



Wird die Vierfeldertafel oder das Baumdiagramm beim Einschätzen von Risiken präferiert? – Ergebnisse einer Eyetracking-Studie

Julia Sirock · Georg Bruckmaier · Stefan Krauss · Markus Vogel

Eingegangen: 15. Januar 2025 / Angenommen: 3. Juli 2026
© The Author(s) 2026

Zusammenfassung In der vorliegenden Studie werden die Präferenz und die Performanz von Schüler:innen der Sekundarstufe bei der Nutzung der beiden schulüblichen Visualisierungen Baumdiagramm und Vierfeldertafel sowie einer rein textuellen Darstellung zur Lösung Bayesianischer Aufgaben untersucht. Bayesianisches Denken ist ein zentraler Bestandteil der Risikokompetenz (Risk Literacy) und essenziell für fundierte Entscheidungen, etwa in Fragen der Rechtsprechung oder bei der Beurteilung von Gesundheitsfragen. Visualisierungen können dabei helfen, die Problemzusammenhänge zu verdeutlichen. Jedoch ist derzeit noch unklar, welche Visualisierungen Problemlöser:innen von sich aus bevorzugen.

In der vorliegenden Studie wurden zur Beantwortung dieser Frage mittels Eyetracking und Kurzinterviews die Präferenzen hinsichtlich je zwei gleichzeitig präsentierten Darstellungsarten von insgesamt 36 Schüler:innen der Klassenstufen 5/6 und 10 in einem dreifaktoriellen experimentellen Design untersucht. Dabei wurden empirisch die Präferenzvergleiche a) Baumdiagramm vs. Vierfeldertafel, b) Baumdiagramm vs. Text, und c) Vierfeldertafel vs. Text implementiert.

Die Ergebnisse zeigen, dass das Baumdiagramm insgesamt gegenüber der Vierfeldertafel präferiert wird. Die rein textuelle Darstellung der Informationen wird nur sehr vereinzelt einer der beiden Visualisierungen vorgezogen. Die Präferenz, aber auch die Performanz werden zudem von der Klassenstufe beeinflusst. So zeigt sich zwar das Baumdiagramm in beiden Klassenstufen als bevorzugte Visualisierung, in Klassenstufe 10 wird jedoch individuell auch häufig die Vierfeldertafel präferiert,

✉ Julia Sirock · Markus Vogel
Pädagogische Hochschule Heidelberg, Heidelberg, Deutschland
E-Mail: sirock@ph-heidelberg.de

Georg Bruckmaier
Pädagogische Hochschule FHNW, Windisch, Schweiz

Stefan Krauss
Universität Regensburg, Regensburg, Deutschland

während in Klassenstufe 5/6 gelegentlich auch die rein textuelle Präsentation der statistischen Informationen präferiert wird.

Schlüsselwörter Bayesianisches Denken · Visualisierungen · Visuelle Aufmerksamkeit · Methodentriangulation · Klassenstufenunterschiede

Preference for 2×2-Tables Versus Tree Diagrams in Risk Assessment: An Eye-Tracking Study

Abstract This study examined the preferences and performance of secondary school students when using the two common visualization methods taught in schools—tree diagrams and 2×2 tables—as well as a purely textual representation to solve Bayesian problems. Bayesian thinking is a central component of risk literacy and essential for making informed decisions, for example, in legal matters or when assessing health issues. The use of visualizations can help clarify the relationships within a problem. However, it is currently unclear which visualizations problem solvers prefer on their own. To answer this question, with a three-factor experimental design the present study used eye-tracking and brief interviews to examine the preferences of a total of 36 students in grades 5/6 and 10 regarding two simultaneously presented representation types. The study empirically compared preferences for (a) a tree diagram versus a 2×2 table (b) a tree diagram versus text, (c) a 2×2 table versus text. The results show that, overall, the tree diagram was preferred over the 2×2 table. Very rarely, the purely textual presentation of the information was preferred over either of the two visualizations. Both preference and performance were also influenced by grade level. While the tree diagram emerged as the preferred form of visualization in both grade levels, 10th-grade students often preferred the 2×2 table on an individual basis, whereas 5th- and 6th-grade students occasionally preferred the purely textual presentation of statistical information.

Keywords Bayesian thinking · Visualizations · Visual attention · Methodological triangulation · Grade-level-differences

1 Einleitung

Risk Literacy (Aven 2024) spielt sowohl im Alltag als auch in Domänen wie Medizin, Rechtsprechung oder Wirtschaftswissenschaften eine große Rolle (Bruckmaier et al. 2021). Insbesondere die Kompetenz, Wahrscheinlichkeiten korrekt zu interpretieren und darauf basierende Entscheidungen zu treffen, ist ein zentraler Bestandteil von Risikokompetenz, die entscheidend für eine mündige Teilhabe am gesellschaftlichen Leben ist (Eichler und Vogel 2015; Gigerenzer 2012). In einer zunehmend digitalisierten Informationsgesellschaft stehen wissenschaftlich fundierte Risikoinformationen neben einer Vielzahl algorithmisch kuratierter Inhalte, die auf Aufmerksamkeit und emotionale Aktivierung ausgerichtet sind. Plattformen wie TikTok, Instagram oder YouTube operieren häufig nach Mechanismen kurzfristiger

Reizverstärkung. Entscheidend ist daher, seriöse Risikoinformationen so zu gestalten, dass sie nicht nur verständlich, sondern auch visuell ansprechend und emotional anschlussfähig sind. Vor diesem Hintergrund kommt der sachlich verständlichen und visuell ansprechenden Aufbereitung von Risikoinformationen besondere Bedeutung zu. Denn wer Risiken nicht adäquat einschätzen kann, ist anfälliger für Fehlentscheidungen in Bereichen wie Gesundheit, Finanzen oder politischer Urteilsbildung.

Ein zentraler Bestandteil von Risk Literacy ist dabei die Fähigkeit zum sogenannten „Bayesianischen Denken“ (Gigerenzer 2015). Diese Fähigkeit wird im schulischen Stochastikunterricht insbesondere beim Thema bedingte Wahrscheinlichkeiten verortet. Bayesianisches Denken ist bei Aufgaben erforderlich, die mit dem Satz von Bayes gelöst werden können. Alternativ können Aufgaben dieser Art auch mit den Pfadregeln, durch die Übersetzung der statistischen Informationen in natürliche Häufigkeiten (z. B. durch Angaben der Art „4 von 10“ statt „40 %“, vgl. Abschn. 2.2.1) oder auf der Basis von Visualisierungen gelöst werden (Steib et al. 2025). Die Förderung entsprechender Kompetenzen stellt jedoch eine didaktische Herausforderung dar, da Lernende bei der Interpretation bedingter Wahrscheinlichkeiten häufig systematische Fehlvorstellungen zeigen. Vor diesem Hintergrund rückt die Frage in den Fokus, wie probabilistische Informationen so dargestellt werden können, dass sie das Verständnis und die korrekte Urteilsbildung unterstützen. Ein zentraler Aspekt des vorliegenden Beitrags ist dabei die avisierte möglichst frühzeitige Förderung, die schulisch bereits ohne das Wahrscheinlichkeitskonzept möglich ist. Auch jüngere Schüler:innen können bereits in der Unterstufe erste propädeutische Erfahrungen mit der Grundstruktur der Bayesianischen Denkweise sammeln – noch ohne den Wahrscheinlichkeitsformalismus, sondern im Sinne des Spiralcurriculums auf der Basis absoluter und natürlicher Häufigkeiten (Krauss et al. 2020).

