathcal{F}$, let $\mathbf{1}_{\mathcal{F}}$ represent the indicator function for the set family $\mathcal{F}$; that is, for any $A \subseteq Q$ we have $\mathbf{1}_{\mathcal{F}}(A) = 1$ if $A \in \mathcal{F}$ and $\mathbf{1}_{\mathcal{F}}(A) = 0$ otherwise. We need to show that (3.8) holds. Note that, since $I_{\{q\}}^+$ is an upper set, $\mathbf{1}_{I_{\{q\}}^+}$ is a monotonically increasing function for any $q \in Q$. Next, observe that by combining (3.30) and (3.31) it follows that (3.27) is satisfied. Setting $f(K) = \mathbf{1}_{I_{\{q\}}^+}(K)$ and $g(K) = \mathbf{1}_{I_{\{r\}}^+}(K)$, from (3.28) we then get

$$\left(\sum_{K \in \mathcal{K}} \mathbf{1}_{I_{\{q\}}^+}(K) \cdot \mathbf{1}_{I_{\{r\}}^+}(K) \cdot m(K)\right)\left(\sum_{K \in \mathcal{K}} m(K)\right) \geq \\ \left(\sum_{K \in \mathcal{K}} \mathbf{1}_{I_{\{q\}}^+}(K) \cdot m(K)\right)\left(\sum_{K \in \mathcal{K}} \mathbf{1}_{I_{\{r\}}^+}(K) \cdot m(K)\right). \quad (3.32)$$

Starting with the first term of (3.32) we have

$$\begin{aligned} \sum_{K \in \mathcal{K}} \mathbf{1}_{I_{\{q\}}^+}(K) \cdot \mathbf{1}_{I_{\{r\}}^+}(K) \cdot m(K) &= \sum_{K \in I_{\{q\}}^+ \cap I_{\{r\}}^+} m(K) \\ &= \sum_{K \in I_{\{q\}}^+ \cap I_{\{r\}}^+ \cap I_B^+ \cap I_C^-} P(K) \qquad \text{by (3.31)} \\ &= P\left(I_{\{q\}}^+ \cap I_{\{r\}}^+ \cap I_B^+ \cap I_C^-\right). \end{aligned} \quad (3.33)$$

Using similar arguments, we can also show that

$$\begin{aligned}\sum_{K\in\mathcal{K}} \mathbf{1}_{I^+_{\{q\}}}(K)\cdot m(K) &= \sum_{K\in I^+_{\{q\}}} m(K)\\ &= \sum_{K\in I^+_{\{q\}}\cap I^+_B\cap I^-_C} P(K) \qquad \text{by (3.31)}\\ &= P\left(I^+_{\{q\}}\cap I^+_B\cap I^-_C\right) \end{aligned}\tag{3.34}$$

and

$$\begin{aligned}\sum_{K\in\mathcal{K}} \mathbf{1}_{I^+_{\{r\}}}(K)\cdot m(K) &= \sum_{K\in I^+_{\{r\}}} m(K)\\ &= \sum_{K\in I^+_{\{r\}}\cap I^+_B\cap I^-_C} P(K) \qquad \text{by (3.31)}\\ &= P\left(I^+_{\{r\}}\cap I^+_B\cap I^-_C\right). \end{aligned}\tag{3.35}$$

Finally, applying (3.31) once again we have

$$\begin{aligned}\sum_{K\in\mathcal{K}} m(K) &= \sum_{K\in I^+_B\cap I^-_C} P(K)\\ &= P\left(I^+_B\cap I^-_C\right). \end{aligned}\tag{3.36}$$

Combining (3.33)–(3.36) with (3.32) results in the inequality

$$P\left(I^+_{\{q\}}\cap I^+_{\{r\}}\cap I^+_B\cap I^-_C\right)\cdot P\left(I^+_B\cap I^-_C\right) \geq \\ P\left(I^+_{\{q\}}\cap I^+_B\cap I^-_C\right)\cdot P\left(I^+_{\{r\}}\cap I^+_B\cap I^-_C\right).$$

Assuming $P\left(I^+_{\{r\}}\cap I^+_B\cap I^-_C\right) > 0$, we can rearrange the terms to get

$$\begin{aligned} & \frac{P\left(I^+_{\{q\}}\cap I^+_{\{r\}}\cap I^+_B\cap I^-_C\right)}{P\left(I^+_{\{r\}}\cap I^+_B\cap I^-_C\right)} \geq \frac{P\left(I^+_{\{q\}}\cap I^+_B\cap I^-_C\right)}{P\left(I^+_B\cap I^-_C\right)}\\ \iff \quad & P\left(I^+_{\{q\}}\,\middle|\, I^+_{B\cup\{r\}}, I^-_C\right) \geq P\left(I^+_{\{q\}}\,\middle|\, I^+_B, I^-_C\right). \end{aligned}$$

Thus, (3.8) holds, as required.

□

We now prove Theorem 3.18.

Proof of Theorem 3.18. Let $B, C \subseteq Q$ and $K, L \in \mathcal{K}$. We need to show that (3.30) holds. To start, suppose that at least one of K or L is not in $\{I_B^+ \cap I_C^-\}$. In such a case, the right-hand side of (3.30) contains at least one probability computed over an empty set family, and by (3.6) it follows that such a probability is equal to zero—thus, the right-hand side equals zero and the inequality holds.

Next, consider the case when $K, L \in \{I_B^+ \cap I_C^-\}$. As such, both K and L contain all the items in B and none of the items in C. Let $b \in B$. Since $b \in K$ and $b \in L$, it follows that $b \in K \cup L$ and $b \in K \cap L$; as this holds for any $b \in B$, we have both $K \cup L \in I_B^+$ and $K \cap L \in I_B^+$. Next, let $c \in C$. Since $c \notin K$ and $c \notin L$, it follows that $c \notin K \cup L$ and $c \notin K \cap L$; as this holds for any $c \in C$, we have both $K \cup L \in I_C^-$ and $K \cap L \in I_C^-$. Combining the results in this paragraph, we have now shown that both $K \cup L$ and $K \cap L$ are contained in $\{I_B^+ \cap I_C^-\}$. We then have

$$\begin{aligned}
&P\left(\{K \cup L\} \cap I_B^+ \cap I_C^-\right) \cdot P\left(\{K \cap L\} \cap I_B^+ \cap I_C^-\right) \\
&\quad = P\left(K \cup L\right) \cdot P\left(K \cap L\right) \\
&\quad \geq P(K) \cdot P(L) \qquad \text{by (3.29)} \\
&\quad = P\left(\{K\} \cap I_B^+ \cap I_C^-\right) \cdot P\left(\{L\} \cap I_B^+ \cap I_C^-\right).
\end{aligned}$$

We have now shown that (3.30) holds in general. Applying Lemma 3.19, it follows that P satisfies the property of positive knowledge correlation. $\square$

Based on Theorem 3.18, we next give an example that satisfies the property of positive knowledge correlation.

Example 3.20. Let $\mathcal{K}$ be a knowledge structure on Q that is closed under both union and intersection, and let P be a uniform probability distribution on the states in $\mathcal{K}$; that is, we have

$$P(K) := \begin{cases} \dfrac{1}{|\mathcal{K}|}, & \text{if } K \in \mathcal{K} \\ 0, & \text{if } K \notin \mathcal{K}. \end{cases} \tag{3.37}$$

We need to show that (3.29) holds. Let K and L be states in $\mathcal{K}$. Since $\mathcal{K}$ is closed under both union and intersection, it follows that both $K \cup L$ and $K \cap L$ are states in $\mathcal{K}$. We then have

$$P(K) = P(L) = P(K \cup L) = P(K \cap L) = \frac{1}{|\mathcal{K}|}. \tag{3.38}$$

Based on these values, we can then see that (3.29) trivially holds—thus, by Theorem 3.18 it follows that P satisfies the property of positive knowledge correlation.

Our next result shows that, under the assumptions of Theorem 3.18, positive knowledge correlation and the conditions given by (3.29) are actually equivalent.

