STUDENTS' USE OF DEFINITIONS OF INDEPENDENCE ACROSS DIFFERENT VISUALIZATIONS

LEON WIEMANN
Paderborn University
leon.wiemann@uni-paderborn.de

KARIN BINDER
Paderborn University
karin.binder@uni-paderborn.de

MICHAEL RÖßNER
Paderborn University
roessner@math.uni-paderborn.de

ABSTRACT

The independence of two events can be defined using either joint probabilities or conditional probabilities. Although mathematically equivalent, these definitions highlight different probabilistic structures that may not be equally transparent and accessible across different visualizations. In this study, 319 secondary school students in Germany solved three probabilistic problems involving two binary events, each presented with a fully completed visualization (2×2 table, double tree, or net diagram) in two different information formats (probabilities or natural frequencies). Students' performance in correctly analyzing the (in)dependence of events across these versions and their use of the definitions of independence were analyzed. The participants showed a particularly low performance in problems presented with probability double trees, even though the relevant probabilities are explicitly displayed. Moreover, 49% of students consistently applied the joint probability definition of independence across all visualizations. These findings suggest that many students rely on stable rather than flexible (or even adaptive) patterns of definition use, highlighting the need to explicitly connect multiple properties of independence in instruction.

Keywords: *Independence; Conditional probabilities; Visualizations; Natural frequencies*

1. INTRODUCTION

Independence is a fundamental idea in probability theory and in the teaching of probability (Batanero et al., 2016; Heitele, 1975). Misunderstandings or false assumptions about the independence or dependence of events can have serious consequences beyond the classroom. For example, in some famous trials, multiple cases of sudden infant death syndrome (SIDS) within the same family were wrongly treated as independent events, leading to underestimating the probability of multiple SIDS cases occurring within the same family. However, the occurrence of one case of SIDS within a family is associated with an increased likelihood of a second case. This false assumption of independence contributed to the wrongful convictions of women (O'Grady, 2022; Schneps & Colmez, 2013). Given the importance of the concept of independence, it is encouraging that independence is included in curricular recommendations (e.g., Australian Curriculum, Assessment and Reporting Authority, 2015; Kultusministerkonferenz, 2012; National Council of Teachers of Mathematics, 2000).

The present article focuses on the independence of two binary events, which can be assessed using two mathematically equivalent definitions. Two events A and B are independent if and only if $P(A) \bullet P(B) = P(A \cap B)$, or equivalently, two events A and B are independent if and only if $P(A) = P(A|B)$, provided $P(B) > 0$ (Díaz et al., 2010). Depending on which of these statements is used as a definition, the other follows as a provable proposition. Although mathematically equivalent, these definitions highlight different probabilistic structures. Prior research suggested that introducing independence via conditional probabilities might be more accessible for learners because it directly expresses the idea

© International Association for Statistical Education (IASE/ISI), 2026

that the occurrence of one event does not change the probability of the other event (Ahlgren & Garfield, 1991; Batanero & Borovcnik, 2016; Díaz et al., 2010; Tarr & Lannin, 2005).

Notably, in Bavaria, Germany, where this study was conducted, all textbooks approved for use in public schools introduce independence via conditional probability before stating the joint probability relation (see Table 2 in Section 2.4). Presenting both statements may be instructionally beneficial because different visualizations (e.g., 2×2 tables and tree diagrams) make different aspects of probabilistic information salient. For instance, the joint probability definition may align more closely with the structure of 2×2 tables, whereas the conditional probability definition can be applied efficiently in visualizations that explicitly display conditional probabilities (Krüger et al., 2015; Wiesner & Binder, 2024). From an instructional perspective, focusing on different properties of independence across visualizations may foster students' flexible and adaptive strategy use (Selter, 2009) in assessing the independence of two events. At the same time, it is possible that once introduced, students predominantly rely on the joint probability relation, which textbooks often present as the rule for determining independence in practice. For example, one textbook in Bavaria refers to "$P(A) \bullet P(B) = P(A \cap B)$" as the "multiplication rule" (Distel, 2023, p. 87).

The present quantitative empirical study investigated how students assess the (in)dependence of two binary events across different visualizations. In particular, the study examined which definition of independence students applied when confronted with different visualizations and whether students showed persistent patterns of strategy use.

2. THEORETICAL BACKGROUND

2.1. DEFINING INDEPENDENCE

In probability theory, the concept of independence was originally introduced as the idea describing the absence of a connection or influence between events (Mošna, 2023). For example, Abraham de Moivre characterized independent events as those for which "the happening of one neither forwards nor obstructs the happening of the other" (de Moivre, 1756, p. 6). This characterization directly connects independence to conditional probability, which expresses how probability judgements are revised when new information becomes available (Borovcnik & Bentz, 1991). Given two events A and B in a sample space Ω, the conditional probability of A given B can be defined as $P(A|B) = P(A \cap B) / P(B)$, assuming the marginal probability fulfills $P(B) > 0$. Accordingly, the events A and B are defined to be independent for $P(B) > 0$ if and only if $P(A) = P(A|B)$, that is, the occurrence of event B does not change the probability of event A (Tarr & Lannin, 2005). In the present paper, this is referred to as the *conditional probability definition of independence*.

Because conditional probability and independence are closely related concepts (Batanero & Borovcnik, 2016), Tarr and Jones (1997) proposed a developmental framework to assess students' thinking about conditional probability and independence. Their framework distinguishes four qualitatively different but continuous levels of thinking. At *level 1 (subjective)*, students' thinking is shaped by subjective beliefs. Numerical information is largely ignored, and consecutive events are typically perceived as being related. Students at *level 2 (transitional)* recognize that probabilities may change under certain conditions. However, probability judgements are frequently based on representativeness rather than systematic comparisons. In contrast, at *level 3 (informal quantitative)*, students attend to changes in the composition of the sample space and can judge whether one event influences another. They start to distinguish independent and dependent events, although this distinction remains imprecise. Finally, at *level 4 (numerical)*, students assign numerical values to the probabilities of events under given circumstances and differentiate between independent and dependent events (Tarr & Jones, 1997; Tarr & Lannin, 2005).

At the same time, two events A and B in a sample space Ω are independent if and only if $P(A) \bullet P(B) = P(A \cap B)$, stating a relation between the marginal probabilities of A and B and their joint probability. In the following, this is referred to as the *joint probability definition of independence*.

Either definition of independence—conditional probability or joint probability—can serve as the definition of (two) independent events, with the other serving as a provable proposition that follows directly from $P(A \cap B) = P(A|B) \bullet P(B) = P(A) \bullet P(B)$ (Mošna, 2023). However, the joint probability definition has three potential advantages compared to the conditional probability definition. First, it can

be generalized easily to multiple events. Second, it avoids the requirement that $P(B)>0$, which may constitute an additional cognitive load for learners (see also Mošna, 2023). Third, it explicitly highlights that independence is a symmetric relation between events. Therefore, it seems reasonable that the joint probability definition is frequently adopted in introductory probability textbooks (i.e., Blitzstein & Hwang, 2019; Olofsson & Andersson, 2012). However, due to the close relationship between independence and conditional probability, as well as the interpretation of independence as invariance of probabilities under new information, previous research in probability education commonly recommended introducing independence via the conditional probability definition $P(A)=P(A|B)$ (Ahlgren & Garfield, 1991; Díaz et al., 2010; Mošna, 2023; Wiesner & Binder, 2024). Accordingly, emphasizing the relation between independence and conditional probability is considered crucial for developing a conceptual understanding of independence (Batanero & Borovcnik, 2016; Tarr & Lannin, 2005).

2.2. PERCEPTIONS OF INDEPENDENCE AND CONDITIONAL PROBABILITY

The framework proposed by Tarr and Jones (1997) provides a lens for understanding the development of students' thinking about independence in relation to conditional probability. Typical patterns of thinking observed at different levels of the framework have been described in the literature as preconceptions (Batanero et al., 2016) or cognitive biases (Nabbout-Cheiban, 2017). In the following, we briefly review selected biases in students' thinking about independence and conditional probability.

Following Truran and Truran (1996), it is useful to distinguish between *trial independence* and *classical independence*. Whereas trial independence refers to the repetition of random experiments under identical conditions, the latter concerns the relationship between events within a given sample space. Referring to trial independence, research in probability education suggested that the assumption of independence is often abandoned by students. Instead, previously observed outcomes or perceived patterns are used for probability judgements (Mošna, 2023). For example, Fischbein and Schnarch (1997) reported that after observing three consecutive heads in a series of coin tosses, 20% of the participating Grade 9 and 10% of the participating Grade 11 students judged the probability of heads on the next toss to be lower than that of tails. This tendency to assign a lower probability to a recently frequent outcome is known as the *negative recency effect* (e.g., Kahneman & Tversky, 1972; Wiemann et al., 2025).

The present paper focuses on classical independence as the relationship between two (binary) events in a sample space Ω (Truran & Truran, 1996). Preconceptions related to classical independence are discussed below. First, the independence of events is often confused with causation (Díaz et al., 2010). Students may believe that independent events lack any causal relation, or conversely, that dependent events necessarily imply causation. Truran and Truran (1996) argued that in the case of independence, the common suggestion "'A has no effect on B' […] is at the heart of [this] confusion" (p. 5). Similarly, Díaz et al. (2010) pointed out that textbook explanations of the joint probability definition $P(A)\bullet P(B)=P(A\cap B)$ often justified the multiplication by stating that "two probabilities may be multiplied together because 'A has no effect on B'" (p. 3). Nasser (2007) asked 48 university students to provide a verbal definition of independence. Although only two students referred to the conditional probability definition and four to the joint probability definition, 18 students (38%) provided responses classified as confusing independence with causality. For example, one student stated that dependence holds if "one of the events is the cause; if it is not, it is independent of the other" (Nasser, 2007, p. 155).

Moreover, independence can be confused with a temporal relation. In a qualitative interview study with undergraduate university students, Plaxco (2011) found that "students think about independence of two events by assuming the occurrence of one event and then considering whether this changes the probability of a second event that occurs chronologically after the first" (p. 282). In fact, independence is a timeless conception; it constitutes a bidirectional relation between events (Biehler & Engel, 2015). However, when introducing a temporal relation, independence may be interpreted as unidirectional. As a consequence, students may believe that an event cannot condition another event that precedes it, known as the *time axis fallacy* (Díaz et al., 2010; Falk, 1979).

Lastly, students often equate independence with mutual exclusivity (e.g., $A\cap B=\varnothing$), arguing that two events are independent if they do not occur together (Andam et al., 2025; Cordani & Wechsler, 2006; D'Amelio, 2009; Díaz et al., 2010). In fact, two mutually exclusive events A and B with positive

probabilities are always dependent because $P(A \cap B) = 0$ (Wiesner & Binder, 2024). Nasser (2007) reported that 14 out of 48 university students (29%) defined the independence of events by stating that they are mutually exclusive. Because "independent" and "mutually exclusive" are often used synonymously in everyday language (Nabbout-Cheiban, 2017; Plaxco, 2011), this confusion may also reflect the influence of everyday language on probabilistic thinking.

Within the developmental framework proposed by Tarr and Jones (1997), it can be argued that students who have reached a numerical level of thinking about independence and conditional probability are less likely to exhibit the cognitive biases described above. Nevertheless, prior research indicates that students' numerical reasoning appears to be sensitive to characteristics of the problem and context. For instance, students may impose an underlying temporal structure on events when no such structure is warranted. In a study with secondary school students, Mošna (2023) found that students' preconceptions of independence became apparent in more challenging problems, in which students relied less on formally acquired knowledge. This finding suggests that students' thinking about independence might be influenced not only by the availability of formal knowledge but also by how this knowledge is activated and applied in specific contexts. In particular, the way probabilistic problems are visualized may therefore play an important role.

2.3. ANALYZING INDEPENDENCE IN DIFFERENT VISUALIZATIONS

The visualization used is an important contextual factor that can impact probabilistic thinking in problems with two binary events (e.g., Arkes & Harkness, 1983; Batanero et al., 1996; Binder et al., 2015; Rößner et al., 2025), in particular because different visualizations highlight different probabilistic relations. Accordingly, the definition students use to assess independence may depend on the probabilistic relations made salient by the visualization. In the following, we focus on three visualizations that display different probabilistic relations: 2×2 tables, double trees, and net diagrams. As illustrated in Table 1 (see also, e.g., Rößner et al., 2025; Stegmüller et al., 2024) and Figure 1, 2×2 tables (which are frequently used in textbooks in Germany) typically display marginal and joint probabilities but no conditional probabilities, whereas double trees typically feature marginal and conditional probabilities but no joint probabilities. Unlike (single) tree diagrams, double trees display all eight conditional probabilities (Binder et al., 2023; Krüger et al., 2015; see Figure 1). For this reason, we consider the double tree as an extension of the conventional tree diagram. Lastly, net diagrams display marginal, joint, and conditional probabilities (Binder et al., 2023).

Table 1. Typically displayed probabilities in different visualizations

	Typically displayed probabilities		
Visualization	Marginal probabilities	Joint probabilities	Conditional probabilities
2×2 table	✓	✓	
Double tree	✓		✓
Net diagram	✓	✓	✓

In addition to probabilities, all three visualizations can present information in the form of so-called "natural frequencies" (e.g., Batanero et al., 1996; Rößner et al., 2025). Natural frequencies express probabilities as pairs of integers within a reference class (Krauss et al., 2020). For example, a 10% probability of having a disease can be expressed as a natural frequency of "10 out of 100" (Gigerenzer & Hoffrage, 1995). Natural frequencies can foster students' probabilistic thinking in problems featuring two binary events (e.g., McDowell & Jacobs, 2017). Figure 1 illustrates the three visualizations in both information formats for an illustrative, probabilistic problem. Here, students are asked to determine whether taking the bus ("Bus") and arriving late for school ("Late") are independent.