Visualisierungen werden im vorliegenden Beitrag nicht nur als Mittel zur Strukturierung komplexer Informationen und als eine zentrale Unterstützungsmöglichkeit für Bayesianisches Denken betrachtet (Spiegelhalter 2014), sondern auch als eine ansprechende und schnell lesbare Aufbereitung statistischer Informationen. Binder et al. (2015) konnten im Hinblick auf die beiden schulüblichen Visualisierungen Vierfeldertafel und Baumdiagramm bereits zeigen, dass bei Bayesianischen Aufgaben beide Visualisierungen besonders dann verständnisfördernd wirken, wenn sie mit absoluten statt mit relativen Häufigkeiten gefüllt sind (auch wenn Baumdiagramme mit Häufigkeiten in den Knoten im schulischen Stochastikunterricht eher unüblich sind). Allerdings sind Wirkung und Nutzen von Visualisierungen nicht für alle Personen universell gleich, sondern auch individuelle Faktoren wie Vorwissen und kognitive Fähigkeiten beeinflussen, wie gut Lernende mit bestimmten Visualisierungen umgehen können (Bruckmaier et al. 2021). Vor dem Hintergrund individueller kognitiver Voraussetzungen sowie digital geprägter Informationsumgebungen stellt sich die Frage, wie Risikoinformationen so gestaltet und vermittelt werden können, dass sie auch unter Bedingungen begrenzter Aufmerksamkeit verständlich und kognitiv anschlussfähig bleiben.

Während es seit etwa 30 Jahren eine große Fülle an Studien zur Performanz bei Bayesianischen Aufgaben gibt, die die Wirkung unterschiedlicher Informationsformate wie relative oder absolute Häufigkeiten sowie verschiedene Visualisierungen untersuchen (für eine Metaanalyse siehe McDowell und Jacobs 2017), existieren bis-

lang nur wenige Untersuchungen dazu, welche Visualisierung beim Lösen solcher Aufgaben *präferiert* wird (für eine Ausnahme im Bereich bedingter Wahrscheinlichkeiten siehe z.B. Rach 2021; vgl. Abschn. 2.2.3). Da solche Bayesianischen Aufgaben zwar nur eine spezielle Variante aus dem schulischen Aufgabenset zu bedingten Wahrscheinlichkeiten darstellen, diese aber gerade für die Risikokompetenz von hervorgehobener Bedeutung sind (Gigerenzer und Hoffrage 1995), ist es nicht nur wichtig, die Performanz bei gegebenen Visualisierungen zu untersuchen, sondern auch die intuitive Präferenz zwischen verschiedenen Darstellungen statistischer Informationen.

Im vorliegenden Beitrag werden sowohl die Präferenz als auch die Performanz (mit der dann präferierten Visualisierung) im Hinblick auf drei häufigkeitsbasierte Informationsdarstellungen (Baumdiagramm, Vierfeldertafel und rein textuelle Darstellung) im Kontext Bayesianischer Aufgaben untersucht. Als Messverfahren kommen neben Eyetracking ergänzend Kurzinterviews zum Einsatz, um die den Präferenzen und Performanzdaten zugrunde liegenden kognitiven Prozesse besser verstehen und die gewonnenen Eyetracking-Daten gegebenenfalls validieren zu können. Ziel ist es, durch die Kombination der Analyse von Präferenz und entsprechender Performanz Einblicke in das Bayesianische Denken von Schüler:innen zu gewinnen, die durch die alleinige Analyse nur eines dieser beiden Kriterien nicht erzielt werden können. Letztendlich könnten hierdurch Visualisierungen identifiziert werden, die Risiken nicht nur transparent kommunizieren, sondern aufgrund ihrer höheren Attraktivität in öffentlichen Diskursen auch zu größerer Aufmerksamkeit führen könnten.

Im Folgenden wird zunächst der theoretische Hintergrund zu den Themen Risk Literacy, Bayesianisches Denken und zur Bedeutung von Visualisierungen skizziert (Abschn. 2). Nach der Formulierung von Forschungsfragen (Abschn. 3) wird die Methode einschließlich Stichprobe, Materialien, Durchführung und Pilotierung beschrieben (Abschn. 4). Die Ergebnisse der Analysen zu Präferenzen und Performanzen werden in Abschnitt 5 dargestellt. In Abschnitt 6 werden die Befunde diskutiert sowie theoretische und praktische Implikationen erörtert.

2 Theoretischer Hintergrund

2.1 Risikokompetenz (Risk Literacy)

Risikokompetenz (engl.: Risk Literacy) beschreibt die Fähigkeit, Risiken zu verstehen, zu bewerten und fundierte Entscheidungen zu treffen (für eine detailliertere Aufteilung in die Begriffe Risiko, Ambiguität, Unsicherheit und Nicht-Berechenbarkeit siehe Gigerenzer 2024). Risk Literacy ist in vielen Lebensbereichen von Bedeutung, etwa bei Gesundheits-, Sicherheits- oder finanziellen Risiken (Lusardi 2015). Dies lässt sich anhand eines Beispiels illustrieren: Die Prävalenz einer bestimmten Krankheit beträgt in der Bevölkerung 1 %, das heißt, im Durchschnitt sind z. B. 100 von 10.000 Menschen betroffen. Solche Basisraten bilden eine zentrale Grundlage, um individuelle Risiken einschätzen zu können. Um beispielsweise nach einem medizinischen Test das tatsächliche Erkrankungsrisiko korrekt zu bestimmen, müssen weitere Wahrscheinlichkeiten wie Richtig-Positiv-Rate und Falsch-Positiv-

Rate eines medizinischen Tests einbezogen und im Verhältnis zur Basisrate interpretiert werden. Wer Risikokompetenz besitzt, versteht, dass eine positive Diagnose (z. B. ein positives Testergebnis) nicht automatisch bedeuten muss, dass eine Krankheit auch mit hoher Wahrscheinlichkeit vorliegt.

Solche Fragestellungen machen deutlich, dass Risikokompetenz eng mit probabilistischem Denken und insbesondere mit Bayesianischem Denken (vgl. Abschn. 2.2) verbunden ist, da hierfür ein Verständnis für Wahrscheinlichkeiten und deren korrekte Interpretation erforderlich ist (Cokely et al. 2012). Visualisierungen wie die Vierfeldertafel oder das Baumdiagramm helfen dabei, die relevanten Informationen (z. B. Anzahl der Erkrankten, falsch-positive und richtig-positive Testergebnisse) so darzustellen, dass sie leichter verständlich und korrekt interpretierbar sind (Büchter et al. 2022a). Dennoch fällt es vielen Menschen schwer, Risiken realistisch einzuschätzen, was häufig zu Fehlentscheidungen führt. Dies unterstreicht die Notwendigkeit, Risikokompetenz in der Schule zu fördern (Till 2014) und damit idealerweise bereits möglichst frühzeitig zu beginnen.

Insbesondere bei der Vermittlung von Risikokompetenz spielen Visualisierungen wie Vierfeldertafeln und Baumdiagramme eine Schlüsselrolle, die beide folgerichtig auch in Schulcurricula verankert sind. Empirische Studien zeigen, dass solche Darstellungen, insbesondere in Verbindung mit natürlichen Häufigkeiten, das Verständnis komplexer Informationen fördern können (Gigerenzer und Hoffrage 1995; im Detail vgl. Abschn. 2.2.2).