Theorem 3.21. Let Q be a set of items and P be a probability distribution on $\mathcal{P}(Q)$, where $P(\varnothing) > 0$ and $P(Q) > 0$. Furthermore, assume that P satisfies the property of positive knowledge correlation. Then, for any $K, L \in \mathcal{K}$, we have

$$P(K \cup L)P(K \cap L) \geq P(K)P(L). \tag{3.39}$$

Proof. We show the claimed result by proving the contrapositive holds. That is, if (3.39) fails to hold, we show that P does not satisfy the property of positive knowledge correlation. Thus, we start by assuming there exist $K, L \in \mathcal{K}$ such that

$$P(K \cup L)P(K \cap L) < P(K)P(L). \tag{3.40}$$

We next rewrite the terms as the following probabilities.

$$\begin{aligned} P(K \cup L) &= P\left(I^{+}_{K\cap L} \cap I^{-}_{(K\cup L)^{\mathsf{c}}} \cap I^{+}_{L\setminus K} \cap I^{+}_{K\setminus L}\right) > 0 \\ P(K \cap L) &= P\left(I^{+}_{K\cap L} \cap I^{-}_{(K\cup L)^{\mathsf{c}}} \cap I^{-}_{L\setminus K} \cap I^{-}_{K\setminus L}\right) > 0 \\ P(K) &= P\left(I^{+}_{K\cap L} \cap I^{-}_{(K\cup L)^{\mathsf{c}}} \cap I^{-}_{L\setminus K} \cap I^{+}_{K\setminus L}\right) > 0 \\ P(L) &= P\left(I^{+}_{K\cap L} \cap I^{-}_{(K\cup L)^{\mathsf{c}}} \cap I^{+}_{L\setminus K} \cap I^{-}_{K\setminus L}\right) > 0 \end{aligned}$$

Note that since (3.40) is a strict inequality, it implies $K \not\subseteq L$, $L \not\subseteq K$, $P(K) > 0$, and $P(L) > 0$. We can then derive the following conditional probabilities, which are well-defined as either K or L is contained in each set family represented in the denominators.

$$\frac{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^+_{L\setminus K}\cap I^+_{K\setminus L}\right)}{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^+_{K\setminus L}\right)} = P\left(I^+_{L\setminus K}\,\middle|\, I^+_{K}, I^-_{(K\cup L)^\mathsf{c}}\right)$$

$$\frac{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^-_{L\setminus K}\cap I^-_{K\setminus L}\right)}{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^-_{K\setminus L}\right)} = P\left(I^-_{L\setminus K}\,\middle|\, I^+_{K\cap L}, I^-_{L^\mathsf{c}}\right)$$

$$\frac{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^-_{L\setminus K}\cap I^+_{K\setminus L}\right)}{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^+_{K\setminus L}\right)} = P\left(I^-_{L\setminus K}\,\middle|\, I^+_{K}, I^-_{(K\cup L)^\mathsf{c}}\right)$$

$$\frac{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^+_{L\setminus K}\cap I^-_{K\setminus L}\right)}{P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^-_{K\setminus L}\right)} = P\left(I^+_{L\setminus K}\,\middle|\, I^+_{K\cap L}, I^-_{L^\mathsf{c}}\right)$$

Using these conditional probabilities and dividing each side of (3.40) by

$$P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^+_{K\setminus L}\right)\cdot P\left(I^+_{K\cap L}\cap I^-_{(K\cup L)^\mathsf{c}}\cap I^-_{K\setminus L}\right),$$

we can rewrite (3.40) as

$$P\left(I^+_{L\setminus K}\,\middle|\, I^+_{K}, I^-_{(K\cup L)^\mathsf{c}}\right)\cdot P\left(I^-_{L\setminus K}\,\middle|\, I^+_{K\cap L}, I^-_{L^\mathsf{c}}\right) < P\left(I^-_{L\setminus K}\,\middle|\, I^+_{K}, I^-_{(K\cup L)^\mathsf{c}}\right)\cdot P\left(I^+_{L\setminus K}\,\middle|\, I^+_{K\cap L}, I^-_{L^\mathsf{c}}\right). \quad (3.41)$$

From (3.41) we can see that at least one of the following (strict) inequalities must hold.

$$P\left(I^+_{L\setminus K}\,\middle|\, I^+_{K}, I^-_{(K\cup L)^\mathsf{c}}\right) < P\left(I^+_{L\setminus K}\,\middle|\, I^+_{K\cap L}, I^-_{L^\mathsf{c}}\right) \quad (3.42)$$

$$P\left(I^-_{L\setminus K}\,\middle|\, I^+_{K\cap L}, I^-_{L^\mathsf{c}}\right) < P\left(I^-_{L\setminus K}\,\middle|\, I^+_{K}, I^-_{(K\cup L)^\mathsf{c}}\right) \quad (3.43)$$

Note that both of these inequalities are in violation of the property of positive knowledge correlation—as such, the probability distribution P does not satisfy positive knowledge correlation, and the claimed result follows.

□

3.3. Empirical Analysis of Positive Knowledge Correlation

In this section we investigate the validity of the property of positive knowledge correlation using data from the ALEKS system. ALEKS, which stands for "**A**ssessment and **LE**arning in **K**nowledge **S**paces," is an artificially intelligent adaptive learning and assessment system that is based on KST (McGraw Hill ALEKS, 2021). For this analysis, we use a data set composed of 3,301,368 ALEKS assessments taken over a period of roughly 10 years, beginning in 2011. These assessments are from the ALEKS Placement, Preparation, and Learning (ALEKS PPL) product, where each assessment functions as a placement test for an incoming college or university student. The ALEKS PPL assessment uses a knowledge space consisting of 314 items that cover material from elementary mathematics to precalculus.

Recall that (3.8) has the form given below.

$$P\left(I_{\{q\}}^{+} \,\middle|\, I_{B\cup\{r\}}^{+}, I_{C}^{-}\right) \geq P\left(I_{\{q\}}^{+} \,\middle|\, I_{B}^{+}, I_{C}^{-}\right)$$

In order to estimate the above conditional probabilities, we make use of a specific feature of the ALEKS assessment. Each ALEKS assessment—ALEKS PPL included—asks an *extra problem* that is chosen uniformly at random from the available items, and the student's response to the extra problem does not have any effect on the outcome of the assessment or the selecting of the subsequent questions to ask. Instead, the data from the extra problem are typically used to evaluate the performance of the assessment and make further improvements. For this analysis, we use the correct answer rates on the extra problems to estimate the conditional probabilities in (3.8).

Next, in order to properly account for the conditional events in (3.8) when computing our estimates, we need to make use of the item classifications coming from the ALEKS assessment. Specifically, each ALEKS PPL assessment partitions the 314 items into the following three categories.

– items that are most likely in the student's knowledge state (in-state)

– items that are most likely not in the student's knowledge state (out-of-state)

– the remaining items (uncertain)

For this analysis, we use the in-state and out-of-state classifications on the items when conditioning on r, B, and C in (3.8). To estimate the right-hand-side of (3.8), we find all the assessments in which the items in B are all classified in-state and all the items in C are classified as out-of-state; for these assessments, we then compute the rate at which q is answered correctly when it appears as the extra problem. Next, to estimate the left-hand-side of (3.8), we find all the assessments in which the items in $B \cup \{r\}$ are all classified in-state, while the items in C are all classified as out-of-state; for these assessments, we again compute the rate at which q is answered correctly when it appears as the extra problem.

Before moving on, a comment must be made on our procedure for estimating the conditional probabilities. The reader may wonder why we have chosen to use the correct answer rate on the extra problem as our estimate, rather than the proportion of times that q is classified as in-state. The reason for this is that simultaneously using the assessment classification information for q, B, and C can, in many cases, heavily bias the estimates. For example, suppose that, due to the specifics of the knowledge space, q is a *prerequisite* of an item $b \in B$; specifically, q is in any state that contains b. In such a case, an estimate of $P\left(I^+_{\{q\}} \,\middle|\, I^+_B, I^-_C\right)$ that uses the in-state information on q would always return a value of one; that is, since q is a prerequisite of an item in B, whenever all the items in B are classified as in-state, q would by necessity also be classified as in-state. Thus, by computing the conditional probability estimates based on the correct answer rate to the extra problem, we are hoping to adjust for biases such as these. While it is true that, due to careless errors, a correct answer rate can be different from the probability of actually knowing an item, we submit that the correct answer rate is a reasonable proxy for this probability in the scenario under consideration.