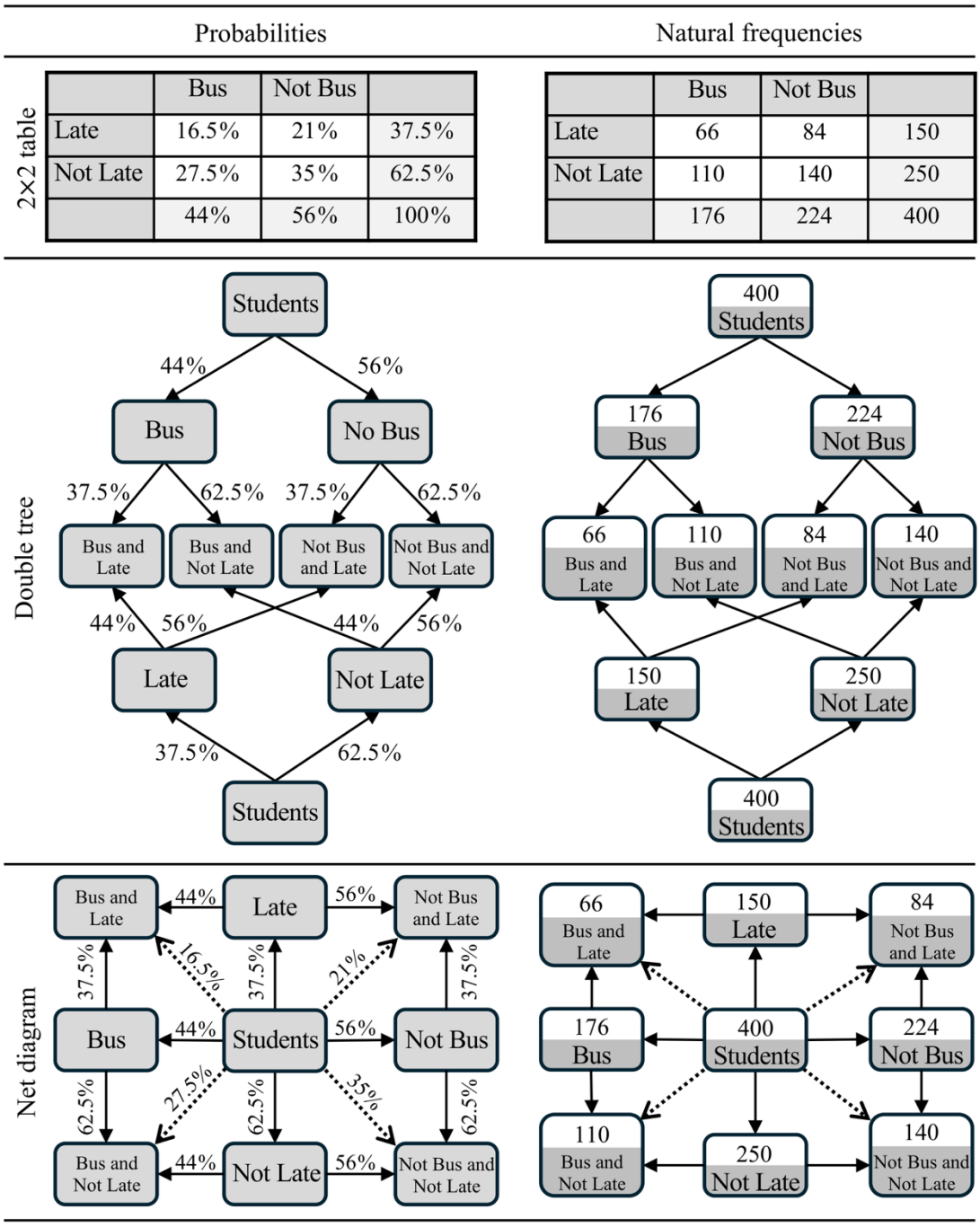


Figure 1. Visualizations in both information formats in an exemplary 'school arrival' context with two binary events for students to determine whether the events are independent

In the following, we describe how independence can be analyzed in each visualization and information format, with a particular focus on how applying the conditional probability and joint probability definition of independence may be facilitated or hindered by the respective visualization and information format.

Analyzing independence in 2×2 tables Prior research has focused on students' assessment of the association of two binary events in 2×2 tables presenting natural frequencies (e.g., Arkes & Harkness, 1983; Batanero et al., 1996, 2015). Figure 2 illustrates in general terms a typical 2×2 table with natural frequencies (see also Batanero et al., 1996).

	A	$\overline{A}$	
B	a	b	a+b
$\overline{B}$	c	d	c+d
	a+c	b+d	a+b+c+d

Figure 2. A 2×2 table presenting natural frequencies in general terms

A 2×2 table with natural frequencies allows several approaches to assess whether the events A and B are independent. For example, independence can be examined using the conditional probability definition by comparing a marginal probability, such as $P(A)=(a+c)/(a+b+c+d)$, with a corresponding conditional probability, such as $P(A|B)=a/(a+b)$. Equivalently, corresponding conditional probabilities can be compared directly, for instance $P(B|A)=a/(a+c)$ and $P(B|\overline{A})=b/(b+d)$ (Batanero et al., 1996; Jenkins & Ward, 1965). Moreover, the joint probability definition of independence can be applied by comparing the product of two marginal probabilities, for example, $P(A)=(a+c)/(a+b+c+d)$ and $P(B)=(a+b)/(a+b+c+d)$, with their joint probability, in this case $P(A\cap B)=a/(a+b+c+d)$.

However, empirical evidence suggests that students' thinking appears to be influenced by selective attention to specific table components. Arkes and Harkness (1983) demonstrated that assessments of association and independence were particularly influenced by the inner cell containing the largest frequency, which in many cases is the cell representing the joint occurrence of both events (i.e., the cell with entry a in Figure 2). Similarly, Batanero et al. (1996) reported that many students relied on incomplete strategies, such as focusing on a single cell or comparing absolute numbers. In their study with secondary school students, 14% of students only used one cell to analyze the association of two events. Moreover, Arkes and Harkness (1983) found that participants changed strategies in response to problem characteristics such as labeling, ordering, and the salience of events.

A 2×2 table in probability format typically displays marginal and joint probabilities, but not conditional probabilities. Therefore, this visualization may directly facilitate the application of the joint probability definition by comparing a product of two marginal probabilities with the probability of the events' joint occurrence, which is displayed in the cell at the intersection of the corresponding row and column (see Figure 3A). Although conditional probabilities are not shown in such tables, they can be calculated by dividing a joint probability by the corresponding marginal probability. This allows the conditional probability definition to be applied by comparing a conditional probability with the already displayed marginal probability of the conditioned event (see Figure 3B). However, students may struggle to derive and calculate conditional probabilities from 2×2 tables presenting probabilities (Binder et al., 2020).

A

	Bus	Not Bus	
Late	16.5%	21%	37.5%
Not Late	27.5%	35%	62.5%
	44%	56%	100%

$0.44 \bullet 0.375 = 0.165$

B

	Bus	Not Bus	
Late	16.5%	21%	37.5%
Not Late	27.5%	35%	62.5%
	44%	56%	100%

$\frac{0.165}{0.375} = 0.44$

Figure 3. Theoretical applications of the joint probability definition on the left (A) and conditional probability definition on the right (B) in 2×2 tables with probabilities displaying two independent binary events

Analyzing independence in double trees The present study focused on double trees rather than (single) tree diagrams. The latter display only one direction of conditioning and therefore do not contain

all marginal and conditional probabilities (Binder et al., 2023; Krüger et al., 2015). Double trees presenting probabilities feature all four marginal and all eight conditional probabilities, but no joint probabilities. Consequently, applying the joint probability definition requires calculating a joint probability, for instance by using $P(A \cap B) = P(A) \bullet P(B|A)$, and comparing this with the product of the corresponding marginal probabilities. In contrast, using the conditional probability definition does not require any calculation because all relevant probabilities are displayed in the visualization. Assessing (in)dependence can then proceed by identifying and comparing appropriate conditional and marginal probabilities. Accordingly, different comparisons are possible. First, a conditional probability, such as $P(A|B)$, can be compared with the marginal (unconditional) probability, $P(A)$. However, Wiesner and Binder (2024) pointed out that these probabilities were displayed on opposite sides of the double tree. Alternatively, two conditional probabilities shown on parallel branches of the tree can be compared, for instance, $P(A|B)$ and $P(A|\bar{B})$. Further, double trees in probability format representing two independent events contain only two distinct pairs of probabilities, as illustrated in Figure 4. This structural property may support thinking about independence. Examining whether students who are not familiar with such a visualization recognize and use this structure to analyze independence may thus provide insights into their understanding of independence beyond procedural familiarity.

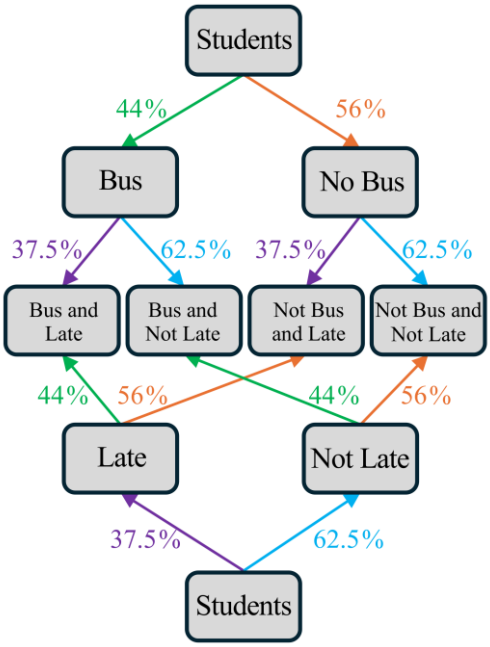


Figure 4. Double tree with probabilities displaying two independent binary events

In double trees with natural frequencies, joint frequencies are explicitly displayed (see Figure 1). This allows marginal, joint, and conditional probabilities to be calculated by dividing two displayed frequencies. In particular, the same applications of the definitions of independence discussed in the context of 2×2 tables with natural frequencies can, in principle, be used because both visualizations present the same frequencies (Krüger et al., 2015). However, double trees differ from 2×2 tables in how these frequencies are structured. In a double tree, joint and marginal frequencies that are required for applying the joint probability definition are not directly connected by branches (e.g., "Students" in Figure 1). In contrast, all relevant frequencies needed to compute marginal and conditional probabilities for applying the conditional probability definition of independence are directly linked through branches.

Analyzing independence in net diagrams Net diagrams presenting probabilities directly display marginal, joint, and conditional probabilities. Therefore, both the joint probability and conditional probability definitions of independence can be used. To apply the joint probability definition, the product of two marginal probabilities can be compared with the respective joint probability found on the branch directly between the nodes displaying the marginal probabilities, as illustrated in Figure 5.

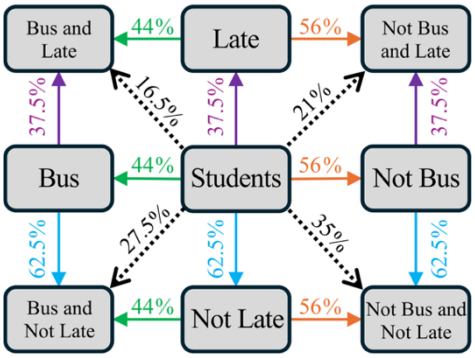


Figure 5. Net diagram with probabilities displaying two independent binary events

As with double trees, applying the conditional probability definition to assess independence does not require calculation because all marginal and conditional probabilities are explicitly displayed. Further, net diagrams with probabilities representing independent events show identical probabilities along parallel branches (see Figure 5). Wiesner and Binder (2024) emphasized that this structure might help students understand that independence links marginal and conditional probabilities. In net diagrams presenting natural frequencies, the same frequency values are displayed in 2×2 tables and double trees. Therefore, the conditional probability and joint probability definition of independence can be applied as previously discussed.

To summarize, analyzing the independence of two events may be affected by the given visualization with respect to the information displayed and the way probabilistic relations are structured. In particular, double trees and net diagrams in probability format are particularly noteworthy because they require no numerical operations to assess independence. At this point, however, it should be noted that independence often serves as a modeling assumption and is rarely observed in empirical data (Borovcnik, 2012; Díaz et al., 2010). By contrast, typical instructional problems present perfect independence. Consequently, assessing (in)dependence typically involves applying formal criteria in idealized situations.