Der rationale Umgang mit Risiken kann erlernt werden. Gigerenzer (2020) argumentiert, dass Wissen das effektivste Mittel gegen Ängste ist und dass jede Person lernen kann, in einer Welt voller Unsicherheiten kluge Entscheidungen zu treffen. Anhand konkreter Beispiele, wie etwa der Corona-Pandemie oder dem Fukushima-Unglück, zeigt er, dass es notwendig ist, Unsicherheiten zu akzeptieren anstatt nach absoluter Sicherheit zu suchen. Dies erfordert jedoch frühzeitige Bildungsmaßnahmen, um Risikokompetenz gezielt zu fördern. Obwohl das Thema Risiko ein „hot topic“ ist (Spiegelhalter und Gage 2014, S. 1), wird es im schulischen Mathematikunterricht jedoch nur selten explizit aufgegriffen. Eichler und Vogel (2015) plädieren dafür, das Thema Risiko im Unterricht durch Modellierungsaufgaben und visuelle Darstellungen stärker zu thematisieren. Durch den Einsatz von realitätsnahen Aufgaben und Visualisierungen können Schüler:innen lernen, abstrakte Wahrscheinlichkeiten besser zu verstehen, insbesondere können junge Schüler:innen bereits in der Unterstufe auf der Basis von Häufigkeiten Einblick in derartige Denkweisen gewinnen.

2.2 Die menschliche Fähigkeit zum Bayesianischen Denken

Bayesianisches Denken als ein spezifisches Beispiel für grundlegende Risikokompetenz bezeichnet die Fähigkeit, Wahrscheinlichkeitseinschätzungen anhand neuer Informationen zu aktualisieren, was eine Grundvoraussetzung für rationales Handeln darstellt. Eine a-priori-Wahrscheinlichkeit – beispielsweise die Basisrate $p(K)$ einer Krankheit – soll mit Hilfe der Richtig-Positiv-Rate $p_K(T+)$ und der Falsch-Positiv-Rate $p_{\bar{K}}(T+)$ eines medizinischen Tests zur a-posteriori-Wahrscheinlichkeit $p_{T+}(K)$ „upgedatet“ werden (für eine diesbezügliche Anwendung des Satzes

Tab. 1 Gegenüberstellung des Wahrscheinlichkeits- und des Häufigkeitsformats bei Bayesianischen Aufgaben

Bayesianische Aufgabe	Wahrscheinlichkeitsformat	Häufigkeitsformat
Basisrate	Die Wahrscheinlichkeit, dass eine Frau, die an einem Screening teilnimmt, Brustkrebs hat, ist 1 %.	Von 10.000 Frauen, die am Screening teilnehmen, haben 100 Brustkrebs.
Richtig-Positiv-Rate	Die Wahrscheinlichkeit, dass eine Frau ein positives Mammogramm erhält, wenn sie Brustkrebs hat, beträgt 80 %, wobei 10 % der Frauen, die keinen Brustkrebs haben, ebenso ein positives Ergebnis erhalten.	80 dieser 100 Frauen erhalten ein positives Testergebnis, wobei 990 der 9900 gesunden Frauen ebenso ein positives Ergebnis erhalten.
Falsch-Positiv-Rate		
Frage	Wie wahrscheinlich ist es, dass eine Frau mit positivem Testergebnis tatsächlich erkrankt ist?	
Lösungsalgorithmus	$ppV = \frac{80\% \cdot 1\%}{80\% \cdot 1\% + 10\% \cdot 99\%} \approx 7,5\%$	$ppV = \frac{80}{80+990} = \frac{80}{1070} \approx 7,5\%$
Lösung (Positiv-prädiktiver Wert)	Die Wahrscheinlichkeit beträgt ca. 7,5 %.	80 der 1070 positiv getesteten Frauen haben Brustkrebs.
Visualisierung (exemplarisch: Baumdiagramm)		

von Bayes vgl. Tab. 1, unten links). Unter „Bayesianischem Denken“ versteht man im didaktischen und kognitionspsychologischen Diskurs aber nicht nur die bloße Anwendung des Satzes von Bayes, sondern die Fähigkeit, neue Evidenz systematisch in bestehende Wahrscheinlichkeitsurteile einzubeziehen. Gerade da Bayesianisches Denken in vielen Bereichen wie der Wahrnehmungspsychologie oder der Entscheidungsfindung eine Rolle spielt (Chater et al. 2010, 2006), stellt sich – auch unabhängig vom schulischen Stochastikunterricht – die Frage, ob Menschen intuitiv in der Lage zum Bayesianischen Denken sind.

Kahneman und Tversky identifizierten im Rahmen ihres heuristics-and-biases-Forschungsprogramms in Bezug auf Bayesianische Aufgaben kognitive Fehlschlüsse, darunter den Basisratenfehler (base-rate fallacy), bei dem Menschen die Grundwahrscheinlichkeiten (a-priori-Wahrscheinlichkeiten) von Ereignissen ignorieren bzw. deren Bedeutung unterschätzen (Kahneman und Tversky 1972; Tversky und Kahneman 1973). Ein Beispiel für die Vernachlässigung der Basisrate wird aus dem typischen Fehlschluss bei der berühmten Mammographie-Aufgabe deutlich (vgl. auch Tab. 1): Bei einer Brustkrebsprävalenz von 1 %, einer Richtig-Positiv-Rate der Mammographie von 80 % und einer Falsch-Positiv-Rate von 10 % überschätzen viele Menschen die Wahrscheinlichkeit deutlich, dass eine Frau mit positivem Testergebnis auch wirklich erkrankt ist. Tatsächlich steigt die Wahrscheinlichkeit für Brustkrebs, wenn eine Frau positiv getestet wird (positiver prädiktiver Wert ppV), bei dieser Basisrate nur auf etwa $ppV = 7,5\%$, bleibt aber weit entfernt von den häufig fälschlich vermuteten 80–90 % (Gigerenzer und Hoffrage 1995). Dieses Beispiel illustriert dramatische Abweichungen menschlicher Einschätzungen von tatsächlichen Risiken.

Diese und ähnliche Befunde legten zunächst nahe, dass der menschliche Verstand nicht dafür gemacht ist, probabilistische Regeln zu befolgen (Gilovich et al. 2002). Allerdings wurden diese Studien aufgrund ihrer methodischen Ansätze vielfach kritisch hinterfragt. Kritiker wie Hertwig und Erev (2009) betonten, dass Laboraufgaben oft abstrakt und realitätsfern gestaltet waren. In erfahrungsbasierten Paradigmen (experience-based paradigms), bei denen Teilnehmende Wahrscheinlichkeiten aus Beobachtungen und wiederholten Ereignissen lernen, schnitten Menschen deutlich besser ab (Schulze und Hertwig 2021). Diese Ergebnisse verdeutlichen zum einen, dass neben dem Kontext auch die Art des „Erlebens“ von Informationen eine Schlüsselrolle bei probabilistischen Schlussfolgerungen spielt (Fiedler und von Sydow 2015; Kohler 1996). Da nur Häufigkeiten, aber keine Wahrscheinlichkeiten direkt erlebt werden können, weisen diese Resultate zum anderen bereits auf die intuitivere Verarbeitung von Häufigkeiten hin.