Now, let $(Q, \mathcal{K})$ be the knowledge space used in the ALEKS PPL assessment. For each $q \in Q$, we use the overall correct answer rate for q to estimate $P(I^+_{\{q\}})$; the results are shown in Figure 1, where we display a relative frequency histogram of the correct answer rate for each item. The correct answer rates are computed with data from the responses to an average of 10,490 extra problems; the mean and median values of the correct answer rates are 0.52 and 0.50, respectively, with the values ranging from a minimum of 0.02 to a maximum of 0.98. For our first evaluation of the positive knowledge correlation property, we can compare the overall correct answer rates to those computed conditional on whether another item is classified as in-state or out-of-state. Formally, for any $q, r \in Q$, with $q \neq r$, we

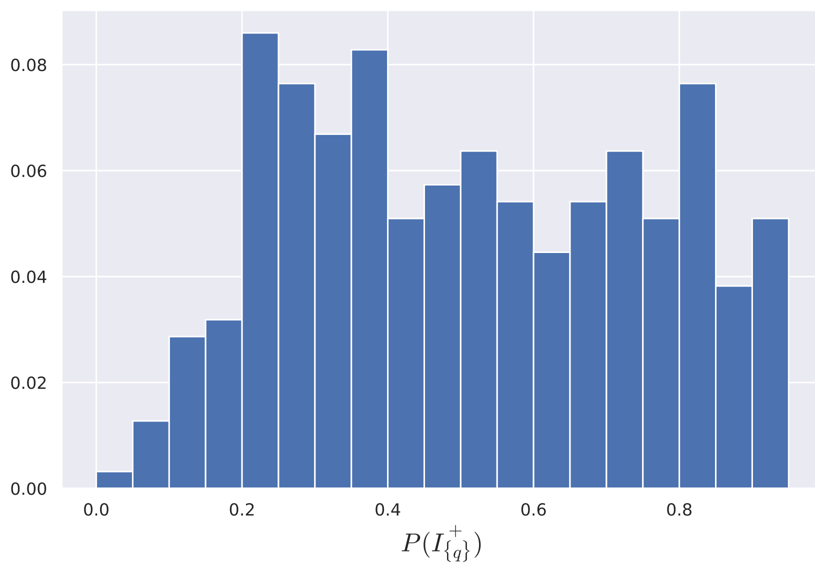

Figure 1: Relative frequency histogram of $P\left(I^{+}_{\{q\}}\right)$, as estimated by the extra problem correct answer rate for each of the 314 items. An average of 10,490 extra problems are used to compute each estimate.

compute the "positive" difference

$$P\left(I^{+}_{\{q\}} \,\middle|\, I^{+}_{\{r\}}\right) - P\left(I^{+}_{\{q\}}\right) \tag{3.44}$$

whenever the conditional probability can be estimated with at least 5,000 data points. Specifically, we find example assessments where r is classified as in-state and q is asked as the extra problem; based on the correct answer rate to q on these assessments, we have our estimate of $P\left(I^{+}_{\{q\}} \,\middle|\, I^{+}_{\{r\}}\right)$. Similarly, we also compute the "negative" difference

$$P\left(I^{+}_{\{q\}} \,\middle|\, I^{-}_{\{r\}}\right) - P\left(I^{+}_{\{q\}}\right) \tag{3.45}$$

whenever the conditional probability can be estimated with at least 5,000 data points. Specifically, we find example assessments where r is classified as out-of-state and q is asked as the extra problem; based on the correct answer rate to q on these assessments, we have our estimate of $P\left(I^{+}_{\{q\}} \,\middle|\, I^{-}_{\{r\}}\right)$.

Note that as (3.44) and (3.45) can be computed for each ordered pair of distinct items in the domain, there are a total of $314 \times 313 = 98,282$ conditional probabilities for us to consider in each of the two cases. Using all the assessments in our data set, we find 46,972 positive conditional probability estimates with sample sizes of 5,000 or more, and none of these are smaller than the corresponding unconditional probability $P\left(I^{+}_{\{q\}}\right)$. Then, there are 34,837 negative examples with sample sizes of 5,000 or more, and none of these have a value larger than the corresponding unconditional probability $P\left(I^{+}_{\{q\}}\right)$. The resulting relative frequency histograms for these estimates of (3.44) and (3.45) are shown in Figure 2.

Our next set of results looks at slightly more complex conditional probabilities. For $q, r, s \in Q$ we are interested in estimating the conditional probability $P\left(I^{+}_{\{q\}} \,\middle|\, I^{+}_{\{r\}}, I^{-}_{\{s\}}\right)$; that is, the probability of knowing q given that r is known and s is not known. To estimate this quantity, we look at all assessments where r is classified as in-state, s is classified as out-of-state, and q is asked as the extra problem. For our "positive" case we compare this estimate to the probability of knowing q given that s is not known, as follows:

$$P\left(I^{+}_{\{q\}} \,\middle|\, I^{+}_{\{r\}}, I^{-}_{\{s\}}\right) - P\left(I^{+}_{\{q\}} \,\middle|\, I^{-}_{\{s\}}\right). \tag{3.46}$$

The above difference is computed whenever we have 5,000 data points for each conditional probability estimate. Similarly, we can compute the "negative"

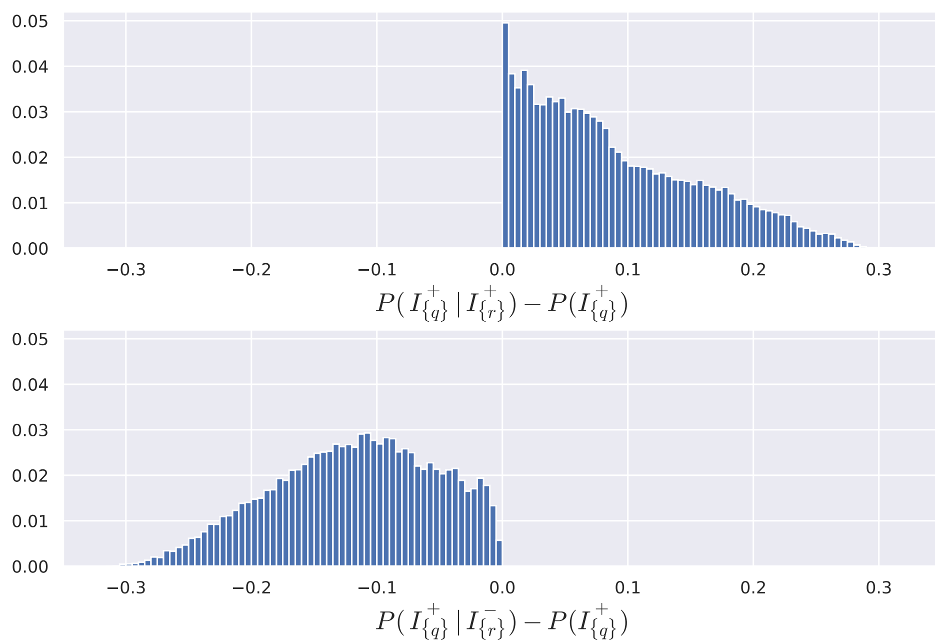

Figure 2: Comparison of the item base rates and conditional probabilities. The top graph shows a relative frequency histogram of the estimates for (3.44), while the bottom graph shows a relative frequency histogram of the estimates for (3.45). All of the 46,972+34,837=81,809 estimates are consistent with the property of positive knowledge correlation.

difference by comparing to the probability of knowing q, given that r is known:

$$P\left(I^+_{\{q\}} \,\middle|\, I^+_{\{r\}}, I^-_{\{s\}}\right) - P\left(I^+_{\{q\}} \,\middle|\, I^+_{\{r\}}\right). \tag{3.47}$$

As before, the above difference is computed whenever we have 5,000 data points for each conditional probability estimate.