2.4. CURRICULAR BACKGROUND AND INSTRUCTIONAL CONTEXT IN BAVARIA

In Bavaria, a federal state of Germany, where the present study was conducted, conditional probabilities and independence are introduced in Grade 11 (Staatsinstitut für Schulqualität und Bildungsforschung [ISB], 2025). In this state, only three different textbooks are permitted for mathematics at this level (Bayerisches Staatsministerium für Unterricht und Kultus, 2025). Table 2 summarizes how independence is introduced and which visualizations are used in these three textbooks. The textbooks A and B are widely used, whereas textbook C appears to be used less frequently.[1]

According to the curriculum, students are expected to assess the (in)dependence of events using 2×2 tables and tree diagrams. However, the curriculum does not explicitly prescribe a particular definition or property to introduce independence (ISB, 2025). All textbooks include both definitions of independence and the two visualizations prescribed by the curriculum. Further, all textbooks introduce independence via conditional probabilities, which is consistent with recommendations from research in probability education (Díaz et al., 2010; Mošna, 2023). When definitions of independence are used explicitly, the joint probability statement is typically used with 2×2 tables, whereas the conditional probability definition is predominantly applied in tree diagrams. Although 2×2 tables are presented in both information formats, tree diagrams are only rarely presented with natural frequencies.

[1] Although no official statistics on textbook use are available, professional exchanges with teachers from multiple schools suggest that textbooks A and B are widely used in schools.

Table 2. Independence in Grade 11 mathematics textbooks in Bavaria, Germany

Textbook	Introduction of independence	Formal definition(s) of independence	Visualizations used/ Definition(s) used within visualizations
Textbook A (Biburger et al., 2023)	via conditional probabilities (p. 84)	joint probability definition (p. 84)	• 2×2 tables and tree diagrams, both in probabilities and natural frequencies • explicit use of joint probability definition in 2×2 tables with natural frequencies (p. 86) • explicit use of conditional probability definition in tree diagrams with probabilities (p. 86)
Textbook B (Brendel et al., 2024)	via conditional probabilities (p. 90)	conditional probability definition, joint probability definition as implication but in bold print (p. 90)	• 2×2 tables in probabilities and natural frequencies, tree diagrams in probabilities • explicit use of joint probability definition in 2×2 tables with probabilities (p. 90) • explicit use of conditional probability definition in tree diagrams with probabilities (p. 91)
Textbook C (Distel, 2023)	via conditional probabilities (p. 87)	conditional probability definition (p. 87), joint probability definition as "multi-plication rule" (p. 87)	• 2×2 tables in probabilities and natural frequencies, tree diagrams in probabilities • explicit use of both definitions in 2×2 tables (p. 88)

2.5. FLEXIBILITY, ADAPTIVITY, AND THE "EINSTELLUNG EFFECT"

Empirical research on students' thinking about independence in 2×2 tables reveals that students use multiple strategies to consider independence, applying multiple properties and definitions of independent events using the same visualization (Arkes & Harkness, 1983; Batanero et al., 1996). Importantly, these strategies are not fixed but are influenced by problem characteristics. This variability can be examined in terms of *strategy flexibility* and *strategy adaptivity*. Strategy flexibility refers to the ability to access and switch between different strategies within or across problems, whereas strategy adaptivity denotes the ability to select and apply the *most appropriate* or *efficient* strategy for a given problem (Selter, 2009). Further, Star and Rittle-Johnson (2008) distinguished between knowledge of different strategies and their relative efficiency, as well as the observed use of these strategies.

Because the main interest of the present study was how students analyzed independence in response to different visualizations, we focused primarily on strategy flexibility. With respect to strategy flexibility, we examined the extent to which students used different definitions of independence across visualizations. In contrast, assessing strategy adaptivity requires a normative judgement about which strategies are particularly appropriate or efficient for a given problem. Although different definitions of independence may be similarly appropriate in most cases, double trees and net diagrams in probability format constitute an exception. In these, applying the conditional probability definition is particularly efficient because the relevant probabilities can be compared directly without additional calculation. However, because strategy adaptivity can only be assessed for these two visualizations, the primary focus in the following was on strategy flexibility.

Both strategy flexibility and strategy adaptivity are considered valuable learning outcomes (Verschaffel, 2024), and some authors have argued that they are associated with higher performance (e.g., Star et al., 2022). However, empirical evidence suggests that learners often develop preferred strategies for particular types of problems based on their knowledge rather than selecting among available strategies each time they encounter a problem (Threlfall, 2009; Verschaffel, 2024). This may result in strategy fixation rather than flexible and adaptive strategy use. One explanation for the persistence of previously successful strategies is the so-called *Einstellung effect* or *mental set effect* (Luchins, 1942). When solving a problem, people tend to apply a previously learned strategy rather than consider potentially simpler alternatives. Such a mental set can arise through repeated practice and enable quick solutions for problems with similar structures (Schultz & Searleman, 2002). At the same

time, it can restrict adaptivity and reduce sensitivity to more efficient strategies. The Einstellung effect was frequently documented in students' assessment of probabilistic problems. For example, Weber et al. (2018) reported that students often transformed natural frequencies into probabilities before performing calculations. By doing so, learners reverted to a familiar probability-based solution strategy. Similarly, research on probabilistic preconceptions indicated that certain biases could be robust or even increase with levels of education (e.g., Morsanyi et al., 2009), suggesting that instruction may reinforce stable strategy use.

3. RESEARCH QUESTIONS AND HYPOTHESES

In this study, we investigated students' performance in analyzing the (in)dependence of two binary events across three different visualizations (2×2 tables, double trees, and net diagrams) and two different information formats (probabilities and natural frequencies). Additionally, we analyzed how the visualization and information format influenced the choice of definition applied and to what extent participants used definitions of independence flexibly and adaptively across visualizations. Table 3 summarizes the research questions and hypotheses.

Table 3. Research questions and hypotheses

Research question	**Hypotheses**
(RQ1) How do different visualizations and information formats affect students' performance in analyzing the (in)dependence of two binary events (when controlling for students' mathematical achievement)?	(H1a) Students' performance in analyzing the (in)dependence of two events is better in 2×2 tables and double trees than in net diagrams.
	(H1b) In double trees and net diagrams, students' performance in analyzing the (in)dependence of two events is better in probability format rather than in natural frequency format.
(RQ2) Which definition of independence (via joint probability or conditional probability) is applied across different visualizations and information formats? How does the chosen definition relate to students' performance?	(H2a) In 2×2 tables with probabilities and natural frequencies, most students apply the joint probability definition of independence.
	(H2b) In double trees and net diagrams in probability format, most students apply the conditional probability definition of independence (i.e., adaptive use of the most appropriate definition for these visualizations).
	(H2c) In double trees and net diagrams in probability format, students applying the conditional probability definition perform better in analyzing the (in)dependence of events than students applying the joint probability definition.
(RQ3) How flexibly (and adaptively) do students use the definitions of independence when analyzing the (in)dependence of events across different visualizations? How does students' performance vary across different categories of definition use?	(H3a) Most students use the definitions of independence flexibly across visualizations.
	(H3b) Students who use the definitions of independence flexibly across visualizations perform better in analyzing the (in)dependence of events than students who consistently use the same definition of independence across visualizations.

4. METHOD

4.1. PARTICIPANTS

The study was conducted in July 2024 and included $N=319$ school students from 20 classes in 13 secondary schools ("Gymnasium") in Bavaria. Because conditional probabilities and independence of events are introduced in Grade 11, we focused on students at the end of this level. All corresponding teachers reported that independence had already been taught in mathematics prior to the study in accordance with the curriculum and that some time had passed since then. All classes used a textbook

authorized for use in public schools (see section 2.4). Consequently, students were familiar with both statements of independence and with 2×2 tables and tree diagrams.

Of the 319 participants, 183 students identified as male, 132 as female, and 2 as diverse. Two students did not report a gender. Students' ages ranged from 14 to 20 years, with a mean of $M = 16.96$ years ($SD = 0.68$). All students participated voluntarily and gave their written informed consent to participate. Anonymity of responses was ensured because no names or other directly identifying information were collected.

4.2. DESIGN OF THE STUDY

In this study, students had to assess the (in)dependence of two binary events. Each student completed three problems so that each investigated visualization was presented exactly once. The problems were introduced with a short description, and probabilistic information was provided exclusively through a fully completed visualization. Students were asked to decide whether the two displayed events were independent or dependent and to justify their answer with a calculation.

The three problems differed systematically with respect to visualization, information format, event type, and context, as illustrated in Table 4. The visualization was varied within participants. Each student received one problem featuring a 2×2 table, one featuring a double tree, and one featuring a net diagram. Thus, the design comprised three probabilistic problems per participant. Information was presented as probabilities or natural frequencies, with two problems using one information format and the remaining problem the other. Likewise, two problems involved independent events, and one problem involved dependent events, or vice versa. In line with classroom practice, the problems were constructed such that the events were either perfectly independent (i.e., they fulfilled the definitions of independence) or explicitly dependent.[2] In the latter case, problems were designed to yield, for instance, chi-squared tests of independence with $p < 10^{-28}$. Each problem was presented in a different context ("school arrival," "treatment study," or "quality control"). Visualization and information format were the factors of primary interest. However, problem variations in event type and context were varied to reflect the range of probabilistic problems typically encountered in instruction.

The order of problems was systematically varied, resulting in 144 unique testlets that were randomly assigned to participants. Figure 1 illustrates the visualizations for independent events in the "School arrival" context in both information formats. The corresponding visualizations for the other event types and contexts are provided in the Appendix (see Figure A1–Figure A5). The test was administered as a paper-and-pencil test during a regularly scheduled mathematics lesson, with a maximum completion time of 30 minutes. The use of a calculator was permitted. Students were also asked to report their most recent mathematics grade, which in Germany ranges from 1 (best grade) to 6 (worst grade). The reported mean grade was $M = 2.69$ ($SD = 1.09$). The grade was used as a proxy for students' prior mathematics achievement and was included as a covariate in the statistical analyses to control for differences in mathematics achievement between students.

[2] This approach may be considered problematic because it blurs the distinction between model and reality: Although independence serves as an assumption, empirical data rarely satisfy these conditions perfectly (Díaz et al., 2010).

Table 4. Problem construction and formulation

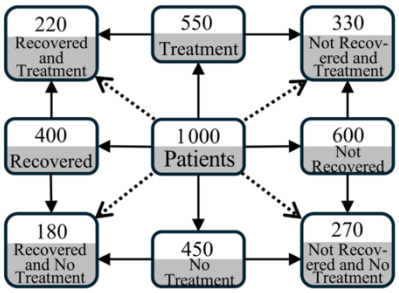
Context	School arrival	Treatment study	Quality control
Problem description	A mathematics teacher is frustrated that the students frequently arrive late to class, especially due to the common excuse "the bus was delayed." For this reason, he wants to determine whether arriving late for school depends on taking the bus. The following information is available and can be taken from the visualization.	In a clinical study, the effect of a new treatment is being investigated. Therefore, it must be evaluated whether patients' recovery depends on whether they received the new treatment or a placebo. The following information is available and can be taken from the visualization.	At a farm's packaging facility, eggs are packaged every day. The managing director is interested in whether the sorting machine is working efficiently. That is, whether rejecting an egg depends on whether it is damaged. The following information is available and can be taken from the visualization.
Visualization	2×2 table, double tree, or net diagram (see Figure 1)		
Information format	Probabilities or natural frequencies within the visualization (see Figure 1)		
Event type	Independent events or dependent events		
Instruction	Indicate whether the events "Bus" and "Late" are independent or dependent. Please show your complete calculation.	Indicate whether the events "Recovered" and "Treatment" are independent or dependent. Please show your complete calculation.	Indicate whether the events "Rejected" and "Damaged" are independent or dependent. Please show your complete calculation.

4.3. CODING

Each response was coded as *correct* (= 1) if the student correctly identified the events as independent or dependent, and *incorrect* (= 0) otherwise. Responses coded as correct included responses in which students explicitly stated the correct classification as "independent" or "dependent," as well as responses in which the correct classification was demonstrated by a calculation, even if no such term was used. In addition, responses were coded as correct when students stated their answer by circling or underlining the corresponding term in the problem instruction, as displayed in Figure 6A.

Moreover, each response was classified according to its use of the definitions of independence (*joint probability definition of independence* and *conditional probability definition of independence*, see section 2.1). A response was coded as *using the joint probability definition* (= 1) if it expressed a relation between the marginal probabilities of both events and their joint probability, either in probability notation (i.e., "$P(A)$," "$P(B)$," and "$P(A \cap B)$"), or by referring to the corresponding quantities from the problem. The stated relation did not have to be mathematically correct for the problem at hand. For example, if a student tried to use the joint probability definition but applied the definition incorrectly (i.e., addition instead of multiplication, "$P(A)+P(B)=P(A \cap B)$"), the answer was still coded as using the joint probability definition. Accordingly, the response shown in Figure 6A was coded as incorrect, but as using the joint probability definition. Similarly, a response was classified as *using the conditional probability definition* (= 1) if it expressed a relation between a marginal and a conditional probability, or between two conditional probabilities, either in formal probability notation or by referring to the numerical values provided. For each problem, we recorded whether the students used exactly one definition, both definitions, or neither definition. For instance, the response shown in Figure 6B was coded as correct using neither definition of independence.

A In a clinical study, the effect of a new treatment is being investigated. Therefore, it must be evaluated if patients' recovery depends on whether they received the new treatment or a placebo.
The following information is available and can be taken from the visualization.



Indicate whether the events "Recovered" and "Treatment" are independent or dependent.
Please show your complete calculation.

Space for notes:

~~600~~ $P(A) \cdot P(B) \overset{\text{in}}{=} P(A \cap B)$
$550 - 600 \overset{\text{in}}{=} 330$
$50 \neq 330$

B At a farm's packaging facility, eggs are packaged every day. The managing director is interested in whether the sorting machine is working efficiently. That is, whether rejecting an egg depends on whether it is damaged.
The following information is available and can be taken from the visualization.