Die empirische Forschung zeigt in der Tat, dass sich Bayesianische Probleme leichter lösen lassen, wenn Wahrscheinlichkeits- oder Prozentangaben in natürliche Häufigkeiten überführt werden (vgl. Abschn. 2.2.1). Oftmals werden zur weiteren Unterstützung auch Visualisierungen wie etwa Vierfeldertafeln oder Baumdiagramme genutzt (vgl. Abschn. 2.2.2).

2.2.1 Bayesianisches Denken und natürliche Häufigkeiten

Gigerenzer und Hoffrage (1995) konnten zeigen, dass die Art der Informationspräsentation erheblichen Einfluss auf die Fähigkeit zum Bayesianischen Schlussfolgern hat. So kann bereits die Darstellung statistischer Informationen in Form sogenannter *natürlicher Häufigkeiten* (vgl. Tab. 1, rechts) anstelle abstrakter Wahrscheinlichkeiten oder Prozentangaben Fehler signifikant reduzieren. Unter natürlichen Häufigkeiten versteht man in der Mathematikdidaktik ein geordnetes Paar natürlicher Zahlen in der Schreibweise *a von b* (mit $a \leq b$; Krauss et al. 2020). Durch die Übersetzung von Prozentangaben (wie z.B. 80%) in natürliche Häufigkeiten (z.B. 8 von 10) können Fehleinschätzungen minimiert werden, da konkrete Anzahlen leichter vorstellbar sind als abstrakte Prozente und darüber hinaus die Rechenanforderungen reduziert werden können. Eichler und Vogel (2015) argumentieren folgerichtig, dass die Förderung von Risikokompetenz und probabilistischem Denken eng mit der Präsentation von statistischen Informationen verknüpft ist.

2.2.2 Bayesianisches Denken und Visualisierungen

Neben natürlichen Häufigkeiten sind Visualisierungen ein weiteres Schlüsselement, um abstrakte Wahrscheinlichkeiten greifbar zu machen. Vor allem unter Einsatz von natürlichen Häufigkeiten, die sich auch direkt in Visualisierungen eintragen lassen (vgl. Tab. 1, unten rechts), machen Visualisierungen die für die Bayes'sche Aktualisierung relevanten Mengenrelationen sichtbar (Böcherer-Linder et al. 2018). Trainingsansätze setzen gezielt auf diese Strategien, indem sie Lernende durch den Einsatz von natürlichen Häufigkeiten und Visualisierungen beim Aufbau Bayesianischen Denkens unterstützen (Büchter et al. 2022b; Steib et al. 2025).

Neben den schulüblichen Visualisierungen wie Baumdiagrammen und Vierfeldertafeln stehen für probabilistische Situationen mit zwei dichotomen Merkmalen prinzipiell noch weitere Visualisierungen zur Verfügung (z.B. Till 2014; Spiegelhalter et al. 2011; Binder et al. 2015). Ein systematisches Literaturreview von Cui et al. (2023) zeigt, dass unterschiedliche Visualisierungstypen, darunter Tabellen, Baumdiagramme und Icon-Arrays, differentielle Effekte auf das Verständnis haben, wobei insbesondere auf natürlichen Häufigkeiten basierende Visualisierungen Bayesianisches Denken unterstützen. Khan et al. (2018) kategorisieren diese Visualisierungen in drei Hauptgruppen: nested style, frequency style und branch style. Visualisierungen des nested style wie etwa Vierfeldertafeln oder Einheitsquadrate organisieren Informationen in verschachtelten, räumlich angeordneten Strukturen, um Relationen von Wahrscheinlichkeiten darzustellen. Beim frequency style handelt es sich um Darstellungen wie zum Beispiel icon arrays, die Häufigkeiten durch Symbole repräsentieren, wobei jedes Symbol eine einzelne Einheit oder einen Fall repräsentiert. Visualisierungen des branch style wie etwa Baum-, Doppelbaum- oder Netzdiagramme (Binder et al. 2023) zeigen Häufigkeiten oder Wahrscheinlichkeiten in verzweigten Strukturen, die Ereignisfolgen und deren Wahrscheinlichkeiten bzw. Anteile hierarchisch darstellen.

Das Lösen Bayesianischer Aufgaben wird in unterschiedlicher Weise von der Vierfeldertafel und vom Baumdiagramm, welche die weithin etabliertesten Visua-

lisierungen im Unterricht darstellen, unterstützt (Binder et al. 2020). Die Vierfeldertafel (nested style) präsentiert Rand- und Schnitthäufigkeiten in einem gemeinsamen Rechteckraster, wodurch die Teil-Ganzes-Relationen sichtbar sind. Studien mit Sekundarstufenschüler:innen zeigen, dass eine Vierfeldertafel in Kombination mit natürlichen Häufigkeiten die Lösungsrate gegenüber einer reinen Textdarstellung signifikant steigert (Binder et al. 2015). Nach dem Spatial-Contiguity-Prinzip (Mayer 2009) reduziert die räumliche Nähe der statistischen Informationen die sachfremde kognitive Belastung, also jene Belastung, die nicht durch den Lerngegenstand selbst entsteht, weil die inhaltlich relevanten Angaben (z.B. erkrankt und Test positiv) unmittelbar verknüpft präsentiert werden. Auf diese Weise werden die verschachtelten Mengen (sog. nested sets) in einer Bayesianischen Aufgabe grafisch transparent dargestellt. Ein Nachteil der Vierfeldertafel liegt darin, dass bedingte Wahrscheinlichkeiten immer nur durch die Division der Zahlen zweier Zellen erhalten werden können (was mit einer Erhöhung der kognitiven Belastung einhergeht).

Das Baumdiagramm (branch style) entfaltet die Stichprobe sequenziell in Pfaden und spiegelt damit einen Prozess (z.B. erst Ereignis A, dann Test B) ikonisch wider. Das sequenzielle Charakteristikum des Baumdiagramms kann bei Aufgaben unterstützen, in denen Pfadwahrscheinlichkeiten schrittweise nachvollzogen werden sollen. Allerdings gibt es hier auch einen Nachteil, da die für die Lösung nötigen Vergleichsgrößen (etwa die Anzahl aller positiv Getesteten bei der oben genannten Mammographie-Aufgabe) im Baumdiagramm über mehrere Pfade verteilt und nicht als nested sets wie bei der Vierfeldertafel abgebildet sind. Wahrscheinlichkeiten von Schnitten müssen erst aktiv identifiziert und mental integriert werden, was zu einer erhöhten kognitiven Belastung gemäß dem Split-Attention-Effekt führt (Mayer 2009). Zudem argumentieren Büchter et al. (2022b), dass redundante Textelemente in Visualisierungen zu Bayesianischen Aufgaben bei Vierfeldertafeln leichter zu vermeiden sind als bei Baumdiagrammen, bei denen an verschiedenen Zweigen teils wiederholte Beschriftungen (z.B. positiv oder negativ) zu finden sind. Eine Vermeidung solcher Wiederholungen wirkt sich nach dem Redundanzprinzip (Mayer 2009) positiv aus.