As (3.46) and (3.47) can be computed for each ordered triple of distinct items in the domain, there are a total of $314 \times 313 \times 312 = 30,663,984$ conditional probability estimates to consider. The results are shown in Figure 3. In this example there are 2,249,553 triples of items $q, r, s \in Q$ for which we have at least 5,000 data points to estimate $P\left(I^+_{\{q\}} \,\middle|\, I^+_{\{r\}}, I^-_{\{s\}}\right)$. Comparing these values to the estimates for $P\left(I^+_{\{q\}} \,\middle|\, I^+_{\{r\}}\right)$, there are only 45 examples for which the difference (3.46) is negative, with all the values being greater than -0.0002; thus, the violations of (3.8) are very minor, both in number and size. Then, for the negative data points, there are only 75 for which the difference (3.47) is positive, with all of the values being less than 0.0003; thus, as before, the violations of positive knowledge correlation in these examples are again very minor both in number and size. Combining the results from all of these figures, overall the examples we've looked at are consistent with the property of positive knowledge correlation—the few specific cases that violated the property did so very minimally, suggesting that these violations may be due simply to uncertainty in our estimates.

4. Positive Correlation and Adaptive Assessments

In the previous section we looked in detail at the property of positive knowledge correlation. However, for all of our previous analyses we assumed that the probability distribution on the states in $\mathcal{K}$ was fixed and unchanging. In this section we look at a related, but distinct, problem. Specifically, we assume that the probability distribution on the states in $\mathcal{K}$ is changing over time—in particular, this is the behavior that results from an adaptive assessment algorithm that updates the distribution on $\mathcal{K}$ with each response from the student taking the assessment. After a brief background on how such an updating algorithm works, we introduce and analyze the concept of a *positively correlated updating rule.*

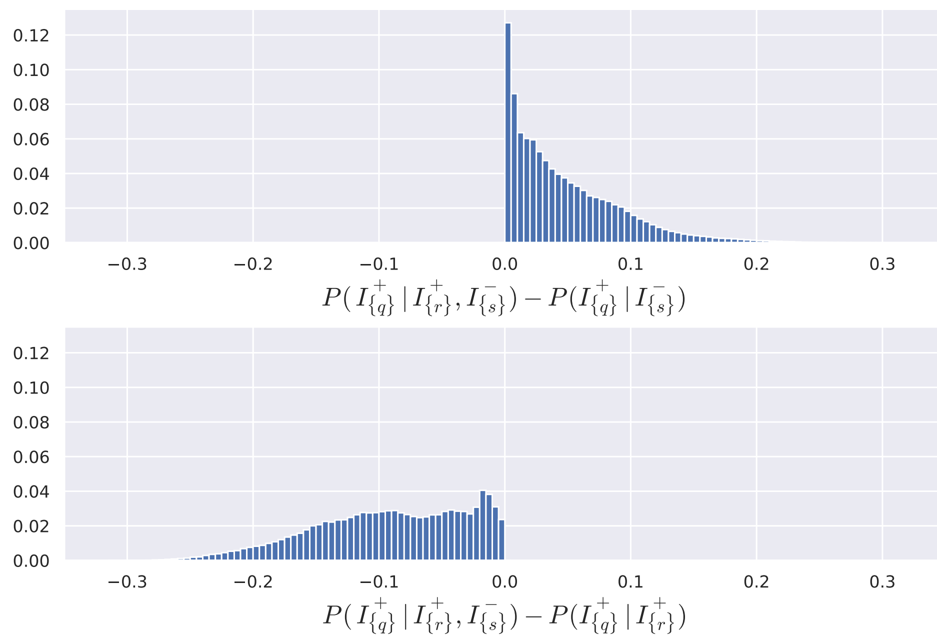

Figure 3: The top graph shows a relative frequency histogram of the estimates for (3.46), where the minimum value is greater than -0.0002 and only 45 of the 2,249,553 data points have negative values. The bottom graph shows a relative frequency histogram of the estimates for (3.47), where the maximum value is less than 0.0003 and only 75 of the 2,249,553 data points have positive values.

4.1. Updating Rule for an Adaptive Assessment

We next discuss a few concepts related to probabilistically modeling and assessing knowledge in KST. This material is adapted mainly from Chapters 11 and 13 in Falmagne and Doignon (2011). We begin with the definition of a *probabilistic knowledge structure*.

Definition 4.1. A *probabilistic knowledge structure* is a triple $(Q, \mathcal{K}, P)$ that satisfies the following.

(i) $(Q, \mathcal{K})$ is a finite knowledge structure with $\cup\mathcal{K} \in \mathcal{K}$.

(ii) The mapping $P : \mathcal{K} \to [0, 1] : K \mapsto P(K)$ is a probability distribution on $\mathcal{K}$; that is, for any $K \in \mathcal{K}$ we have $P(K) \geq 0$ and, additionally, $\sum_{K \in \mathcal{K}} P(K) = 1$.

Note that the above definition has a subtle difference from our previous assumptions in Section 3. In particular, Definition 4.1 assumes a knowledge structure, $\mathcal{K}$, exists, and that the probability distribution is then a function of the states in $\mathcal{K}$. However, in Section 3 our typical assumption was that the knowledge structure was induced by the probability distribution, which implies that $P(K) > 0$ for any $K \in \mathcal{K}$; as such, Definition 4.1 is slightly more general, as $P(K) = 0$ is allowed even if $K \in \mathcal{K}$.

Next, we need to define the concept of an *updating rule*.

Definition 4.2. For $n = 1, 2, \ldots$, let (q_n, r_n) represent a sequence of questions and responses, respectively, that appear during an adaptive assessment. That is, $q_n \in Q$ is the item asked at time n, while $r_n \in \{0, 1\}$ represents the student's response to q_n; note that $r_n = 1$ represents a correct answer to q_n, while $r_n = 0$ signifies a wrong answer. Assume that $(Q, \mathcal{K}, P_n)$ is a probabilistic knowledge structure for each n. Then, an updating rule u is a function satisfying the equation

$$P_{n+1} \stackrel{a.s.}{=} u(r_n, q_n, P_n). \tag{4.1}$$

For a state $K \in \mathcal{K}$ and item $q \in Q$, let $\mathbf{1}_K$ be the indicator function for K, where $\mathbf{1}_K(q)$ is one if $q \in K$ and zero otherwise. Then, it's additionally assumed that u satisfies the following:

$$P_{n+1}(K) = u_K(r_n, q_n, P_n) \begin{cases} > P_n(K) & \text{if } \mathbf{1}_K(q_n) = r_n, \\ < P_n(K) & \text{if } \mathbf{1}_K(q_n) \neq r_n. \end{cases} \tag{4.2}$$

In words, an updating rule increases the probability of (a) any state that contains a correctly answered item, or (b) any state that doesn't contain an incorrectly answered item. Conversely, the updating rule decreases the probability of (c) any state that contains an incorrectly answered item, or (d) any state that doesn't contain a correctly answered item.

4.2. Positively Correlated Updating Rule

We next introduce the concept of a *positively correlated updating rule.*

Definition 4.3. For each $n = 1, 2, \ldots,$ let $(Q, \mathcal{K}, P_n)$ be a probabilistic knowledge structure. Then, $P_n(I^+_{\{a\}})$ is the probability of knowing item $a \in Q$ at question n. A positively correlated updating rule satisfies the following two inequalities:

$$\begin{aligned} r_n = 1 &\implies P_{n+1}\left(I^+_{\{a\}}\right) \geq P_n\left(I^+_{\{a\}}\right), \forall a \in Q, \\ r_n = 0 &\implies P_{n+1}\left(I^+_{\{a\}}\right) \leq P_n\left(I^+_{\{a\}}\right), \forall a \in Q. \end{aligned} \tag{4.3}$$

So, given a correct answer a positively correlated updating rule does not decrease the probability of any individual item; conversely, given an incorrect answer, it does not increase the probability of any individual item. Our goal is to develop a set of sufficient conditions that guarantees an updating rule is positively correlated.

For this analysis, we assume that we have a *multiplicative updating rule*, which we define as follows.