	Rejected	Not Rejected	
Damaged	700	300	1000
Undamaged	1800	7200	9000
	2500	7500	10000

Indicate whether the events "Rejected" and "Damaged" are independent or dependent.
Please show your complete calculation.

Space for notes:

Dependent

Figure 6. Examples of students' written responses to two probabilistic problems, translated into English and recreated to conceal the original handwriting

Responses from ten randomly selected participants were independently coded by two raters and full agreement was achieved. Because the coding categories were unambiguous, the remaining responses were coded by a single researcher using a coding manual. Identifying the independence definitions used was straightforward; students either clearly applied a formula (see Figure 6A) or merely stated whether the events were (in)dependent without providing a justification or calculation (see Figure 6B).

4.4. STATISTICAL ANALYSIS

For (RQ1), students' performance in analyzing the (in)dependence of events was first examined descriptively, separately by visualization and information format. To examine whether observed differences in performance were systematically associated with visualization and information format while taking problem characteristics and student characteristics into account, an inferential statistical model was fitted. Because the outcome variable was binary and each student responded to multiple problems, a generalized linear mixed model (GLMM) with a binomial error distribution and a logit link function was used. Such a mixed-effects approach was chosen because each student responded to multiple problems, so responses within students were not independent (Bolker, 2015; Hilbert et al., 2019). The model was estimated in R (version 2025.09.0+387) using the package "glmmTMB." In the following, p_{ij} denotes the probability that student i correctly analyzed the events in problem j. The logit of this probability, defined as $\log(p_{ij}/(1-p_{ij}))$, was modeled as

$$(1)\quad \log(p_{ij}/(1-p_{ij})) = \beta_0 + \beta_1 \bullet \text{Double_tree}_{ij} + \beta_2 \bullet \text{Net_diagram}_{ij} + \beta_3 \bullet \text{Natural_frequencies}_{ij} + \beta_4 \bullet \text{Independent_events}_{ij} + \beta_5 \bullet \text{Treatment_study}_{ij} + \beta_6 \bullet \text{Quality_control}_{ij} + \beta_7 \bullet \text{Grade}_i + \beta_8 \bullet (\text{Double_tree}_{ij} \times \text{Natural_frequencies}_{ij}) + \beta_9 \bullet (\text{Net_diagram}_{ij} \times \text{Natural_frequencies}_{ij}) + u_i.$$

Double_tree_{ij}, Net_diagram_{ij}, and $\text{Natural_frequencies}_{ij}$ were dummy coded, with 2×2 tables and probability format serving as reference categories, respectively. Event type and context were included as additional problem-level factors. Here, $\text{Independent_events}_{ij}$ was dummy coded, with dependent events as the reference category. The context was dummy coded with "School arrival" as the reference category, resulting in the two indicator variables for the contexts "$\text{Treatment_study}_{ij}$" and "$\text{Quality_control}_{ij}$." Further, Grade_i denotes students' standardized mathematics grade, included as a proxy for students' prior mathematical achievement (e.g., Kuncel et al., 2004). The interactions between visualization and information format were included with respect to hypothesis (H1b). Finally, u_i represents a student-specific random intercept that accounts for the repeated observations per student.

The GLMM estimates the regression coefficients β_1 to β_9. To ease interpretation, regression coefficients were transformed into odds ratios (*OR*), calculated as $OR = e^{\beta}$. Odds ratios express the factor by which the odds of a correct response change for a one-unit increase in a continuous predictor or relative to the reference category for categorical predictors (e.g., Breslow & Clayton, 1993; Twisk, 2019). For instance, an *OR* of 2.00 for double trees compared to 2×2 tables would indicate twice the odds of a correct response for double trees compared to 2×2 tables, holding the other predictors in the GLMM model constant. The 95% confidence interval (CI) and the *p*-value are reported for each *OR* as well. Moreover, the variance explained by the model was evaluated using marginal and conditional R^2 (Nakagawa & Schielzeth, 2013). The marginal R^2 reflects the variance explained by fixed factors only, whereas the conditional R^2 reflects the variance explained by both fixed and random factors.

Based on the fitted GLMM, estimated marginal probabilities of a correct response were obtained using the "emmeans" package in R. These model-based probabilities state students' estimated performance for each visualization and information format, while accounting for event type, context, and students' standardized mathematics grade as included in the fitted model. To test hypotheses (H1a) and (H1b), contrasts were calculated based on these estimated marginal probabilities. In particular, for hypothesis (H1a), contrasts compared 2×2 tables with net diagrams and double trees with net diagrams, while accounting for information format, event type, context, and the standardized mathematics grade. For (H1b), contrasts compared probabilities with natural frequencies within double trees and net diagrams. Contrasts were expressed as differences in predicted probabilities (Δ). To ease interpretation, Δ values are reported as percentage points, together with their 95% confidence intervals and *p*-values.

For research question (RQ2) and hypotheses (H2a) and (H2b), students' use of the definitions of independence was examined descriptively. Frequencies and percentages of definition use were examined separately by visualization and information format for using the joint probability definition, conditional probability definition, both definitions, and no definition of independence.

Hypothesis (H2c) predicts that under adaptive definition use (i.e., for both double trees and net diagrams in probability format), the application of the conditional probability definition is associated with a higher performance rate than the application of the joint probability definition of independence. To test this hypothesis, a GLMM analogous to that used in (RQ1) was fitted to the subset of problems featuring either a double tree or a net diagram with probabilities in which *exactly* one of the two definitions was applied, that is, either the joint probability *or* the conditional probability definition. The probability p_{ij} that student i correctly analyzed the events in problem j as (in)dependent was now modeled as

$$(2)\quad \log(p_{ij}/(1-p_{ij})) = \beta_0 + \beta_1 \bullet \text{Net_diagram}_{ij} + \beta_2 \bullet \text{Conditional_probability_definition}_{ij} + \beta_3 \bullet \text{Independent_events}_{ij} + \beta_4 \bullet \text{Treatment_study}_{ij} + \beta_5 \bullet \text{Quality_control}_{ij} + \beta_6 \bullet \text{Grade}_i + \beta_7 \bullet (\text{Net_diagram}_{ij} \times \text{Conditional_probability_definition}_{ij}) + u_i.$$

Here, Net_diagram_{ij} and $\text{Conditional_probability_definition}_{ij}$ were dummy coded, with double trees and the application of the joint probability definition serving as reference categories, respectively. All other variables were defined as in the model for (RQ1). With respect to (H2c), contrasts compared the

application of the conditional probability with the application of the joint probability definition within the subset of responses described above. The contrasts were again expressed as differences in predicted probabilities (Δ) and are reported in percentage points, together with their 95% confidence intervals and *p*-values.

The third research question (RQ3) examines students' flexibility (and adaptivity) in using the definitions of independence across visualizations. Students were categorized according to their definition use across the three problems into consistent use of the joint probability definition, the conditional probability definition, or both definitions of independence, as well as varying definition use across problems and incomplete or absent definition use. Regarding hypotheses (H3a) and (H3b), these categories were analyzed descriptively with respect to their frequencies and students' mean number of problems in which events were correctly assessed as (in)dependent. To gain insight into students' flexibility in definition use, the specific combinations of applied definitions across visualizations and information formats were additionally examined descriptively for those students who varied their definition use across problems.

5. RESULTS

5.1. (RQ1): STUDENTS' PERFORMANCE IN ANALYZING (IN)DEPENDENCE

Figure 7 shows students' performance in correctly analyzing the (in)dependence of two binary events, separately for each visualization and information format. In probability format, participants performed best with net diagrams (65%). Their performance was similarly high for 2×2 tables (62%) and lowest for double trees (42%). In natural frequencies, performance showed little variation across visualizations, ranging from 51% to 60%. When aggregating across information formats, students' performance was best for the 2×2 table with 61% correct answers (195 out of 319), followed by the net diagram (185 out of 319; 58%), and the double tree (159 out of 319; 50%).

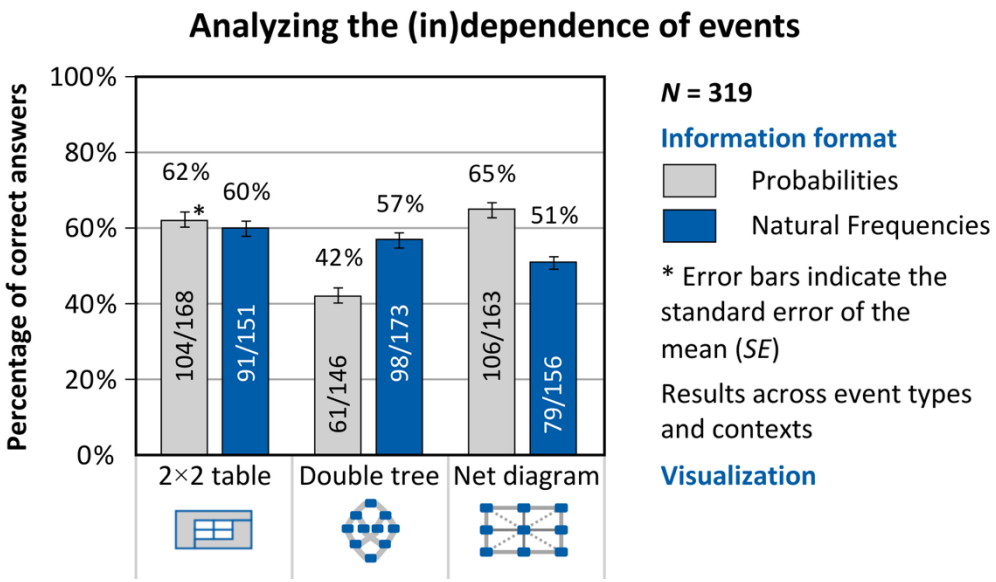


Figure 7. Students' performance in analyzing the (in)dependence of events, separated for visualizations and information formats

The results of the GLMM for (RQ1) (see (1) in section 4.4) are presented in Table 5. For problems presented in probability format (reference category), the odds of correctly analyzing the (in)dependence of events were substantially lower for double trees than for 2×2 tables ($OR = 0.25$, 95% CI [0.13, 0.50], $p < 0.001$). For problems presented in probability format, the *OR* of 0.25 indicates 75% lower odds of a correct response for double trees compared to 2×2 tables. In contrast, the estimated odds were similar for net diagrams compared to 2×2 tables ($OR = 1.08$, 95% CI [0.55, 2.11], $p = 0.83$).

Within 2×2 tables, the odds of a correct response were also similar between both information formats ($OR = 0.86$, 95% CI [0.43, 1.72], $p = 0.66$). However, the interaction between double trees and natural frequencies ($OR = 3.19$, 95% CI [1.14, 8.89], $p = 0.03$) indicates that the disadvantage in analyzing the (in)dependence of events associated with double trees was less pronounced when information was presented as natural frequencies rather than probabilities. The corresponding interaction for net diagrams and natural frequencies was estimated less precisely, reflected in a wider confidence interval ($OR = 0.62$, 95% CI [0.22, 1.73], $p = 0.36$).

Table 5. GLMM results for students' performance in correctly analyzing the (in)dependence of events

Predictor	*OR*	95% CI	*p*
Intercept	2.96	[1.60, 5.49]	< 0.001
Visualization: Double tree (vs. 2×2 table)	0.25	[0.13, 0.50]	< 0.001
Visualization: Net diagram (vs. 2×2 table)	1.08	[0.55, 2.11]	0.83
Information format: Natural frequencies (vs. probabilities)	0.86	[0.43, 1.72]	0.66
Event type: Independent events (vs. dependent events)	0.54	[0.37, 0.78]	0.001
Context: "Treatment study" (vs. "School arrival")	1.37	[0.89, 2.13]	0.16
Context: "Quality control" (vs. "School arrival")	1.27	[0.82, 1.96]	0.28
Mathematics grade (standardized)	0.39	[0.28, 0.54]	< 0.001
Interaction: Double tree × Natural frequencies	3.19	[1.14, 8.89]	0.03
Interaction: Net diagram × Natural frequencies	0.62	[0.22, 1.73]	0.36

Note. Odds ratios (*OR*), calculated as $OR = e^{\beta}$ from the regression coefficient β, express the factor by which the odds of a correct response change for a one-unit increase in a continuous predictor or relative to the reference category for categorical predictors.

As an exploratory finding, independent events were associated with lower odds of a correct response than dependent events ($OR = 0.54$, 95% CI [0.37, 0.78], $p = 0.001$). Further, higher mathematics achievement was associated with higher performance. Because higher mathematics grades represent lower achievement in the German grading system, an increase of one standard deviation in mathematics grade was associated with 61% lower odds of correctly analyzing the (in)dependence of events ($OR = 0.39$, 95% CI [0.28, 0.54], $p < 0.001$). The model explains 15% of the variance in performance through fixed effects and 61% through both fixed and random effects ($R^2_m = 0.15$, $R^2_c = 0.61$). Thus, considerably more variation in performance was captured when between-student differences were considered as well.