Zusammenfassend ergibt sich aus theoretischer Sicht, dass die Vierfeldertafel (im Vergleich zum Baumdiagramm) eine transparente Darstellung eingebetteter Mengen ermöglicht (nested sets) sowie die kognitive Belastung durch die räumliche Nähe bei der Darstellung entscheidungsrelevanter Informationen minimiert. Das Baumdiagramm wiederum kann im Wahrscheinlichkeitsformat die in Bayesianischen Aufgaben typischerweise gegebenen bedingten Wahrscheinlichkeiten direkt wiedergeben. Für die vorliegende Studie wurde durch die Auswahl der beiden schulüblichen Visualisierungen jeweils *eine* Visualisierung aus der nested style- (Vierfeldertafel) und eine aus der branch style-Gruppe (Baumdiagramm) gewählt. Um Präferenzen bereits von Unterstufenschülerinnen und -Schülern zu erfassen, wurden beide Visualisierungen im Häufigkeitsformat implementiert.

2.2.3 Präferenzen bei Darstellungen Bayesianischer Aufgaben

Im Gegensatz zur Performanz auf der Basis von Visualisierungen, die traditionell häufig anhand von Maßen wie Korrektheit von Lösungen und ggfs. Bearbeitungs-

zeit untersucht wird, stehen subjektive Einschätzungen und Präferenzen von Nutzer:innen gegenüber Visualisierungen bislang deutlich seltener im Fokus empirischer Untersuchungen (Saket et al. 2016; Figueiras 2018). Der Begriff der Präferenz findet in unterschiedlichen Disziplinen wie z.B. Psychologie, Politikwissenschaften, Medizin oder Umwelt- und Nachhaltigkeitsforschung sowie im Bereich der Entscheidungstheorie, des Konsumentenverhaltens und der Marktforschung unterschiedlich nuancierte Verwendungen. In Pädagogik und Bildungswissenschaften sind Präferenzen von Schüler:innen von Interesse, da Unterricht auf der Ebene makro- wie mikroadaptiven Lehrkrafthandelns an Lernvoraussetzungen, Bedürfnisse und Interessen der Lernenden anschließen sollte, z.B. durch unterschiedliche Repräsentationen, Methoden oder Aufgabenformate. Die zugrundeliegende Förderannahme ist dabei, dass Lernmaterialien und Aufgabenformate, die Wahlmöglichkeiten eröffnen und an Interessen bzw. subjektive Präferenzen der Lernenden anschließen, Motivation und unter bestimmten Bedingungen auch Lernleistung unterstützen können (Cordova und Lepper 1996; Hardy et al. 2019; Patall et al. 2008).

Mit Rückgriff auf Lumer (2022) lässt sich dabei Präferenz grundlegend definieren als eine „subjektive Einstellung der Art, dass ein Subjekt s einen Sachverhalt p einem Sachverhalt q vorzieht, p also als besser ansieht.“ (S. 178). Das subjektive Moment bedingt, dass Präferenzen individuelle Einstellungsdimensionen zwischen den vorhandenen zwei Alternativen zu einem bestimmten Zeitpunkt widerspiegeln, also im Unterschied zu erworbenen Einstellungen einem Sachverhalt gegenüber keine zeitlich überdauernde Eigenschaft sind (Steiner 2007). Eine Präferenz ist demnach als dynamische Größe zu betrachten, die das Ergebnis wiederholter Prozesse der Wahrnehmung und Entscheidung darstellt und somit wandelbar ist (Hahn 1997). Präferenzen sind situativ und dynamisch, weil sie je nach Aufgabenstellung, verfügbaren Informationen und individuellen Erfahrungen variieren können. Diese Abhängigkeit ist für die Bildungsforschung relevant, da sie zeigt, dass Präferenzen nicht als stabile Dispositionen verstanden werden sollten, sondern als momentane Auswahlentscheidung innerhalb einer spezifischen Lernsituation. Empirische Studien zeigen zudem, dass Darstellungspräferenzen von Lernenden durch Aufgabenmerkmale und Vorwissen beeinflusst werden und sich im Lernprozess verändern können (Acevedo Nistal et al. 2012).

Die Präferenz in Bezug auf Visualisierungen ist für schulische Belange von Bedeutung, wenn beispielsweise im Stochastikunterricht bereits in der Unter- oder Mittelstufe Kontexte und Denkweisen angebahnt werden sollen, deren formale Behandlung erst in der Oberstufe und mit dem Kalkül bedingter Wahrscheinlichkeiten möglich ist (mithilfe natürlicher Häufigkeiten ist eine Behandlung Bayesianischer Denkweisen prinzipiell bereits im Grundschulunterricht möglich; Zhu und Gigerenzer 2006). Vor diesem Hintergrund ist es interessant, welche Visualisierungen Lernende für einen solchen „propädeutischen Erstkontakt“ intuitiv bevorzugen und wie sich solche Präferenzen zwischen Jahrgangsstufen unterscheiden. Es gibt bereits Hinweise darauf, dass Schüler:innen beim Umgang mit bedingten Wahrscheinlichkeiten das Baumdiagramm bevorzugen, gefolgt von der Vierfeldertafel (Rach 2018; das bislang noch wenig bekannte Einheitsquadrat fand in dieser Studie keinen vergleichbaren Anklang).

Die Bestimmung der Präferenzen von Personen ist methodisch anspruchsvoll. Viele Studien, die sich mit Präferenzen beschäftigen, stützen sich auf Notizen und Selbstauskünfte der Proband:innen. Acevedo Nistal et al. (2012) sehen hierbei eine methodische Grenze: Die valide Bestimmung anhand eigener Angaben, welche Visualisierung – z. B. aus einer vorgegebenen Auswahl – auch tatsächlich genutzt wird, ist aufgrund teilweise auch unbewusst ablaufender kognitiver Prozesse nur bedingt aussagekräftig. Eyetracking bietet die Möglichkeit, Blickbewegungen aufzunehmen und dadurch Präferenzen objektiv zu quantifizieren. Eine Studie von Ackermann und Domhardt (2015) zeigte, dass sowohl die Fixationsdauer als auch die Anzahl der Fixationen die verbal geäußerte Präferenz der Proband:innen bestätigen können. Eine Untersuchung von Malone et al. (2020) in einem (nicht-Bayesianischen) mathematischen Problemkontext, in dem es um die Auswahl von Visualisierungen ging (inkl. einer reinen Textversion), ergab, dass bei der parallelen Präsentation von textuellen Informationen sowie Visualisierungen eine längere Fixationsdauer für die Visualisierungen als für den Text ermittelt wurden. Im Hinblick auf Bayesianische Situationen untersuchten Bruckmaier et al. (2019) bereits die Blickbewegungen der Proband:innen mittels Eyetracking, jedoch nicht mit dem Fokus auf Präferenzen, sondern um fehlerhafte Lösungsstrategien bei jeweils gegebener Visualisierung zu identifizieren.

Im vorliegenden Beitrag gehen wir nun einen Schritt weiter und erfassen die Präferenz der Lernenden in einem mixed-method-Ansatz: Zur Analyse werden Eyetracking-Daten (Fixationsdauer, Heatmaps) als quantitative Maße genutzt, diese aber anschließend mit qualitativen Daten aus Interviews abgeglichen, um die Blickmuster mit den subjektiven Aussagen und Begründungen der Schüler:innen zu validieren. Somit werden erstmals zum Thema Bayesianische Aufgaben Eyetracking-Daten methodisch durch Kurzinterviews ergänzt und auf dieser Basis Präferenzen bezüglich Visualisierungen untersucht.