Definition 4.4. Let P_n be a probability distribution on a knowledge structure $\mathcal{K}$ at question n. Given real-valued parameters $\beta^i_n > 1$, $i \in \{0, 1\}$, define sets U^i_n as follows.

$$U^i_n = \left\{ K \in \mathcal{K} \,\middle|\, P_n(K) \to P_{n+1}(K) := \frac{\beta^i_n \cdot P_n(K)}{\beta^i_n \cdot P_n(U^i_n) + 1 - P_n(U^i_n)} \right\} \tag{4.4}$$

We call this a *multiplicative updating rule.* In the event of a correct answer to q_n (i.e., $r_n = 1$), the set U^1_n consists of all the states where the update β^1_n is applied. Then, given an incorrect answer to q_n (i.e., $r_n = 0$), the set U^0_n consists of all the states where the update β^0_n is applied. For any $K \notin U^i_n$, the updated probability is given by

$$P_n(K) \to P_{n+1}(K) := \frac{P_n(K)}{\beta^i_n \cdot P_n(U^i_n) + 1 - P_n(U^i_n)}. \tag{4.5}$$

Our next result shows that for any state $K \in U_n^i$, the updated probability is non-decreasing.

Lemma 4.5. Let u be a multiplicative updating rule with U_n^i defined as in (4.4) and $\beta_n^i > 1$. Suppose that $r_n = i$. Then, $P_{n+1}(K) \geq P_n(K)$ for $K \in U_n^i$ and $P_{n+1}(K) \leq P_n(K)$ for $K \in \mathcal{K} \setminus U_n^i$.

Proof. For $K \in U_n^i$ we have

$$\begin{aligned} P_{n+1}(K) &= \frac{\beta_n^i \cdot P_n(K)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\ &= \frac{P_n(K)}{P_n(U_n^i) + \frac{1 - P_n(U_n^i)}{\beta_n^i}} \\ &\geq P_n(K), \end{aligned}$$

where the last line follows from the fact that for $\beta_n^i > 1$ we have

$$P_n(U_n^i) + \frac{1 - P_n(U_n^i)}{\beta_n^i} \leq 1.$$

Next, for $K \notin U_n^i$ we have

$$\begin{aligned} P_{n+1}(K) &= \frac{P_n(K)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\ &\leq P_n(K), \end{aligned}$$

where the last line follows from the fact that for $\beta_n^i > 1$ we have

$$\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i) \geq 1.$$

□

For our subsequent results, we once again need to make use of the FKG inequality. As in Section 3.2, we let $\Gamma = \mathcal{K}$, where $\mathcal{K}$ is a knowledge structure defined on a set of items Q, and we assume that $\mathcal{K}$ is closed under both union and intersection. Finally, we assume that $m = P$ is a probability distribution on $\mathcal{P}(Q)$. In order to apply the FKG inequality, for any $K, L \in \mathcal{K}$ we need to show that the following holds after a multiplicative update.

$$P\left(K \cup L\right) P\left(K \cap L\right) \geq P(K)P(L) \tag{4.6}$$

The next lemma formulates a set of sufficient conditions that guarantees the inequality in (4.6) holds.

Lemma 4.6. Let u be a multiplicative updating rule with U_n^i defined as in (4.4), and assume that (4.6) currently holds for P_n and all $A, B \in \mathcal{K}$. Then, for $\beta_n^i > 1$ (4.6) continues to hold for P_{n+1} if u satisfies the following conditions.

(a) $A, B \in U_n^i \implies A \cup B, A \cap B \in U_n^i$

(b) $A \in U_n^i \implies \forall B \in \mathcal{K}$, at least one of $A \cup B$ or $A \cap B$ must be in U_n^i

Proof. Let $A, B \in \mathcal{K}$, and assume that (4.6) holds for P_n. Assume both A and B are in U_n^i. Then, by (a) we know $A \cup B$ and $A \cap B$ are both in U_n^i as well; thus, we have

$$\begin{aligned}
P_{n+1}(A \cup B)P_{n+1}(A \cap B) &= \frac{\beta_n^i \cdot P_n(A \cup B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \frac{\beta_n^i \cdot P(A \cap B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\
&\geq \frac{\beta_n^i \cdot P_n(A)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \frac{\beta_n^i \cdot P_n(B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\
&\quad \text{(by (4.6))} \\
&= P_{n+1}(A)P_{n+1}(B).
\end{aligned}$$

Next, without loss of generality assume that only A is in U. By (b) we know at least one of $A \cup B$ or $A \cap B$ is in U_n^i. Thus, we have

$$\begin{aligned}
P_{n+1}(A \cup B)P_{n+1}(A \cap B) &\geq \beta_n^i \frac{P_n(A \cup B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \frac{P_n(A \cap B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\
&\quad \text{(Since at least one of } A \cup B \text{ or } A \cap B \text{ is in } U_n^i) \\
&\geq \beta_n^i \frac{P_n(A)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \frac{P_n(B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\
&\quad \text{(by (4.6))} \\
&= P_{n+1}(A)P_{n+1}(B),
\end{aligned}$$

where the last line follows from the fact that only A is in U_n^i.

Finally, assume that neither A nor B is contained in U_n^i. Note that for an arbitrary $K \in \mathcal{K}$, combining (4.4) and (4.5) with the fact that $\beta_n^i > 1$ gives

$$P_{n+1}(K) \geq \frac{P_n(K)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)}.$$

We then have

$$
\begin{aligned}
P_{n+1}(A \cup B) P_{n+1}(A \cap B) &\geq \frac{P_n(A \cup B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \frac{P_n(A \cap B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\
&\geq \frac{P_n(A)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \frac{P_n(B)}{\beta_n^i \cdot P_n(U_n^i) + 1 - P_n(U_n^i)} \\
&\quad \text{(by (4.6))} \\
&= P_{n+1}(A) P_{n+1}(B),
\end{aligned}
$$

where the last line follows from the fact that we're assuming A and B are not in U_n^i. Thus, we have now shown that (4.6) holds for P_{n+1}. □

One specific updating rule that satisfies (a) and (b) from Lemma 4.6 is the following. In the event of a correct response to an item q, set $U_n^1 = I_{\{q\}}^+$; otherwise, if q is answered incorrectly set $U_n^0 = I_{\{q\}}^-$. The intuition is that if q is answered correctly we want to increase the probabilities of all the states that contain q; conversely, if q is answered incorrectly we want to increase the probabilities of all the states that do not contain q. Such an updating rule is discussed in Section 13.4 of Falmagne and Doignon (2011). More generally, the conditions (a) and (b) are satisfied if, for example, every set in U_n^i contains all the items from some set R. Or, as another example, the conditions are satisfied if every set in U_n^i contains no items from some set R.

We next identify another set of conditions that must be satisfied in order to have a positively correlated updating rule. To do this, we first need to prove the following lemma.

Lemma 4.7. Let u be a multiplicative updating rule with U_n^i defined as in (4.4), and suppose that $r_n = i$. Then, for any $q \in Q$ we have the following.

$$P_{n+1}\left(I_{\{q\}}^+ \,\middle|\, U_n^i\right) = P_n\left(I_{\{q\}}^+ \,\middle|\, U_n^i\right) \tag{4.7}$$

$$P_{n+1}\left(I_{\{q\}}^+ \,\middle|\, \left(U_n^i\right)^{\mathsf{c}}\right) = P_n\left(I_{\{q\}}^+ \,\middle|\, \left(U_n^i\right)^{\mathsf{c}}\right) \tag{4.8}$$

That is, the probability of knowing an item, conditioned on either U_n^i or $\left(U_n^i\right)^{\mathsf{c}}$, is not affected by the update.