Based on the GLMM, estimated marginal probabilities for correctly analyzing (in)dependent events were computed (see section 4.4). These estimated marginal probabilities were used to test contrasts related to hypotheses (H1a) and (H1b), as displayed in Table 6. With respect to hypothesis (H1a), the model estimated a difference in the probability of correctly analyzing the (in)dependence between 2×2 tables and net diagrams of only 3.62 percentage points (95% CI [-5.72, 12.94], $p = 0.45$). In addition, given the *p*-value of 0.45, observing a difference at least as extreme as the one obtained would not be unusual if there were no true difference in the estimated probabilities. Further, the estimated probability for a correct response was 15.34 percentage points lower for double trees than for net diagrams (95% CI [-25.71, -5.05], $p = 0.01$), contrary to the prediction in (H1a).

Table 6. Contrasts for students' performance in analyzing the (in)dependence of events based on estimated marginal probabilities from GLMM (1)

Contrast	Δ	95% CI	*p*
(H1a): 2×2 table vs. net diagram	3.62	[-5.72, 12.94]	0.45
(H1a): Double tree vs. net diagram	-15.34	[-25.71, -5.05]	0.01
(H1b): Double tree: Probabilities vs. natural frequencies	-24.53	[-40.92, -8.14]	0.03
(H1b): Net diagram: Probabilities vs. natural frequencies	14.05	[-1.24, 29.33]	0.07

Note. The second category serves as the reference category. Δ represents the model-estimated difference in predicted probabilities between the two conditions.

For hypothesis (H1b), the estimated probability for correctly analyzing the (in)dependence of events in double trees was 24.53 percentage points lower when probabilities rather than natural frequencies were presented (95% CI [-40.92, -8.14], $p=0.03$). This difference was in the opposite direction to that predicted by (H1b). For net diagrams, students' performance was estimated to be 14.05 percentage points higher in probabilities than in natural frequencies (95% CI [-1.24, 29.33], $p=0.07$). Although the 95% confidence interval is relatively wide, which reflects some uncertainty, under the assumption of no true difference, observing such a difference in estimated probabilities—or a large one—appears rather unusual.

5.2. (RQ2): STUDENTS' USE OF DEFINITIONS OF INDEPENDENCE

Figure 8 presents the percentage and absolute number of problems in which students applied the joint probability definition, conditional probability definition, both definitions, or no definition of independence, separated by visualization and information format. Across conditions, the relative use of definitions varied only moderately. The joint probability definition of independence was used most frequently (53% to 62%), whereas only a minority of students applied the conditional probability definition (6% to 13%). Only 2% to 5% of students applied both definitions, whereas 25% to 36% used no definition at all. Accordingly, hypothesis (H2a), predicting predominant use of the joint probability definition in 2×2 tables, is supported, whereas hypothesis (H2b), stating that the conditional probability definition is primarily used in double trees and net diagrams with probability format (in the sense of an adaptive use of the most appropriate definition in these visualizations), is not supported.

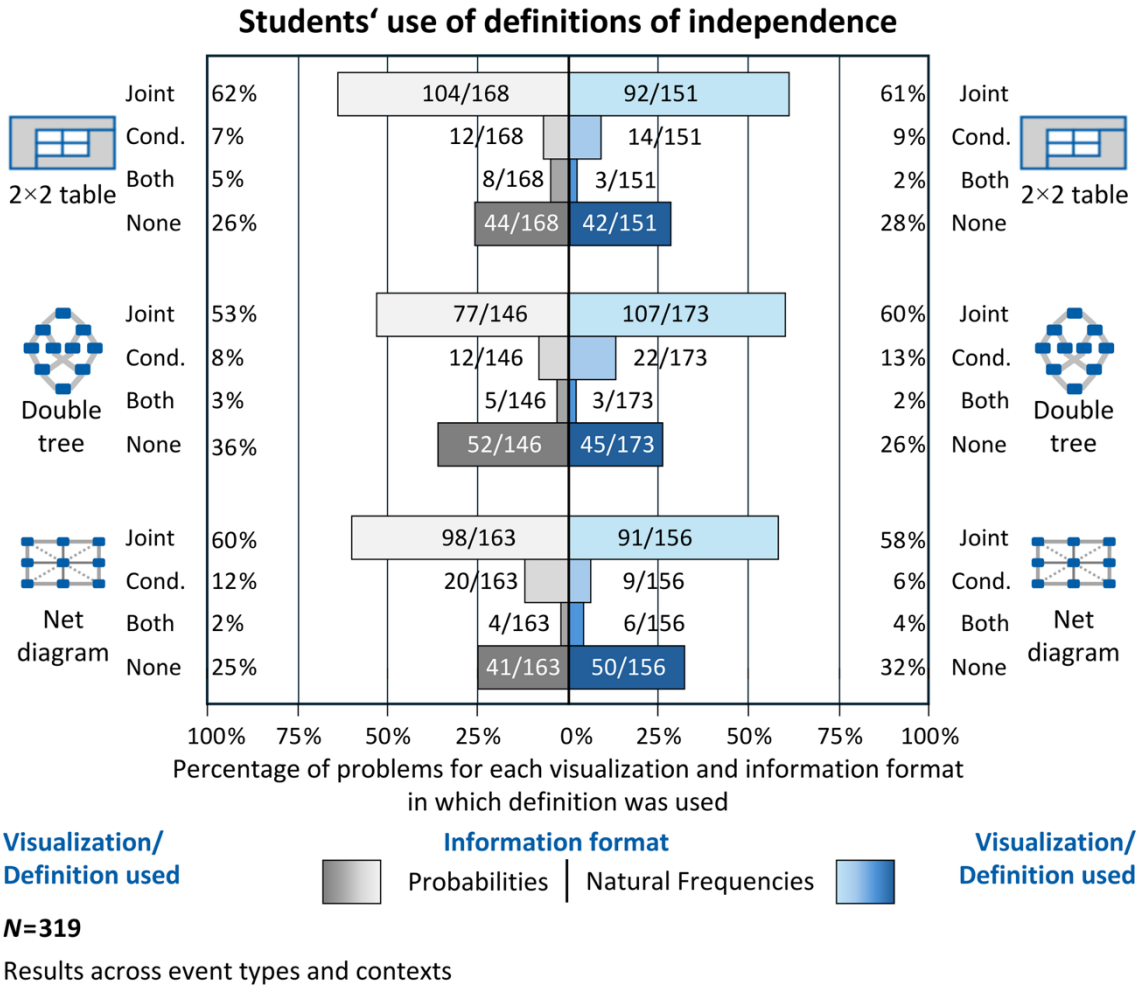


Figure 8. Students' use of definitions of independence, separated by visualizations and information formats

Table 7 presents students' performance rates for correctly analyzing the (in)dependence of events for each visualization and information format, separated by the definition of independence used. The highest overall performance was observed for the conditional probability definition in a natural frequency net diagram ($M = 0.89$); however, only nine students did so. Participants applying the joint probability definition performed best in 2×2 tables ($M \geq 0.80$), whereas students using the conditional probability definition performed best in net diagrams ($M \geq 0.80$). In both cases, the lowest performance was observed in probability double trees ($M = 0.57$ and $M = 0.58$, respectively). With one exception, students who used both definitions achieved intermediate accuracy (M ranging from 0.33 to 0.67 across visualizations and information formats), whereas students using no definition of independence showed the lowest performance, with performance rates ranging from $M = 0.14$ to $M = 0.32$.

Table 7. Mean (M) and standard error (SE) of students' performance in analyzing the (in)dependence of events, separated by definition(s) applied, visualizations, and information formats

		Definition of independence used							
		Joint probability definition		Conditional probability definition		Both definitions		No definition	
Visualization	Information format	*M* (*SE*)	*n*	*M* (*SE*)	*n*	*M* (*SE*)	*n*	*M* (*SE*)	*n*
2×2 table	Probabilities	0.82 (0.03)	104	0.58 (0.15)	12	0.63 (0.18)	8	0.16 (0.06)	44
	Natural frequencies	0.80 (0.04)	92	0.79 (0.11)	14	0.00 (0.00)	3	0.14 (0.05)	42
Double tree	Probabilities	0.57 (0.06)	77	0.58 (0.15)	12	0.40 (0.24)	5	0.15 (0.05)	52
	Natural frequencies	0.68 (0.05)	103	0.68 (0.10)	22	0.33 (0.33)	3	0.27 (0.07)	45
Net diagram	Probabilities	0.77 (0.04)	98	0.80 (0.09)	20	0.50 (0.29)	4	0.32 (0.07)	41
	Natural frequencies	0.66 (0.05)	91	0.89 (0.11)	9	0.67 (0.21)	6	0.14 (0.05)	50

To test hypothesis (H2c), a GLMM was fitted (see (2) in section 4.4) to examine whether applying *exactly* one of the two definitions of independence was associated with differences in performance in double trees and net diagrams in probability format. Based on the estimated marginal probabilities, contrasts for definition use were calculated (see section 4.4). As presented in Table 8, the estimated probability of correctly analyzing the (in)dependence of events was 6.66 percentage points lower when the conditional probability definition rather than the joint probability definition was applied in double trees (95% CI [-41.61, 28.30], $p = 0.71$). For net diagrams, the difference in estimated probabilities was close to zero ($\Delta = 1.03$, 95% CI [-20.01, 22.15], $p = 0.92$). Given the 95% confidence interval, the direction and size of the difference were associated with uncertainty. Overall, these findings provide little evidence that performance differed according to the definition applied. Thus, the results were not consistent with the prediction of hypothesis (H2c).

Table 8. Contrasts for students' performance in analyzing the (in)dependence of events by independence definition applied, based on estimated marginal probabilities from GLMM (2)

Contrast	Δ	95% CI	*p*
(H2c): Double tree in probability format: Conditional probability vs. joint probability definition	-6.66	[-41.61, 28.30]	0.71
(H2c): Net diagram in probability format: Conditional probability vs. joint probability definition	1.03	[-20.01, 22.15]	0.92

Note. The second category serves as the reference category. Δ represents the model-estimated difference in predicted probabilities between the two conditions.

5.3. (RQ3): STUDENTS' USE OF DEFINITIONS ACROSS VISUALIZATIONS

In this study, each student solved three problems, each presented with a different visualization. To examine students' strategy flexibility across visualizations, their definition use in the three problems was categorized as presented in Table 9. Nearly half of the students (49%) consistently applied the joint probability definition across all visualizations, and 5% consistently used the conditional probability definition. In total, 179 out of 319 participants (54%) used the same definition of independence in all visualizations, contradicting hypothesis (H3a). Six students (2%) always applied both definitions of independence. In contrast, 17 students (5%) applied at least one definition in every visualization and used each definition at least once, though not always simultaneously, indicating flexibility in their use of definitions across visualizations. Lastly, 38% of students did not explicitly state any definition in at least one visualization.

Table 9. Categories of students' use of definitions of independence across visualizations, numbers of students within each category, and mean (M) and standard error (SE) of correctly solved problems

	Number of students within the category (N=319)		M and SE of correctly solved problems (out of three)	
Category of students' use of definition(s)	n	%	M	SE
Use of joint probability definition in all visualizations	157	49	2.27	0.07
Use of conditional probability definition in all visualizations	15	5	2.20	0.28
Use of both definitions in all visualizations	6	2	1.17	0.31
Use of at least one definition in every visualization, but not always the same definition and not always both definitions	17	5	2.53	0.21
Use of a definition in at least one, but not all visualizations	59	18	1.05	0.12
No explicit use of definitions in any visualization	65	20	0.59	0.10

Performance also differed descriptively across these categories of definition use, as illustrated in Table 9. Participants who did not explicitly state any definition achieved the lowest mean performance with M=0.59 correct responses out of three problems. Students who consistently used one definition of independence performed considerably better (M=2.27 for using the joint probability definition and M=2.20 for using the conditional probability definition). However, the highest mean performance was observed among the 17 students who varied their use of definitions across visualizations (M=2.53), which is descriptively consistent with (H3b).

Lastly, Table 10 provides a more detailed account of definition use by the 17 students who varied their use across visualizations, that is, students who used at least one definition in every visualization but did not consistently use the same definition or both definitions (see Table 9). The table presents the applied definitions across visualizations and information formats. Because each student received each visualization only in one information format, empty cells indicate that the respective combination of visualization and information format was not presented. Shaded cells indicate correctly solved problems.

Eight of the 17 students (47%; students with numbers 1–8 in Table 10) used the joint probability definition in both 2×2 tables and net diagrams, while using the conditional probability definition in double trees. Seven out of these eight students received the net diagram in probability format. Because this representation also allows conditional probabilities to be compared directly, the continued reliance on the joint probability definition in these cases may indicate limited adaptivity. In contrast, three of the 17 students (18%; students with numbers 9–11 in Table 10) used the joint probability definition in 2×2 tables, and the conditional probability definition in the other two visualizations. This pattern appears more consistent with adaptive definition use; the applied definition aligns with the accessible information structure in the respective visualizations. Simultaneous use of both definitions was observed only in a few cases. Overall, the observed patterns provide only limited evidence of adaptive strategy use among students who demonstrated flexibility in their definition use across problems.

Table 10. Definitions used across visualizations and information formats by the 17 students who varied their use of definitions

No.	Visualization and information format[1]						Correctly solved problems (out of three)
	2×2 table		Double tree		Net diagram		
	Probabilities	Natural frequencies	Probabilities	Natural frequencies	Probabilities	Natural frequencies	
1	Joint			Conditional	Joint		3
2	Joint			Conditional		Joint	1
3		Joint	Conditional		Joint		2
4		Joint	Conditional		Joint		2
5		Joint	Conditional		Joint		3
6		Joint	Conditional		Joint		3
7		Joint		Conditional	Joint		3
8		Joint		Conditional	Joint		3
9	Joint		Conditional			Conditional	3
10	Joint			Conditional	Conditional		3
11		Joint	Conditional		Conditional		3
12		Joint	Joint			Conditional	2
13	Joint			Joint		Both	3
14		Conditional		Joint	Joint		0
15		Conditional		Conditional	Joint		3
16	Both			Joint	Joint		3
17	Both			Joint	Both		3

[1] Empty cells indicate that the visualization in the respective information format was not presented to the student. Shaded cells indicate correctly solved problems.