3 Forschungsfragen

Bisher ist wenig darüber bekannt, welche Präferenzen Schüler:innen bei Visualisierungen Bayesianischer Aufgaben haben und wie sich diese auf die Bearbeitung entsprechender Aufgaben auswirken. Im Folgenden sollen dabei nur noch Häufigkeitsversionen betrachtet werden (Tab. 1, rechts). Konkret ist das Ziel der vorliegenden Studie, die Präferenzen von Schüler:innen der 5./6. und der 10. Klassenstufe zwischen Baumdiagramm und Vierfeldertafel beim Lösen von Bayesianischen Aufgaben, aber auch eine mögliche Bevorzugung dieser Visualisierungen gegenüber einer rein textuellen Darstellung der statistischen Informationen zu untersuchen. Diese Präferenzen sollen anschließend mit den zugehörigen Performanzen in Beziehung gesetzt und dabei auch Bezüge zur Klassenstufe analysiert werden. Die Wahl zweier verschiedener Altersgruppen ermöglicht die Betrachtung der beiden Fälle, dass die Visualisierungen noch nicht (5./6. Klasse) bzw. schon bekannt (10. Klasse) sind (Hessisches Kultusministerium 2010).

Folgende Forschungsfragen (F) werden untersucht:

1. Präferenz

F1 Welche Darstellungsform präferieren Schüler:innen der Sekundarstufe beim Lösen Bayesianischer Aufgaben, wenn die zur Bearbeitung der Aufgabe erforderlichen statistischen Informationen jeweils in zwei von drei Darstellungsformen (Baumdiagramm, Vierfeldertafel, reiner Text) parallel zur Verfügung gestellt werden?

2. Präferenz und Klassenstufe

F2 Hängt die Präferenz von der Klassenstufe (5./6. vs. 10. Klassenstufe) ab?

3. Präferenz und Performanz

F3 Wie unterscheiden sich die Lösungsraten in Abhängigkeit von der präferierten Darstellung (Baumdiagramm, Vierfeldertafel, reiner Text) und welche Interaktion gibt es dabei mit der Klassenstufe?

4 Methode

4.1 Stichprobe

Die Erhebung der Daten erfolgte im Herbst 2024 mit insgesamt $N = 36$ Schüler:innen einer Gesamtschule in Hessen, wovon $N = 18$ Schüler:innen aus der 5. und 6. Klasse (Alter: $M(SD) = 11,78(0,85)$ Jahre; 50 % weiblich) und $N = 18$ Schüler:innen aus der 10. Klasse (Alter: $M(SD) = 16,00(0,76)$ Jahre; 72 % weiblich) waren. Da in Hessen vor allem das Baumdiagramm als Visualisierung (bedingter) Wahrscheinlichkeiten in der 8. Klassenstufe unterrichtet wird (Hessisches Kultusministerium 2010), verfügen nur die 10.-Klässler:innen aus ihrem Mathematikunterricht über Vorwissen zum Baumdiagramm, teilweise auch zur Vierfeldertafel. Die Schüler:innen erhielten keine Entlohnung für die Teilnahme an der Studie, die innerhalb der Unterrichtszeit und mit Zustimmung der Eltern durchgeführt wurde.

4.2 Stimuli

Für die Erhebung am Eyetracker wurden insgesamt 15 Bayesianische Aufgaben konzipiert, wovon drei als Beispielaufgaben und 12 als Testaufgaben fungierten. Für jede Aufgabe wurden die zur Lösung erforderlichen statistischen Informationen im Format natürlicher Häufigkeiten in je zwei (aus insgesamt drei) verschiedenen Darstellungsformen – Vierfeldertafel, Baumdiagramm und Text – simultan präsen-

Frage: Wie viele Brillentragende Teilnehmerinnen einer Studie haben braune Haare?			
Antwort: _____ von _____ Brillentragenden haben braune Haare.			
60 von 100 Teilnehmerinnen an einer Studie haben braune Haare.			
20 der 60 Teilnehmerinnen mit braunen Haaren tragen eine Brille.			
15 der 40 Teilnehmerinnen mit anderer Haarfarbe tragen eine Brille.			
100 Teilnehmerinnen	mit Brille	ohne Brille	
braune Haare	20	...	60
andere Haarfarbe	15	...	40

Abb. 1 Beispielaufgabe mit zwei der insgesamt drei Darstellungsformen (Vierfeldertafel vs. Text)

tiert. Diese drei Darstellungsformen wurden informationsäquivalent präsentiert, d. h., sie beinhalten dieselben statistischen Informationen (konkret: dieselben Zahlen). Man beachte, dass informationsäquivalente Darstellungen nur im Häufigkeitsformat möglich sind, da Vierfeldertafeln im Wahrscheinlichkeitsformat Schnittwahrscheinlichkeiten (und keine bedingten Wahrscheinlichkeiten) enthalten. Aufgrund der Informationsäquivalenz ergeben sich sowohl beim Baumdiagramm als auch bei der Vierfeldertafel nicht ausgefüllte Felder, welche hier durch drei Punkte gekennzeichnet wurden (z.B. ... Teilnehmerinnen ohne Brille mit braunen Haaren; vgl. Abb. 1). Ein weiterer Vorteil der ausschließlichen Verwendung des Formats (natürlicher) Häufigkeiten (z.B. 50 der 200 Kinder, 300 von 500 Kindern) ist, dass sich bei Stimuli im Häufigkeitsformat ausreichend Varianz in den Antworten erwarten lässt (wohingegen Wahrscheinlichkeitsversionen in der Regel nur von weniger als 10 % der Versuchspersonen gelöst werden).

Die Beispielaufgabe in Abb. 1 illustriert neben der räumlichen Anordnung der Fragestellung (jeweils oben) und den darunter nebeneinander abgebildeten zwei Darstellungsformen (hier: Vierfeldertafel bzw. Text) insbesondere, dass die Teilnehmenden zur Lösung der jeweiligen Aufgabe beliebig zwischen den beiden Darstellungsformen wechseln, ggfs. aber auch nur eine davon heranziehen konnten.

Die drei verschiedenen Kombinationen von Darstellungen (Baumdiagramm vs. Vierfeldertafel; Baumdiagramm vs. Text; Vierfeldertafel vs. Text) wurden systematisch mit einer Variation der Positionierung auf den beiden Bildschirmhälften gekreuzt (d.h. Darstellung links und rechts wurde jeweils noch vertauscht).

Die Aufgabenkontexte (z. B. Medizin, Medien, Beruf) waren schüler:innengerecht gewählt und in der Abfolge ebenso systematisch variiert. Da der Fokus der Studie auf der Wahl der Darstellungsform, nicht jedoch auf den Rechenfertigkeiten der Schüler:innen lag, wurden Zahlen lediglich aus dem Zahlenraum bis 10.000 verwendet.

4.3 Durchführung

Für die Eyetracking-Erhebung wurden die Teilnehmenden einzeln in einem separaten Raum ihrer Schule befragt. Zu Beginn wurde den Schüler:innen ein kurzer Überblick über den Ablauf der Erhebung gegeben. Die Aufnahmen erfolgten mit einem Tobii-Eyetracker (Modell: Pro Fusion) mit einer Frequenz von 250 Hz, wobei die Stimuli auf einem 17-Zoll-Monitor im Abstand von ca. 65 cm von den Teilnehmenden präsentiert wurden. Nach der Aufklärung über das Ziel der Studie erfolgte für jede Versuchsperson die Kalibrierung am Eyetracker mittels einer 9-Punkt-Kalibrierung.