Proof. We first prove (4.7).

$$\begin{aligned}
P_{n+1}\left(I_{\{q\}}^{+} \,\middle|\, U_n^i\right) &= \frac{P_{n+1}\left(I_{\{q\}}^{+} \cap U_n^i\right)}{P_{n+1}\left(U_n^i\right)} \\
&= \frac{\sum_{K \in \{I_{\{q\}}^{+} \cap U_n^i\}} \beta_n^i P_n(K)}{\sum_{K \in U_n^i} \beta_n^i P_n(K)} \\
&= \frac{\sum_{K \in \{I_{\{q\}}^{+} \cap U_n^i\}} P_n(K)}{\sum_{K \in U_n^i} P_n(K)} \\
&= \frac{P_n\left(I_{\{q\}}^{+} \cap U_n^i\right)}{P_n\left(U_n^i\right)} \\
&= P_n\left(I_{\{q\}}^{+} \,\middle|\, U_n^i\right) \qquad (4.9)
\end{aligned}$$

Using a similar argument, we next prove (4.8).

$$\begin{aligned}
P_{n+1}\left(I_{\{q\}}^{+} \,\middle|\, \left(U_n^i\right)^{\mathsf{c}}\right) &= \frac{P_{n+1}\left(I_{\{q\}}^{+} \cap \left(U_n^i\right)^{\mathsf{c}}\right)}{P_{n+1}\left(\left(U_n^i\right)^{\mathsf{c}}\right)} \\
&= \frac{\sum_{K \in \{I_{\{q\}}^{+} \cap (U_n^i)^{\mathsf{c}}\}} P_n(K)}{\sum_{K \in (U_n^i)^{\mathsf{c}}} P_n(K)} \\
&= \frac{P_n\left(I_{\{q\}}^{+} \cap \left(U_n^i\right)^{\mathsf{c}}\right)}{P_n\left(\left(U_n^i\right)^{\mathsf{c}}\right)} \\
&= P_n\left(I_{\{q\}}^{+} \,\middle|\, \left(U_n^i\right)^{\mathsf{c}}\right) \qquad (4.10)
\end{aligned}$$

□

In this next theorem, we show that a positively correlated updating rule is obtained by adding a final pair of assumptions on u.

Theorem 4.8. Let u be a multiplicative updating rule with U_n^i and β_n^i defined as in (4.4). Assume that u satisfies the conditions in Lemma 4.6. Then, if $\mathbf{1}_{U_n^1}$ is monotonically increasing and $\mathbf{1}_{U_n^0}$ is monotonically decreasing, it follows that u is a positively correlated updating rule.

Proof. For u to be positively correlated, the difference $P_{n+1}\left(I^+_{\{a\}}\right) - P_n\left(I^+_{\{a\}}\right)$ must satisfy the following two inequalities:

$$r_n = 1 \implies P_{n+1}\left(I^+_{\{a\}}\right) - P_n\left(I^+_{\{a\}}\right) \geq 0, \forall a \in Q, \tag{4.11}$$

$$r_n = 0 \implies P_{n+1}\left(I^+_{\{a\}}\right) - P_n\left(I^+_{\{a\}}\right) \leq 0, \forall a \in Q. \tag{4.12}$$

Using Lemma 4.7, we can rewrite $P_{n+1}\left(I^+_{\{a\}}\right) - P_n\left(I^+_{\{a\}}\right)$ as follows.

$$\begin{aligned}
&P_{n+1}\left(I^+_{\{a\}}\right) - P_n\left(I^+_{\{a\}}\right) \\
&\quad = P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) P_{n+1}\left(U^i_n\right) + P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) P_{n+1}\Big(\left(U^i_n\right)^{\mathsf{c}}\Big) \\
&\qquad - P_n\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) P_n\left(U^i_n\right) - P_n\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) P_n\Big(\left(U^i_n\right)^{\mathsf{c}}\Big) \\
&\quad = P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) P_{n+1}\left(U^i_n\right) + P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) P_{n+1}\Big(\left(U^i_n\right)^{\mathsf{c}}\Big) \\
&\qquad - P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) P_n\left(U^i_n\right) - P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) P_n\Big(\left(U^i_n\right)^{\mathsf{c}}\Big) \\
&\qquad \text{(by Lemma 4.7)} \\
&\quad = P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) \left(P_{n+1}\left(U^i_n\right) - P_n\left(U^i_n\right)\right) \\
&\qquad + P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) \left(P_{n+1}\Big(\left(U^i_n\right)^{\mathsf{c}}\Big) - P_n\Big(\left(U^i_n\right)^{\mathsf{c}}\Big)\right) \\
&\quad = P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) \left(P_{n+1}\left(U^i_n\right) - P_n\left(U^i_n\right)\right) \\
&\qquad + P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) \left(1 - P_{n+1}\left(U^i_n\right) - 1 + P_n\left(U^i_n\right)\right) \\
&\quad = \left(P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) - P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right)\right) \left(P_{n+1}\left(U^i_n\right) - P_n\left(U^i_n\right)\right)
\end{aligned}$$

Note that $P_{n+1}\left(U^i_n\right) - P_n\left(U^i_n\right)$ is positive for $\beta^i_n > 1$; this is easily seen from the updating formula in (4.4). So, (4.11) holds when $P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) - P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) \geq 0$, while (4.12) holds if $P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^i_n\right) - P_{n+1}\left(I^+_{\{a\}} \,\middle|\, \left(U^i_n\right)^{\mathsf{c}}\right) \leq 0$.

Suppose $r_n = 1$. Since $\mathbf{1}_{U^1_n}$ is monotonically increasing, setting $f = \mathbf{1}_{I^+_{\{a\}}}$,

$g = \mathbf{1}_{U_n^1}$, and $m = P_{n+1}$, we can then apply the FKG inequality to get

$$\left(\sum_{K\in\mathcal{K}} \mathbf{1}_{I^+_{\{a\}}}(K)\mathbf{1}_{U_n^1}(K)P_{n+1}(K)\right)\left(\sum_{K\in\mathcal{K}} P_{n+1}(K)\right) \geq \left(\sum_{K\in\mathcal{K}} \mathbf{1}_{I^+_{\{a\}}}(K)P_{n+1}(K)\right)\left(\sum_{K\in\mathcal{K}} \mathbf{1}_{U_n^1}(K)P_{n+1}(K)\right). \tag{4.13}$$

Note that since $\sum_{K\in\mathcal{K}} P_{n+1}(K) = 1$, the the left-hand side of (4.13) simplifies to

$$\sum_{K\in I^+_{\{a\}}\cap U_n^1} P_{n+1}(K) = P_{n+1}\left(I^+_{\{a\}} \cap U_n^1\right).$$

Then, since the right-hand side of (4.13) simplifies to

$$\Bigg(\sum_{K\in I^+_{\{a\}}} P_{n+1}(K)\Bigg)\Bigg(\sum_{K\in U_n^1} P_{n+1}(K)\Bigg) = P_{n+1}\left(I^+_{\{a\}}\right)P_{n+1}\Big(U_n^1\Big),$$

after rearranging terms we have

$$P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U_n^1\right) = \frac{P_{n+1}\left(I^+_{\{a\}} \cap U_n^1\right)}{P_{n+1}\left(U_n^1\right)} \geq P_{n+1}\left(I^+_{\{a\}}\right). \tag{4.14}$$

Next, since $\mathbf{1}_{U_n^1}$ is monotonically increasing, we have that $\mathbf{1}_{(U_n^1)^\mathsf{c}} = 1 - \mathbf{1}_{U_n^1}$ is monotonically decreasing. Applying the FKG inequality once again with $f = \mathbf{1}_{I^+_{\{a\}}}$, $g = \mathbf{1}_{(U_n^1)^\mathsf{c}}$, and $m = P_{n+1}$ gives

$$\left(\sum_{K\in\mathcal{K}} \mathbf{1}_{I^+_{\{a\}}}(K)\mathbf{1}_{(U_n^1)^\mathsf{c}}(K)P_{n+1}(K)\right)\left(\sum_{K\in\mathcal{K}} P_{n+1}(K)\right) \leq \left(\sum_{K\in\mathcal{K}} \mathbf{1}_{I^+_{\{a\}}}(K)P_{n+1}(K)\right)\left(\sum_{K\in\mathcal{K}} \mathbf{1}_{(U_n^1)^\mathsf{c}}(K)P_{n+1}(K)\right). \tag{4.15}$$

Simplifying terms and rearranging, we end up with

$$P_{n+1}\left(I^+_{\{a\}} \,\middle|\, (U_n^1)^\mathsf{c}\right) = \frac{P_{n+1}\left(I^+_{\{a\}} \cap (U_n^1)^\mathsf{c}\right)}{P_{n+1}\Big((U_n^1)^\mathsf{c}\Big)} \leq P_{n+1}\left(I^+_{\{a\}}\right). \tag{4.16}$$

Combining equations (4.14) and (4.16), we get

$$P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^1_n\right) - P_{n+1}\left(I^+_{\{a\}} \,\middle|\, (U^1_n)^{\mathsf{c}}\right) \geq 0;$$

thus, as previously discussed, (4.11) holds.