6. DISCUSSION AND CONCLUSION

6.1. ANALYZING THE (IN)DEPENDENCE OF EVENTS

This study investigated the performance of $N = 319$ students from Bavaria, Germany, in analyzing the (in)dependence of two binary events across three different visualizations (2×2 tables, double trees, and net diagrams) and two information formats (probabilities and natural frequencies). In addition, we examined which definition(s) of independence students applied, how the applied definition(s) affected their performance, and whether students demonstrated flexibility in their definition use across problems.

Although students' performance in analyzing (in)dependent events when information was presented in natural frequencies demonstrated little variation across visualizations, the performance was surprisingly low for double trees in probability format, for which identifying (in)dependence only requires comparing two visually displayed probabilities (e.g., $P(A)$ and $P(A|B)$; Wiesner & Binder, 2024). Participants' relatively strong performance in 2×2 tables with probabilities may also be due to students' familiarity with this visualization in classroom practice (ISB, 2025). However, students' performance in assessing the (in)dependence of events across all visualizations and information formats was highest in the unfamiliar net diagram, which indicates that performance is not systematically associated with the familiarity of visualizations.

Participants showed a strong tendency to rely on the joint probability definition. Nearly half of the students (49%) consistently applied this definition in all three problems. In contrast, only 5% of students consistently relied on the conditional probability definition. Students' persistence in using the joint probability definition suggests that many students demonstrated limited flexibility in the sense of accessing and switching between different strategies (Selter, 2009; Star & Rittle-Johnson, 2008). With respect to strategy adaptivity, that is, selecting the most appropriate strategy for a given context (Selter, 2009), this pattern is particularly noteworthy. Although both definitions can be applied across visualizations, double trees and net diagrams in probability format make conditional relationships salient, rendering the conditional probability definition more efficient. However, students' persistence

in using the joint probability definition in these visualizations suggests limited adaptivity in definition use.

Further insight is provided by the sample's subset of 17 students who used both definitions at least once across the three problems. This group showed the highest mean accuracy in assessing (in)dependence and had the best average mathematics grade. Although this finding may partly reflect their generally higher achievement level in mathematics (e.g., Kuncel et al., 2004), it might also suggest that flexible strategy use is associated with better performance (Verschaffel, 2024). At the same time, their strategy use was not fully adaptive. Within this sample of 17 students, six out of seven participants (86%) who were presented with a double tree in probability format used the conditional probability definition for this visualization. In contrast, only 3 out of 13 students (23%) who received a probability net diagram used the conditional probability definition. Thus, even among higher-achieving students, strategy selection may still be linked to whether joint probabilities are explicitly displayed.

Interestingly, students' definition choice was not systematically associated with differences in performance rates. In double trees with probabilities, 57% of the 77 students applying the joint probability definition correctly analyzed the (in)dependence of events, compared to 58% of the 12 students applying the conditional probability definition. A similarly small difference was found in probability net diagrams (see Table 7). These performance comparisons should be interpreted with caution due to unequal group sizes. However, given the difference in the number of inferential steps required by the two definitions in these visualizations, the similarity in performance rates is surprising: If students were adaptively selecting definitions based on the displayed information, one would expect higher performance among those applying the conditional probability definition.

An additional exploratory finding is that students had more difficulty in identifying *independence* than dependence of events. This general pattern may reflect misleading intuitive categorizations of independence that often lead students to classify events as dependent, despite their actual relation. For instance, students who adopt a chronological interpretation of independence might intuitively suggest that "taking the bus" and "arriving late for school" are dependent because one follows the other in temporal succession (Díaz et al., 2010; Nabbout-Cheiban, 2017). Moreover, identifying events as dependent is less sensitive to computational errors or inaccuracies because any deviation from equality relations suffices to indicate dependence. On the other hand, analyzing the (in)dependence of events requires reasoning about equality of probabilities, which may explain why independence proved more demanding for students. Future research should investigate the contrast in students' thinking about independence and dependence in more detail.

6.2. LIMITATIONS

Because this study relied on a paper-and-pencil design, students' reasoning processes were not captured in detail. The data provide information about written products, but not the underlying cognitive processes. It remains unclear how students who provided no calculation but simply classified events as "independent" or "dependent" arrived at their conclusions. For instance, in probability double trees and probability net diagrams, participants could infer (in)dependence directly from the displayed probabilities without performing any calculation. Such reasoning would be valid but would not be captured by our coding scheme. However, the relatively low performance rate in analyzing (in)dependent events among responses coded as not using an independence definition suggests that this explanation likely applies only to a small subset of cases. Nevertheless, the present design does not allow us to distinguish between intuitive judgements, implicit reasoning, and explicit definitional strategy use. Future research employing qualitative methods, such as think-aloud protocols or process-tracing approaches, would be valuable for investigating how students analyze the independence of events in more detail.

Moreover, because this study was conducted within one federal state of Germany, the findings are necessarily situated within a specific curricular context. However, there is a common overarching curriculum in Germany including the concepts of independence and conditional probabilities (Kultusministerkonferenz, 2012). Because teachers followed the officially authorized textbooks and curricular standards, the instruction of independence can be assumed to reflect common classroom practice. Nevertheless, we did not collect data on how closely teachers followed these guidelines in terms of emphasis, sequencing, and problem selection. Thus, variation in instruction between

participants cannot be ruled out. Future research could investigate more systematically how instruction influences students' strategy use and conceptual understanding of independence.

Finally, students had prior experience with 2×2 tables and (single) tree diagrams, but no prior experience with double trees and net diagrams. Although double trees can be regarded as an extension of tree diagrams, it remains unclear how students perceived this visualization. The present study did not systematically investigate how prior knowledge of tree diagrams may support transfer to related representations. Future research could therefore examine how students' thinking about independence can be leveraged to foster their flexible and adaptive strategy use across related visualizations, such as double trees and net diagrams.

6.3. CONCLUSION

When analyzing the (in)dependence of events, students predominantly relied on the joint probability definition, even in visualizations in which conditional probabilities are made particularly salient. However, based on the curricular standards and textbooks used (Bayerisches Staatsministerium für Unterricht und Kultus, 2025; ISB, 2025), students had been exposed to both definitions in school instruction before. This fact makes the finding especially striking. Although these textbooks follow recommendations in the literature to introduce independence via conditional probabilities (e.g., Ahlgren & Garfield, 1991; Díaz et al., 2010; Tarr & Lannin, 2005; Wiesner & Binder, 2024), they often explicitly or implicitly present the joint probability statement as a "multiplication rule" that can be routinely applied to determine the (in)dependence of events (e.g., Distel, 2023, p. 87). Our findings suggest that the "multiplication rule" may function as a "default strategy" in analyzing independence. Persistent reliance on the joint probability definition across familiar and unfamiliar visualizations is consistent with the notion of an instructional mental set reinforced through repeated procedural use (Schultz & Searleman, 2002).

From an instructional perspective, these results indicate that introducing multiple definitions is not sufficient to foster students' flexible and adaptive strategy use. Rather, teaching may need to more explicitly integrate the joint and conditional probability formulations, emphasize their connections, and engage students in comparing their efficiency across visualizations. In particular, instruction could make explicit which probabilistic relationships are foregrounded by different visualizations and encourage students to select the most appropriate strategy within a context. The presented findings suggest that such a flexible and adaptive use of independence definitions has not yet been sufficiently achieved.

ACKNOWLEDGEMENTS

We would like to thank all participating students. We also thank the editor, Susan Peters, and the reviewers for their valuable feedback on earlier versions of this paper.

REFERENCES

Ahlgren, A., & Garfield, J. (1991). Analysis of the probability curriculum. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 107–134). Springer. https://doi.org/10.1007/978-94-011-3532-0_4

Andam, E. A., Awuah, F. K., & Obeng-Denteh, W. (2025). Analysis of grade 11 students' conceptual understanding in probability concepts: A perspective of Skemp understanding theory. *East African Journal of Education Studies*, *8*(2), 358–374. https://doi.org/10.37284/eajes.8.2.2987

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, *112*(1), 117–135. https://doi.org/10.1037/0096-3445.112.1.117

Australian Curriculum, Assessment and Reporting Authority. (2015). *Mathematics: Sequence of content*. https://docs.acara.edu.au/resources/Mathematics_-_Sequence_of_content.pdf

Batanero, C., & Borovcnik, M. (2016). *Statistics and probability in high school*. Sense Publishers. https://doi.org/10.1007/978-94-6300-624-8

Batanero, C., Chernoff, E. J., Engel, J., Lee, H. S., & Sánchez, E. (2016). *Research on teaching and learning probability*. Springer. https://doi.org/10.1007/978-3-319-31625-3_1

Batanero, C., Contreras, J. M., Díaz, C., & Sánchez, E. (2015). Prospective teachers' semiotic conflicts in computing probabilities from a two-way table. *International Electronic Journal of Mathematics Education*, *10*(1), 3–16. https://doi.org/10.29333/iejme/288

Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, *27*(2), 151–169. https://doi.org/10.2307/749598

Bayerisches Staatsministerium für Unterricht und Kultus. (2025). *Zugelassene Lernmittel* [Approved learning materials]. https://www.km.bayern.de/unterrichten/unterrichtsalltag/lernmittel

Biburger, J., Frantzke, A., Herbst, M., Kestler, C., Seibold, C., Sy, B., Thiessen, T., & Weber, M. (2023). *Lambacher Schweizer Mathematik für Gymnasium 11: Ausgabe Bayern ab 2023* [Lambacher Schweizer mathematics for advanced academic high school 11: Bavaria edition 2023]. Ernst Klett.

Biehler, R., & Engel, J. (2015). Stochastik: Leitidee Daten und Zufall [Stochastics: Core ideas of data and chance]. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Eds.), *Handbuch der Mathematikdidaktik* (pp. 221–251). Springer. https://doi.org/10.1007/978-3-642-35119-8_8

Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information – an empirical study on tree diagrams and 2×2 tables. *Frontiers in Psychology*, *6*, Article 1186. https://doi.org/10.3389/fpsyg.2015.01186

Binder, K., Krauss, S., & Wiesner, P. (2020). A new visualization for probabilistic situations containing two binary events: The frequency net. *Frontiers in Psychology*, *11*, Article 750. https://doi.org/10.3389/fpsyg.2020.00750

Binder, K., Steib, N., & Krauss, K. (2023). Von Baumdiagrammen über Doppelbäume zu Häufigkeitsnetzen – kognitive Überlastung oder didaktische Unterstützung? [Moving from tree diagrams to double trees to net diagrams – cognitively overwhelming or educationally supportive?] *Journal für Mathematik-Didaktik*, *44*(2), 471–503. https://doi.org/10.1007/s13138-022-00215-9

Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429428357

Bolker, B. M. (2015). Linear and generalized linear mixed models. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), *Ecological statistics: Contemporary theory and application* (1st ed., pp. 309–333). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199672547.003.0014

Borovcnik, M. (2012). Perspectivas múltiples sobre el concepto de probabilidad condicional [Multiple perspectives on the concept of conditional probability]. *Avances de Investigación en Educación Matemática*, *2*, 5–27. https://doi.org/10.35763/aiem.v1i2.32

Borovcnik, M., & Bentz, H.-J. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 73–105). Springer. https://doi.org/10.1007/978-94-011-3532-0_3

Brendel, A., Schmidt-Kessel, M., & Eisentraut, F. (2024). *Mathe.delta: Mathematik für das Gymnasium 11 Bayern* [Mathe.delta: Mathematics for advanced academic high school grade 11 (Bavaria)] (2nd ed.). C. C. Buchner.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(421), 9–25. https://doi.org/10.1080/01621459.1993.10594284

Cordani, L. K., & Wechsler, S. (2006). Teaching independence and exchangeability. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh International Conference on Teaching Statistics (ICOTS-7), Salvador, Brazil*. International Statistical Institute. https://www.stat.auckland.ac.nz/~iase/publications/17/3I1_CORD.pdf

D'Amelio, A. (2009). Undergraduate student difficulties with independent and mutually exclusive events concepts. *The Mathematics Enthusiast*, *6*(1–2), 47–56. https://doi.org/10.54870/1551-3440.1133

de Moivre, A. (1756). *The doctrine of chances* (3rd ed.). A. Millar.

Díaz, C., Batanero, C., & Contreras, J. M. (2010). Teaching independence and conditional probability. *Boletín de Estadística e Investigación Operativa*, *26*(2), 149–162.

Distel, B. (2023). *Fundamente der Mathematik–Bayern–Gymnasium–11. Jahrgangsstufe* [Foundations of mathematics–Bavaria–advanced academic high school–grade 11]. Cornelsen.