Zu Beginn der eigentlichen Erhebung am Eyetracker wurde den Schüler:innen der Aufbau von Vierfeldertafel und Baumdiagramm anhand standardisierter Folien vermittelt; die Darstellungsform Text wurde, da sie allen Schüler:innen ohnehin bekannt war, nicht gesondert thematisiert. Dadurch waren alle Schüler:innen mit beiden Visualisierungen im Grundsatz vertraut, selbst wenn sie eine oder beide Visualisierungen im Unterricht noch nicht kennengelernt haben sollten. Vor der Bearbeitung der (Beispiel-)Aufgaben wurden die Schüler:innen explizit auf die Informationsäquivalenz der beiden jeweils gegebenen Darstellungsformen hingewiesen. Zu den drei Beispielaufgaben konnten die Schüler:innen dann auch die Lösung einsehen, um den folgenden Ablauf und das erwartete Format der Antworten zu illustrieren. Anschließend wurden ihnen 12 Aufgabenstellungen zu 12 Kontexten präsentiert (vgl. Abb. 1), wobei diese den Teilnehmenden zur Vermeidung von systematischen Reihenfolgeeffekten zufällig in einer von drei ausbalancierten Versionen – hinsichtlich Reihenfolge, Kombination und Positionierung der drei Darstellungsformen (links vs. rechts) – gezeigt wurden. Die Schüler:innen gaben ihre Lösungen jeweils an-

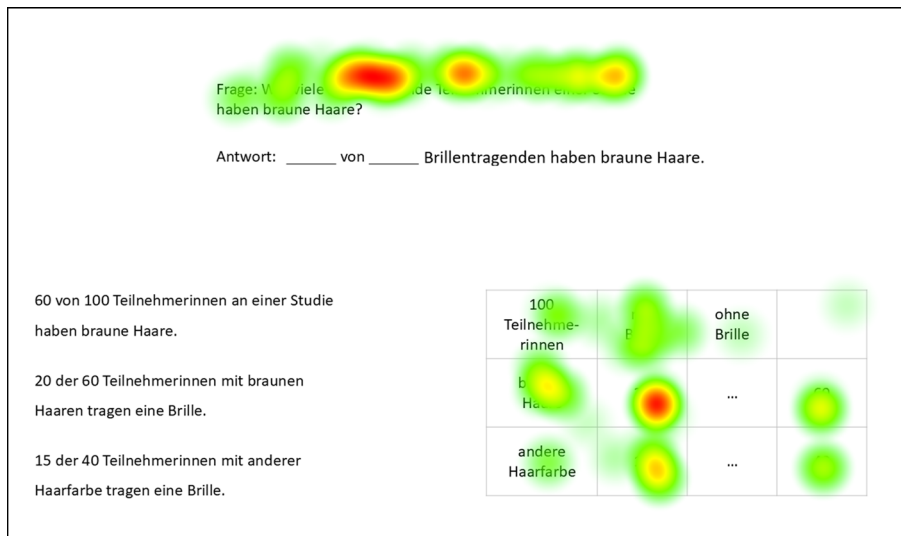


Abb. 2 Exemplarische Heatmap eines/r Schüler:in bei einer der 12 Aufgaben. (In der gedruckten Schwarz-Weiß-Fassung gilt: Je dunkler eine Stelle, desto länger die Fixationszeit. Eine farbige Ansicht ist in der online-Version dieses Artikels zu finden.)

schließend in mündlicher Form an. Die Teilnehmenden konnten die Geschwindigkeit beim Lösen der Aufgaben selbst bestimmen, indem sie per Mausklick zur nächsten Aufgabe wechselten.

Im Anschluss an die Bearbeitung der Aufgaben wurden die Teilnehmenden in einem per Tonaufnahme festgehaltenen Kurzinterview zunächst nach ihren allgemeinen Präferenzen bezüglich der Darstellungsformen und den entsprechenden Gründen befragt. Dabei wurde explizit die Angabe einer Reihenfolge der drei Darstellungsformen nach ihrer Präferenz eingefordert (jede Darstellungsform wurde bei den 12 Aufgaben insgesamt acht Mal gesehen). Anschließend wurden noch einige demografische Angaben (u. a. Alter, Geschlecht) ebenfalls mündlich erhoben. Zum Abschluss der Befragung, die insgesamt ca. 30 min in Anspruch nahm, kehrten die Schüler:innen in den regulären Unterricht zurück.

4.4 Operationalisierung des Begriffs Präferenz und Auswertung

Für die Auswertung der Blickverläufe wurde die Präferenz für eine Visualisierung in Anlehnung an Rach (2018) anhand zweier Indikatoren bestimmt: Zum einen diente die per Eyetracking ermittelte Wahl der Darstellungsform als Indikator, zum anderen die daran anschließende Angabe der Präferenz in den Kurzinterviews. Letztere diente vor allem zur Validierung der Daten aus den Eyetracking-Erhebungen.

Als Kriterium für die über Eyetracking bestimmte Präferenz einer Darstellung gegenüber einer anderen wurde die summierte Fixationszeit pro Darstellung als Vergleichsmaß herangezogen (vgl. Abschn. 4.5; Ackermann und Domhardt 2015). Das Ermitteln einer Präferenz aus zwei vorgegebenen Optionen, also das Treffen einer binären Entscheidung, entspricht letztlich der Methode des sog. „comparative judgement“ (Jones und Davies 2024; Jones und Inglis 2015), bei der eine Person mit zwei Repräsentationen (hier: statistische Darstellungsformen) konfrontiert wird und entscheiden soll, welche einem bestimmten Kriterium mehr entspricht (hier: welche sie – gemessen mittels Blickbewegungen – präferiert). Diese Methode ermöglicht es nicht nur, individuelle Präferenzen für bestimmte Darstellungsformen zu bestimmen, sondern darüber hinaus auch, Zusammenhänge mit Performanzen zu erkennen. Die diesbezüglichen Auswertungen erfolgten mit der Software Tobii Pro Lab.

Zur ersten visuellen Analyse der Blickverteilungen (und damit potenzieller Präferenzen) hinsichtlich der jeweils zwei parallel präsentierten Darstellungsformen wurden die Heatmaps aller 12 Aufgaben ausgewertet. Hierbei wurden sowohl individuelle Heatmaps für jede Aufgabe und jede:n Schüler:in (vgl. exemplarisch Abb. 2) als auch aggregierte Heatmaps erstellt – letztere sowohl über alle Teilnehmenden hinweg als auch differenziert nach Klassenstufen.

Zur quantitativen Analyse der Fixationszeiten kam der I-VT-Algorithmus (Velocity-Threshold Identification; Salvucci und Goldberg 2000) mit dem Tobii I-VT Fixationsfilter (Tobii Pro AB 2014) zum Einsatz. Augenbewegungen mit unter 30°/s wurden als Fixationen klassifiziert, schnellere Bewegungen als Sakkaden. Für die Analyse wurden die Fixationen auf zuvor definierten AOIs (Area Of Interest) um die beiden Darstellungsformen untersucht, um die Betrachtungsdauer pro Darstellungsform zu ermitteln und Vergleiche hinsichtlich der Präferenz zwischen den je zwei Darstellungsformen pro Aufgabe zu ermöglichen.