Next, assume $r_n = 0$. In this case $\mathbf{1}_{U^0_n}$ is monotonically decreasing, while $\mathbf{1}_{(U^0_n)^{\mathsf{c}}} = 1 - \mathbf{1}_{U^0_n}$ is monotonically increasing. Applying the FKG inequality with $f = \mathbf{1}_{I^+_{\{a\}}}$, $g = \mathbf{1}_{(U^0_n)^{\mathsf{c}}}$, and $m = P_{n+1}$ gives

$$P_{n+1}\left(I^+_{\{a\}} \,\middle|\, (U^0_n)^{\mathsf{c}}\right) = \frac{P_{n+1}\left(I^+_{\{a\}} \cap (U^0_n)^{\mathsf{c}}\right)}{P_{n+1}\left((U^0_n)^{\mathsf{c}}\right)} \geq P_{n+1}\left(I^+_{\{a\}}\right), \tag{4.17}$$

while using $f = \mathbf{1}_{I^+_{\{a\}}}$, $g = \mathbf{1}_{U^0_n}$, and $m = P_{n+1}$ gives

$$P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^0_n\right) = \frac{P_{n+1}\left(I^+_{\{a\}} \cap U^0_n\right)}{P_{n+1}\left(U^0_n\right)} \leq P_{n+1}\left(I^+_{\{a\}}\right). \tag{4.18}$$

Combining equations (4.17) and (4.18), we have

$$P_{n+1}\left(I^+_{\{a\}} \,\middle|\, U^0_n\right) - P_{n+1}\left(I^+_{\{a\}} \,\middle|\, (U^0_n)^{\mathsf{c}}\right) \leq 0,$$

from which (4.12) follows. □

Based on Theorem 4.8, we get the following corollary, which shows the connection between positive knowledge correlation and a positively correlated updating rule.

Corollary 4.9. Let Q be a set of items and P_n be a probability distribution on $\mathcal{P}(Q)$, where $P_n(\varnothing) > 0$ and $P_n(Q) > 0$. Assume that P_n satisfies the property of positive knowledge correlation. Let u be a multiplicative updating rule with U^i_n and β^i_n defined as in (4.4), and assume that (a) and (b) from Lemma 4.6 are satisfied. Then, if $\mathbf{1}_{U^1_n}$ is monotonically increasing and $\mathbf{1}_{U^0_n}$ is monotonically decreasing, it follows that u is a positively correlated updating rule.

Proof. Theorem 3.21 tells us that (4.6) holds for any $K, L \in \mathcal{K}$. Thus, the requirements for Lemma 4.6 are satisfied, and the result then follows from Theorem 4.8. □

We next give an example of a positively correlated updating rule.

Example 4.10. Let $\mathcal{K}$ be a knowledge structure that is closed under both union and intersection. Let P_0 be defined as in (3.37); that is, P_0 is a uniform probability distribution on the states in $\mathcal{K}$. As such, it's easy to see that (4.6) holds for P_0. Let q_n and r_n represent the item asked and response given, respectively, at time n, and define our updating rule as follows.

$$U_n^1 = I^+_{\{q_n\}}$$
$$U_n^0 = I^-_{\{q_n\}}$$

That is, in the event of a correct answer to q_n all the states containing q_n are updated; conversely, in the event of an incorrect answer all the states that do not contain q_n are updated. We first note that it's easily checked that (a) and (b) from Lemma 4.6 hold for both $I^+_{\{q_n\}}$ and $I^-_{\{q_n\}}$.

Next, observe that $\mathbf{1}_{U_n^1} = \mathbf{1}_{I^+_{\{q_n\}}}$ is monotonically increasing as a function on $\mathcal{K}$, while $\mathbf{1}_{U_n^0} = \mathbf{1}_{I^-_{\{q_n\}}}$ is monotonically decreasing as a function on $\mathcal{K}$. Thus, from Theorem 4.8 it follows that the updating rule is positively correlated.

In the remainder of this section we look at the specific case of an updating rule on a non-ordinal learning space—i.e., a learning space that is not closed under intersection. In doing so, we need to make use of the following definition.

Definition 4.11. Let Q be a nonempty set and let $\mathcal{F}$ be a family of of subsets of Q. For an item $q \in Q$, an *atom at* q is a minimal set—where 'minimal' is defined with respect to set inclusion—of $\mathcal{F}$ containing q. A set $X \in \mathcal{F}$ is called an *atom* if it is an atom at q for some $q \in Q$.

The following theorem gives a useful property of atoms (for a proof see, for example, Theorem 5.4.1 in Falmagne and Doignon, 2011).

Theorem 4.12. Let $\mathcal{K}$ be a well-graded knowledge space on a set of items Q. Then, for any atom B at q, where $q \in Q$, the set $B \setminus \{q\}$ is a state.

We are now ready to prove our next result, which shows that a non-ordinal learning space is not a distributive lattice—as such, the techniques developed in this section are not directly applicable to non-ordinal learning spaces.

Lemma 4.13. Any non-ordinal learning space $\mathcal{L}$ is not a distributive lattice.

Proof. Since $\mathcal{L}$ is a non-ordinal learning space, there exist $Y, Z \in \mathcal{L}$ such that $Y \cap Z \notin \mathcal{L}$. Note that if $Y \wedge Z$ does not exist in $\mathcal{L}$, then $\mathcal{L}$ is not a lattice; thus, assume that $Y \wedge Z$ exists for each $Y, Z \in \mathcal{L}$. We want to show that the operations of "meet" (i.e., "set union") and "join" do not distribute. To do this, we need to find $X \in \mathcal{L}$ such that

$$X \cup (Y \wedge Z) \neq (X \cup Y) \wedge (X \cup Z). \tag{4.19}$$

To that end, we first note that since $Y \cap Z \notin \mathcal{L}$, it must be the case that $Y \wedge Z \subset Y \cap Z$; thus, it follows that $C = (Y \cap Z) \setminus (Y \wedge Z)$ is non-empty. We next claim that, without loss of generality, for at least one $q \in C$ there exists an atom at q, A, such that $A \in \mathcal{L}$ and $A \subseteq Y$, but $A \nsubseteq Z$. We show this by contradiction. That is, for each $q \in C$ and for every atom at q, A_q, suppose we have $A_q \in \mathcal{L}$, $A_q \subseteq Y$, and $A_q \subseteq Z$. This means that

$$Y \cap Z = (Y \wedge Z) \bigcup_{q \in C} A_q \in \mathcal{L},$$

contradicting the assumption that $Y \cap Z \notin \mathcal{L}$. Thus, we can now assume that such an A exists.

Next, let $X = A \setminus \{q\}$; by Theorem 4.12, $X \in \mathcal{L}$. Observe that $q \notin X \cup (Y \wedge Z)$, as $q \in C = (Y \cap Z) \setminus (Y \wedge Z)$. Note, however, that since $q \in Z$ we have $A = X \cup \{q\} \subseteq X \cup Z$. Combined with the fact that $A \subseteq Y$, it follows that $A \subseteq (X \cup Y) \wedge (X \cup Z)$. Thus, we have now shown that $q \in (X \cup Y) \wedge (X \cup Z)$ but $q \notin X \cup (Y \wedge Z)$, from which (4.19) follows. $\square$

On the one hand, the preceding result shows that we cannot use the results from this section to decide if an updating rule on a non-ordinal learning space is positively correlated. On the other hand, it's still possible that an updating rule on a non-ordinal learning space can be positively correlated under certain conditions. However, our next example shows that any such hypothetical set of conditions must necessarily be different from the ones developed in this section for distributive lattices.