Falk, R. (1979). Revision of probabilities and the time axis. In D. Tall (Ed.), *Proceedings of the third conference for the Psychology of Mathematics Education* (pp. 64–66). International Group for the Psychology of Mathematics Education.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105. https://doi.org/10.2307/749665

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, *6*(2), 187–205. https://doi.org/10.1007/BF00302543

Hilbert, S., Stadler, M., Lindl, A., Naumann, F., & Bühner, M. (2019). Analyzing longitudinal intervention studies with linear mixed models. *TPM–Testing, Psychometrics, Methodology in Applied Psychology*, *26*(1), 101–119. https://doi.org/10.4473/TPM26.1.6

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, *79*(1), 1–17. https://doi.org/10.1037/h0093874

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Krauss, S., Weber, P., Binder, K., & Bruckmaier, G. (2020). Natürliche Häufigkeiten als numerische Darstellungsart von Anteilen und Unsicherheit—Forschungsdesiderate und einige Antworten [Natural frequencies as a numerical representation of proportions and uncertainty—research gaps and some answers]. *Journal für Mathematik-Didaktik*, *41*(2), 485–521. https://doi.org/10.1007/s13138-019-00156-w

Krüger, K., Sill, H.-D., & Sikora, C. (2015). *Didaktik der Stochastik in der Sekundarstufe I* [Teaching stochastics in lower secondary education]. Springer. https://doi.org/10.1007/978-3-662-43355-3

Kultusministerkonferenz. (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife* [Educational standards in mathematics for the general higher education entrance qualification]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*(1), 148–161. https://doi.org/10.1037/0022-3514.86.1.148

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*(6), i–95. https://doi.org/10.1037/h0093502

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, *143*(12), 1273–1312. https://doi.org/10.1037/bul0000126

Morsanyi, K., Primi, C., Chiesi, F., & Handley, S. (2009). The effects and side-effects of statistics education: Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology*, *34*(3), 210–220. https://doi.org/10.1016/j.cedpsych.2009.05.001

Mošna, F. (2023). Secondary and university students' understanding of independence and conditional probability. *Scientia in Educatione*, *14*(1), 15–26. https://doi.org/10.14712/18047106.3013

Nabbout-Cheiban, M. (2017). Intuitive thinking and misconceptions of independent events: A case study of US and French pre-service teachers. *International Journal of Research in Undergraduate Mathematics Education*, *3*(2), 255–282. https://doi.org/10.1007/s40753-016-0038-x

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nasser, R. (2007). Deciding whether there is statistical independence or not? *Journal of Mathematics and Statistics*, *3*(3), 151–162. https://doi.org/10.3844/jmssp.2007.151.162

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/Principles,-Standards,-and-Expectations/

O'Grady, C. (2022). Unlucky numbers. *Science*, *379*(6629), 228–233. https://doi.org/10.1126/science.adg6746

Olofsson, P., & Andersson, M. (2012). *Probability, statistics, and stochastic processes*. John Wiley & Sons. https://doi.org/10.1002/9781118231296

Plaxco, D. (2011). The temporal conception: Student difficulties defining probabilistic independence. In L. R. Wiest & T. Lamberg (Eds.), *Proceedings of the 33rd annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education, Reno, Nevada* (pp. 276–283). University of Nevada.

Rößner, M., Binder, K., Geier, C., & Krauss, S. (2025). Students' performance and typical errors in filling empty probabilistic visualizations with probabilities or frequencies. *Educational Studies in Mathematics*, *120*(1), 137–167. https://doi.org/10.1007/s10649-024-10372-y

Schneps, L., & Colmez, C. (2013). *Math on trial: How numbers get used and abused in the courtroom*. Basic Books.

Schultz, P. W., & Searleman, A. (2002). Rigidity of thought and behavior: 100 years of research. *Genetic, Social, and General Psychology Monographs*, *128*(2), 165–207.

Selter, C. (2009). Creativity, flexibility, adaptivity, and strategy use in mathematics. *ZDM–Mathematics Education*, *41*(5), 619–625. https://doi.org/10.1007/s11858-009-0203-7

Staatsinstitut für Schulqualität und Bildungsforschung. (2025). *LehrplanPLUS Bayern Gymnasium Jahrgangsstufe 11 Mathematik Fachlehrpläne* [LehrplanPLUS Bavaria advanced academic high school grade 11 mathematics subject curricula]. https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/11/mathematik

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, *18*(6), 565–579. https://doi.org/10.1016/j.learninstruc.2007.09.018

Star, J. R., Tuomela, D., Joglar-Prieto, N., Hästö, P., Palkki, R., Abánades, M. Á., Pejlare, J., Jiang, R. H., Li, L., & Liu, R.-D. (2022). Exploring students' procedural flexibility in three countries. *International Journal of STEM Education*, *9*(1), Article 4. https://doi.org/10.1186/s40594-021-00322-y

Stegmüller, N., Binder, K., & Krauss, S. (2024). How general is the natural frequency effect? The case of joint probabilities. *Frontiers in Psychology*, *15*, Article 1296359. https://doi.org/10.3389/fpsyg.2024.1296359

Tarr, J. E., & Jones, G. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, *9*(1), 39–59. https://doi.org/10.1007/BF03217301

Tarr, J. E., & Lannin, J. K. (2005). How can teachers build notions of conditional probability and independence? In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 215–238). Springer. https://doi.org/10.1007/0-387-24530-8_10

Threlfall, J. (2009). Strategies and flexibility in mental calculation. *ZDM–Mathematics Education*, *41*(5), 541–555. https://doi.org/10.1007/s11858-009-0195-3

Truran, J. M., & Truran, K. M. (1996). Statistical independence – One concept or two? Implications for research and for classroom practice. In B. Phillips (Ed.), *Papers on statistical education presented at ICME-8 (International Congress on Mathematics Education-8)* (pp. 87–100). Swinburne Press.

Twisk, J. W. R. (2019). *Applied mixed model analysis: A practical guide* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108635660

Verschaffel, L. (2024). Strategy flexibility in mathematics. *ZDM–Mathematics Education*, *56*(1), 115–126. https://doi.org/10.1007/s11858-023-01491-6

Weber, P., Binder, K., & Krauss, S. (2018). Why can only 24% solve Bayesian reasoning problems in natural frequencies: Frequency phobia in spite of probability blindness. *Frontiers in Psychology*, *9*, Article 1833. https://doi.org/10.3389/fpsyg.2018.01833

Wiemann, L., Schroeder, S., & Halverscheid, S. (2025). The influence of mathematical achievement on the development of intuitive probabilistic reasoning in lower secondary school students. In C. Cornejo, P. Felmer, D. M. Gómez, P. Dartnell, P. Araya, A. Peri, & V. Randolph (Eds.),

Proceedings of the 48th conference of the International Group for the Psychology of Mathematics Education: Research reports (Vol. 2, pp. 387–394). PME.

Wiesner, P., & Binder, K. (2024). Stochastische Unabhängigkeit: Die Rolle der gewählten Definition und verschiedener Visualisierungen [Stochastic independence: The role of the chosen definition and different visualizations]. *Stochastik in der Schule*, *44*(3), 21–29.

LEON WIEMANN
Paderborn University
Warburger Str. 100
33098 Paderborn
Germany

APPENDIX: VISUALIZATIONS USED IN THIS STUDY

In the following, the visualizations with information formats used in this study are presented, separated by context ("School arrival," "Treatment study," and "Quality control"), and event type (independent and dependent events). Figure 1 illustrates the visualizations and information formats for the "School arrival" context with independent events.

Visualizations in the 'school arrival' context with dependent events

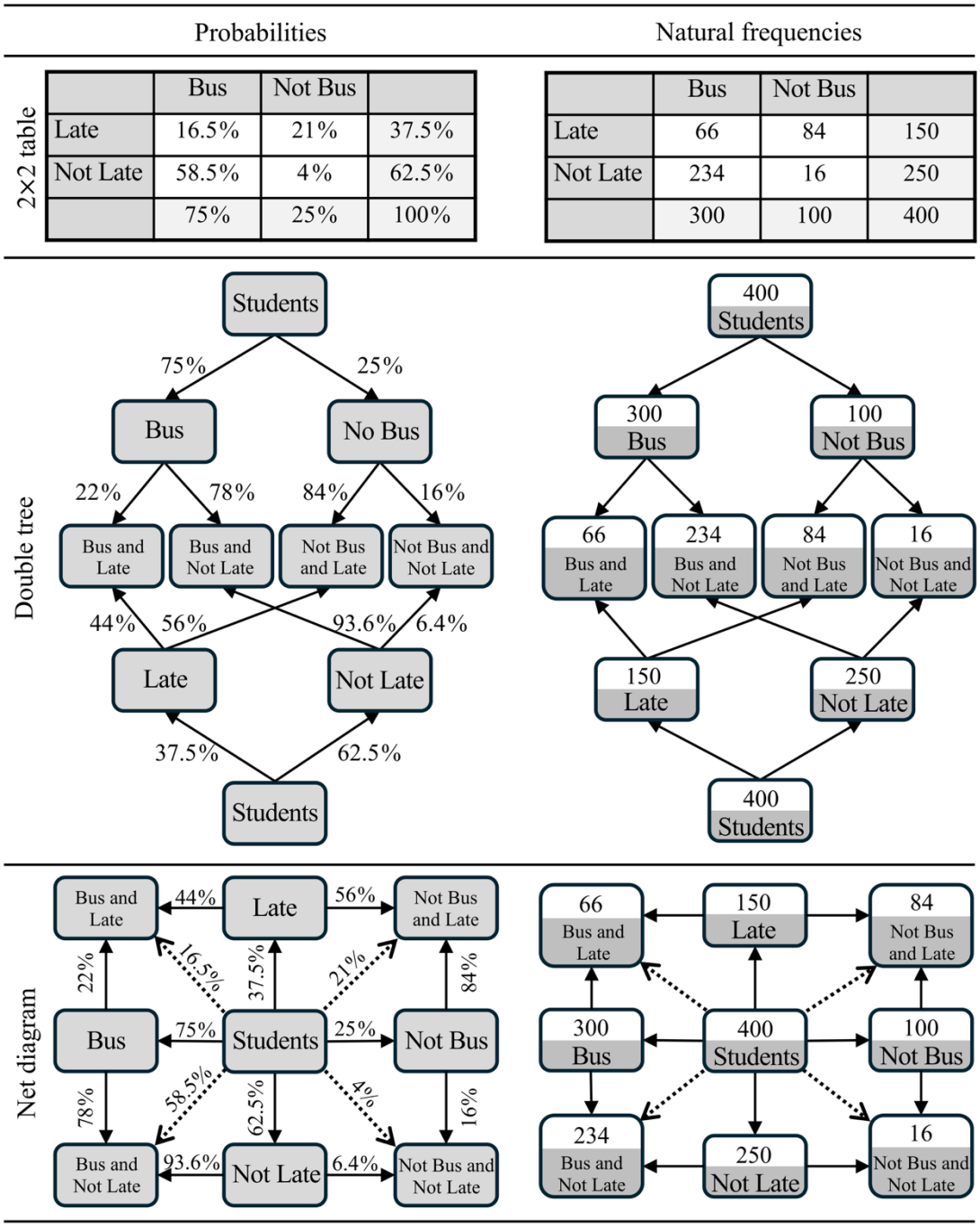


Figure A1. Visualizations in both information formats in the 'school arrival' context with two binary dependent events

Visualizations in the 'treatment study' context with independent events

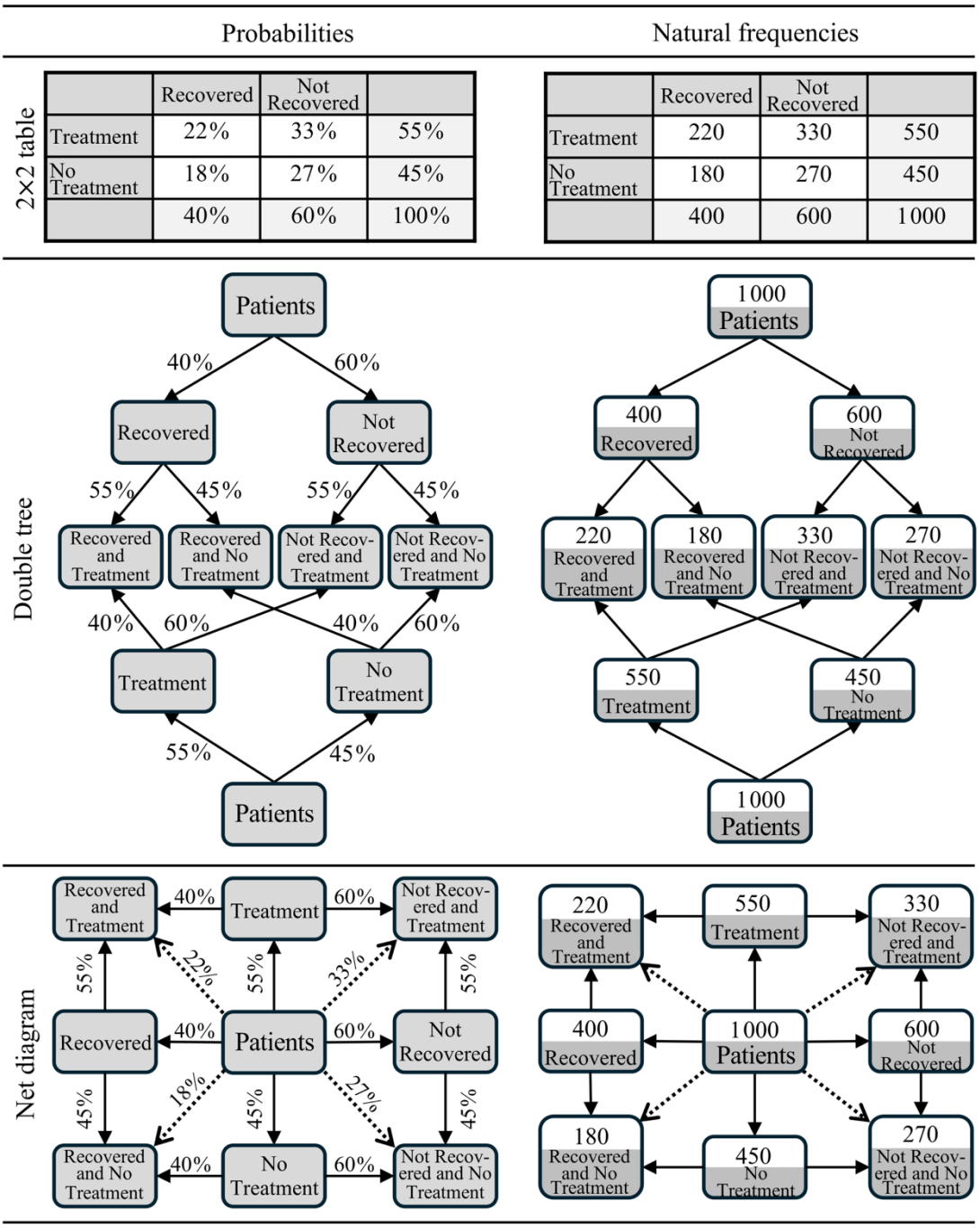


Figure A2. Visualizations in both information formats in the 'treatment study' context with two binary independent events

Visualizations in the 'treatment study' context with dependent events

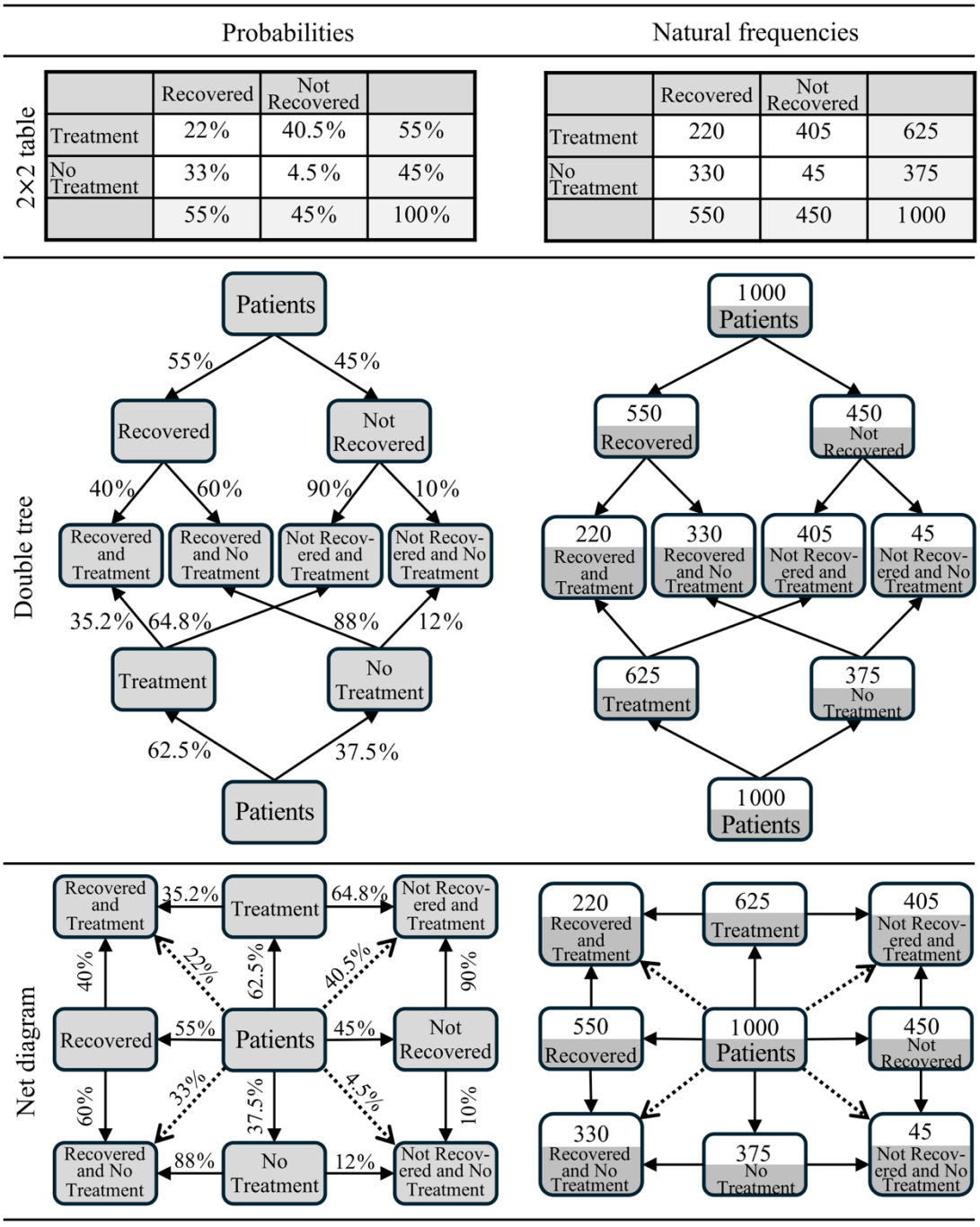


Figure A3. Visualizations in both information formats in the 'treatment study' context with two binary dependent events

Visualizations in the 'quality control' context with independent events

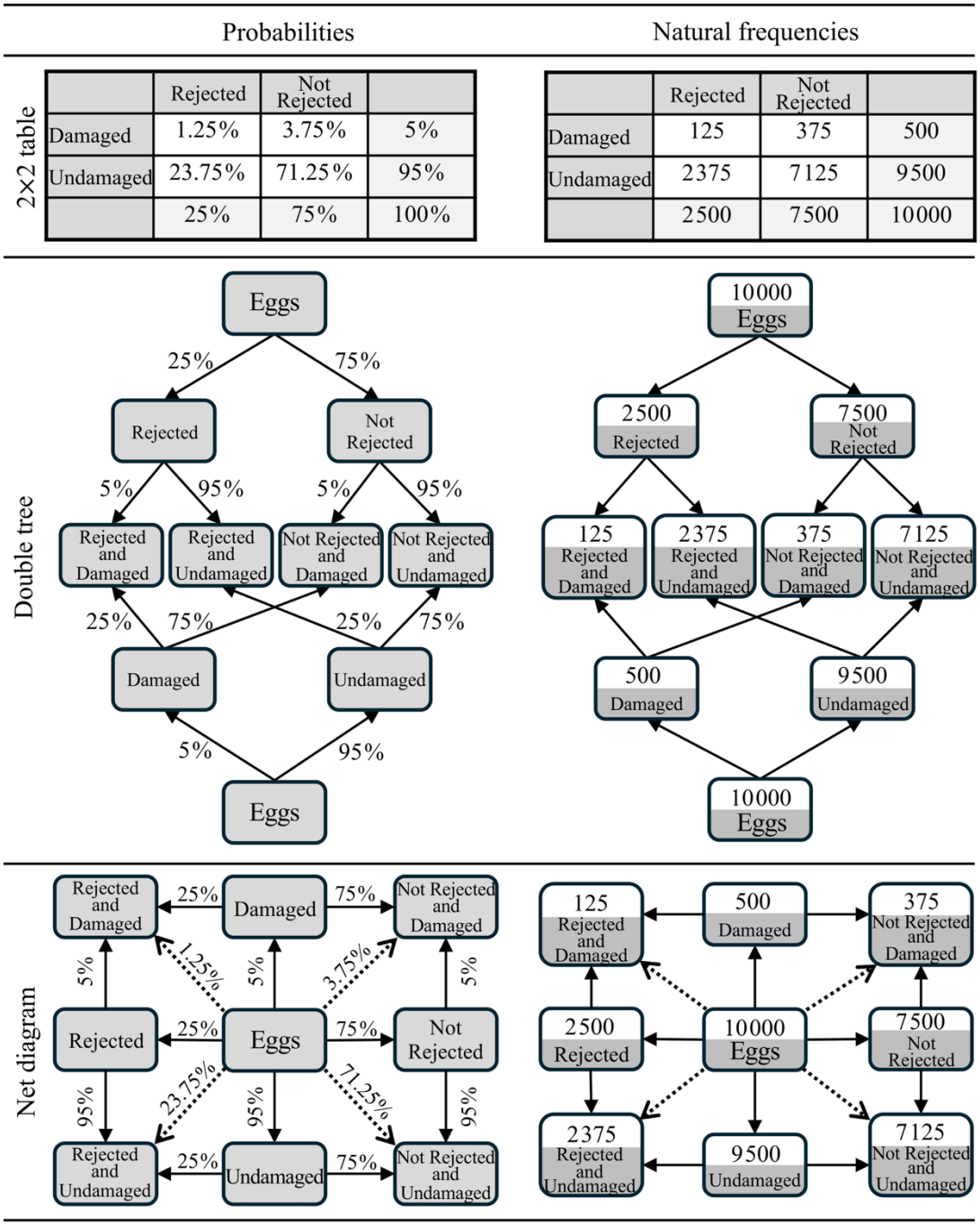


Figure A4. Visualizations in both information formats in the 'quality control' context with two binary independent events

Visualizations in the 'quality control' context with dependent events

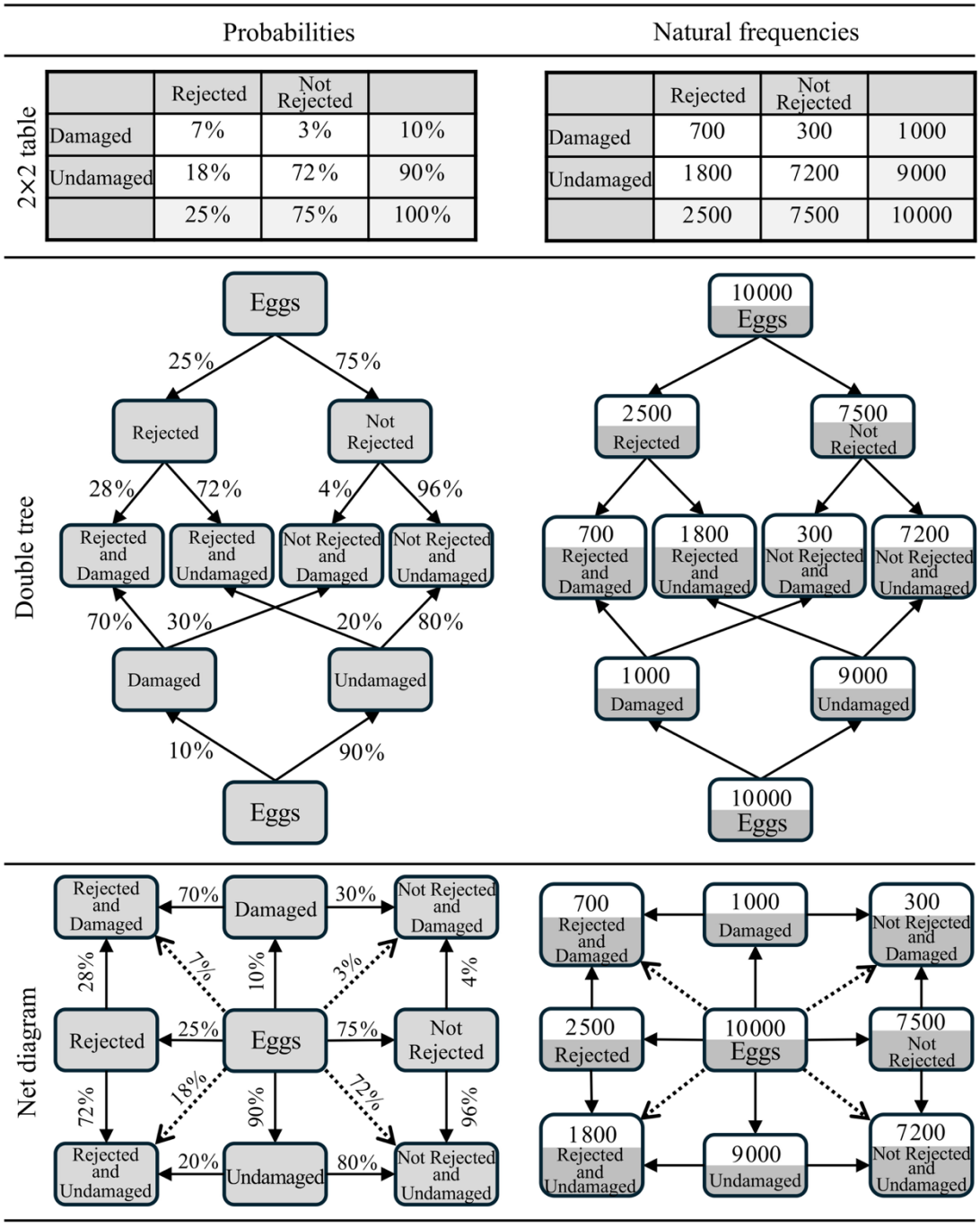


Figure A5. Visualizations in both information formats in the 'quality control' context with two binary dependent events