Example 4.14. Consider the following family of sets defined on $Q = \{a, b, c\}$.

$$\mathcal{K} = \{\varnothing, \{a\}, \{b\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

Note that this family is a learning space, as it can be checked that it satisfies both [LS] and [LC]. However, as $\{a,c\} \cap \{b,c\} = \{c\} \notin \mathcal{K}$, it is not intersection-closed; thus, $\mathcal{K}$ is a non-ordinal learning space. Suppose P_0 is a uniform probability distribution on the states in $\mathcal{K}$; that is, $P_0(K) = \frac{1}{7}$ for each $K \in \mathcal{K}$. We then have

$$\begin{aligned} P_0(I_b^+) &= P_0(\{b\}) + P_0(\{a,b\}) + P_0(\{b,c\}) + P_0(\{a,b,c\}) \\ &= \frac{4}{7} \end{aligned}$$

and

$$\begin{aligned} P_0(I_a^+) &= P_0(\{a\}) + P_0(\{a,b\}) + P_0(\{a,c\}) + P_0(\{a,b,c\}) \\ &= \frac{4}{7}. \end{aligned}$$

Now, suppose we apply an updating rule with parameter $\beta_n^1 = \beta_1^1 > 1$ and $U_n^1 = U_1^1 = I_a^+$; that is, assume that a is answered correctly. As we've seen previously, such an updating rule is positively correlated on a knowledge space closed under union and intersection. For any state $K \in \mathcal{K}$ such that $a \in K$ we have

$$\begin{aligned} P_1(K) &= \frac{\beta_1^1 \cdot P_0(K)}{\beta_1^1 \cdot P_0(I_a^+) + 1 - P_0(I_a^+)} \\ &= \frac{\frac{\beta_1^1}{7}}{\beta_1^1 \cdot \frac{4}{7} + \frac{3}{7}} \\ &= \frac{\beta_1^1}{4\beta_1^1 + 3}, \end{aligned}$$

while for any state $L \in \mathcal{K}$ such that $a \notin L$ we get

$$\begin{aligned} P_1(K) &= \frac{P_0(L)}{\beta_1^1 \cdot P_0(I_a^+) + 1 - P_0(I_a^+)} \\ &= \frac{\frac{1}{7}}{\beta_1^1 \cdot \frac{4}{7} + \frac{3}{7}} \\ &= \frac{1}{4\beta_1^1 + 3}. \end{aligned}$$

We then have

$$
\begin{aligned}
P_1(I_b^+) &= P_1(\{b\}) + P_1(\{a,b\}) + P_1(\{b,c\}) + P_1(\{a,b,c\}) \\
&= 2 \cdot \frac{1}{4\beta_1^1 + 3} + 2 \cdot \frac{\beta_1^1}{4\beta_1^1 + 3} \\
&= \frac{2\beta_1^1 + 2}{4\beta_1^1 + 3},
\end{aligned}
\tag{4.20}
$$

which is a strictly decreasing function of β_1^1. Since (4.20) equals $\frac{4}{7}$ for $\beta_1^1 = 1$, it follows that for any $\beta_1^1 > 1$ we have

$$P_1(I_b^+) < \frac{4}{7} = P_0(I_b^+).$$

Thus, the updating rule is not positively correlated.

5. Discussion

In this work we introduced and examined multiple properties related to the modeling of student knowledge in knowledge structures and knowledge spaces. We began by looking at the implications of the forgetting consistency condition [FC], a condition that was introduced with the goal of allowing the forgetting of items in a knowledge structure to occur in a systematic way. We showed that, when combined with the learning smoothness condition [LS] from Cosyn and Uzun (2009), the resulting knowledge structure must be closed under intersection. Next, we introduced the more general concept of positive knowledge correlation. Under the intuition that knowing more should not make it less likely a student knows a particular item, we derived several implications resulting from this condition. In particular, a knowledge structure that satisfies the conditions of positive knowledge correlation satisfies both learning consistency and forgetting consistency. Furthermore, we showed that such a knowledge structure is also necessarily closed under both union and intersection, a strong and slightly surprising result.

To evaluate the concept of positive knowledge correlation, we described the results of an empirical analysis using data from the ALEKS system. For a few different scenarios we saw evidence supporting the concept, as there were no substantial violations of positive knowledge correlation for the examples we evaluated. Finally, we introduced and discussed the related concept of a positively correlated updating rule. In doing so, we derived results giving sufficient conditions for an updating rule to satisfy this property.

One common theme that emerged from our investigation of these concepts is the property of being closed under intersection. That is, being intersection-closed is a consequence of both the forgetting consistency and positive knowledge correlation properties. Additionally, we showed that, in general, non-ordinal learning spaces—i.e., learning spaces that are not intersection-closed—do not satisfy the conditions to have a positively correlated updating rule. Thus, being intersection-closed is either a consequence of, or closely related to, several of the properties discussed in this manuscript.

Previously, some algorithms for constructing knowledge spaces have attempted to relax the condition of being closed under intersection with the goal of using the extra flexibility to reduce the size of the resulting knowledge space (see, for example, Section 11 in Doignon and Falmagne, 2016). The ultimate motivation is that, all else being equal, running an adaptive assessment is easier on a smaller knowledge space. However, recent work has shown that knowledge spaces can have extremely large numbers of states, and as such reducing the sizes of these spaces by even several orders of magnitude might not have much of a practical effect Matayoshi (2021). Thus, taking all of these results together, it could perhaps be argued that the benefits of modeling student knowledge with intersection-closed knowledge spaces outweigh any of the potential drawbacks.

References

Averell, L. and Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology*, 55:25–35.

Cosyn, E. and Uzun, H. (2009). Note on two sufficient axioms for a well-graded knowledge space. *Journal of Mathematical Psychology*, 53(1):40–42.

Cosyn, E., Uzun, H., Doble, C., and Matayoshi, J. (2021). A practical perspective on knowledge space theory: ALEKS and its data. *Journal of Mathematical Psychology*, 101.

de Chiusole, D., Stefanutti, L., Anselmi, P., and Robusto, E. (2020). Stat-knowlab. assessment and learning of statistics with competence-based knowledge space theory. *International Journal of Artificial Intelligence in Education*, 30(4):668–700.

Doble, C., Matayoshi, J., Cosyn, E., Uzun, H., and Karami, A. (2019). A data-based simulation study of reliability for an adaptive assessment based on knowledge space theory. *International Journal of Artificial Intelligence in Education*, 29:258–282.

Doignon, J.-P. and Falmagne, J. (2016). Knowledge spaces and learning spaces. *arXiv: Combinatorics*, pages 1–51.

Doignon, J.-P. and Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23:175–196.

Doignon, J.-P. and Falmagne, J.-C. (1997). Well-graded families of relations. *Discrete mathematics*, 173:35–44.

Ebbinghaus, H. (1885; translated by Henry A. Ruger and Clara E. Bussenius (1913)). *Memory: A Contribution to Experimental Psychology*. Originally published by Teachers College, Columbia University, New York.

Falmagne, J.-C., Albert, D., Doble, C., Eppstein, D., and Hu, X., editors (2013). *Knowledge Spaces: Applications in Education*. Springer-Verlag, Heidelberg.

Falmagne, J.-C. and Doignon, J.-P. (2011). *Learning Spaces*. Springer-Verlag, Heidelberg.

Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103.

Hockemeyer, C., Held, T., and Albert, D. (1997). RATH-a relational adaptive tutoring hypertext WWW-environment based on knowledge space theory.

Lynch, D. and Howlin, C. P. (2014). Real world usage of an adaptive testing algorithm to uncover latent knowledge. In *Proceedings of the 7th International Conference of Education, Research and Innovation*, pages 504–511.

Matayoshi, J. (2021). Approximately counting and sampling knowledge states. *British Journal of Mathematical and Statistical Psychology*.

Matayoshi, J., Granziol, U., Doble, C., Uzun, H., and Cosyn, E. (2018). Forgetting curves and testing effect in an adaptive learning and assessment system. In *Proceedings of the 11th International Conference on Educational Data Mining*, pages 607–612.

Matayoshi, J., Uzun, H., and Cosyn, E. (2019). Deep (un)learning: Using neural networks to model retention and forgetting in an adaptive learning system. In *Artificial Intelligence in Education-20th International Conference, AIED 2019*, pages 258–269.

Matayoshi, J., Uzun, H., and Cosyn, E. (2020). Studying retrieval practice in an intelligent tutoring system. In *Proceedings of the Seventh ACM Conference on Learning @ Scale*, pages 51–62.

McGraw Hill ALEKS (2021). What is ALEKS? `https://www.aleks.com/about_aleks`.