Entscheidend für die Bestimmung einer individuellen Präferenz in Bezug auf eine Aufgabe war das Verhältnis der jeweils summierten Fixationszeiten auf den beiden Darstellungsformen dieser Aufgabe. Ein Verhältnis der Fixationsdauer von beispielsweise 17 zu 3 s ergibt ein prozentuales Präferenzverhältnis von 85:15. Ab welchem Anteil der Fixationsdauer eine Darstellungsform als präferiert gegenüber der anderen Darstellungsform angesehen wird (z. B. ab >50 % oder >70 % oder >90 %), ist aufgrund der dürftigen Datenlage bezüglich der Eyetracking-Methode zur Untersuchung von Präferenzen zu einem gewissen Grad arbiträr. Als Entscheidungskriterium wurde festgelegt, dass eine Darstellungsform dann gegenüber der anderen, parallel dargebotenen Darstellungsform präferiert wird, wenn sie mindestens zu 70 % (d. h. Verhältnis mind. 70:30) betrachtet wurde. Dieses Kriterium wurde bereits in einer der Hauptstudie vorangehenden Pilotstudie (vgl. Abschn. 4.5) erprobt und konnte durch die erhobenen Interviewdaten validiert werden. Wie vergleichende Analysen unserer Daten mit alternativen Präferenzkriterien (bspw. 60:40 und 80:20) zudem zeigten, führen diese nicht zu substantiell anderen Ergebnissen. Darüber hinaus ist das von uns gewählte Kriterium strenger als in vergleichbaren Studien anderer Disziplinen (Ackermann und Domhardt 2015).

Unter Einbezug aller 12 pro Person bearbeiteter Aufgaben wurde anhand der Häufigkeit der Präferenzen unter diesen Aufgaben eine intraindividuelle Präferenz über alle Aufgaben ermittelt. Sobald eine der drei Darstellungsformen – jede kam in jeweils 8 von 12 Aufgaben vor – mindestens zwei Mal öfter als jede der anderen beiden Darstellungen präferiert wurde (z. B. 7:4:1 oder 6:4:2), wurde sie als die von dieser Person präferierte Darstellung betrachtet. Dieser gewählte Abstand (von mindestens zwei Präferenzen) erlaubt die pragmatische Unterscheidung zwischen relativ klaren Präferenzen und annähernd gleichen Präferenzen (von zwei Darstellungen); auch konservativere oder liberalere Kriterien wären denkbar gewesen, entsprechende vergleichende Analysen führen aber ebenfalls nicht zu substantiell abweichenden Ergebnissen. Auch dieses Kriterium wurde in der Pilotstudie bereits angewandt und durch die Interviewdaten validiert (siehe Abschn. 4.5). Bei einem geringeren Abstand (nur maximal einmal öfter) wurden die beiden am häufigsten präferierten als gleichermaßen präferierte Darstellungsformen gewertet (z. B. 6:5:1 oder 5:5:2). Wenn alle drei Darstellungsformen nahezu gleich häufig präferiert wurden (z. B. 5:4:3 oder 4:4:4), wurde dies als keine Präferenz zugunsten einer Darstellung gewertet.

Neben deskriptiven Ergebnissen (mittlere relative und absolute Lösungsraten inkl. Standardabweichungen: M , SD) werden trotz der kleinen Stichprobenzahl auch inferenzstatistische Analysen der Eyetrackingdaten durchgeführt, die nur als ergänzende Information verstanden werden sollen (vgl. Abschn. 5.3). Hierbei werden neben Varianzanalysen inklusive Post-Hoc-Tests auch Effektstärken d für Mittelwertsunterschiede berechnet. Dabei handelt es sich gemäß Cohen (1992) bei $d=0,2$ um kleine, bei $d=0,5$ um mittlere und bei $d=0,8$ um große Effekte.

Der zweite Indikator ist die von den Schüler:innen in den Interviews angegebene Präferenzreihenfolge der drei Darstellungsformen. Die als am hilfreichsten eingeschätzte Darstellungsform wurde als die präferierte Darstellungsform gewertet. In aller Regel gaben die Schüler:innen eine eindeutige, unmissverständliche Antwort (die zudem audiografiert wurde). Im Fall uneindeutiger Antworten wurde

vom Interviewer eine Rückfrage dazu gestellt, welche der drei bzw. welche beiden Darstellungen sie insgesamt präferieren. Daraufhin ergaben sich immer klare Antworten.

Zur Bestimmung der intraindividuellen Präferenz wurde die Passung zwischen der Präferenz, die anhand der Fixationsdauer bestimmt wurde, und der Präferenz, die im Interview genannt wurde, überprüft. Dabei wurde als eindeutige Präferenz gezählt, wenn entweder beide Präferenzen (Fixationsdauer und Interview) eine eindeutige Übereinstimmung ergaben oder wenn eine der beiden Präferenzen sich auf zwei Darstellungen bezog, die andere jedoch eindeutig war (z.B. Fixationsdauer uneindeutig zwischen Baumdiagramm und Vierfeldertafel, Interview hingegen eindeutig für Baumdiagramm: Dies ergab eine Präferenz für das Baumdiagramm). So konnte auch bei vielen Teilnehmenden mit zwei präferierten Darstellungen bei einem der beiden Kriterien (Fixationsdauer oder Interview) eine finale Entscheidung für eine Präferenz getroffen werden.

4.5 Pilotierung

Vor der Durchführung der Studie wurde eine Pilotierung mit $N=30$ Proband:innen (Studierende, Promovierende und Mitarbeiter:innen) an der Universität Regensburg und der Pädagogischen Hochschule Heidelberg durchgeführt (Alter: $M(SD)=28,67(10,12)$ Jahre; 61 % weiblich). Die Stichprobe war heterogen gewählt – vor allem in Hinblick auf den mathematischen Hintergrund der Teilnehmenden –, um möglichst differenzierte Hinweise für die darauffolgende Hauptstudie mit Schüler:innen zweier Altersstufen zu erhalten.

Neben der technischen Funktionalität des Eyetracking-Verfahrens zeigte sich in der Pilotierung bereits eine deutliche Präferenz zugunsten der beiden Visualisierungen (Baumdiagramm oder Vierfeldertafel) gegenüber dem Text. Die intraindividuellen Präferenzen in der Pilotierungsstudie sind in Tab. 2 sowohl anhand des quantifizierbaren Verhältnisses der kumulierten Fixationszeiten der Darstellungsformen als auch anhand der anschließenden Interviews operationalisiert (vgl. Abschn. 4.4). Insgesamt ließ sich (ausschließlich) anhand der Fixationsdauer bei 22 der 30 Teilnehmenden der Pilotierung und (ausschließlich) anhand des folgenden Interviews sogar bei 29 Teilnehmenden eine eindeutige Präferenz zugunsten einer bestimmten Darstellungsform feststellen.

Schließlich wurde noch die Bestimmung einer intraindividuellen Präferenz durch die Kombination von Eyetracking und Interview erprobt. Dabei wurde – wie dann auch in der Hauptstudie – eine Präferenz als eindeutig festgelegt, wenn entweder beide Präferenzen (Fixationsdauer und Interview) eine eindeutige Übereinstimmung ergaben oder wenn eine Präferenz sich auf zwei Darstellungen bezog, die andere jedoch eindeutig war (z.B. Fixationsdauer vergleichbar für Baumdiagramm und Vierfeldertafel, Interview hingegen eindeutig für Baumdiagramm ergab: Präferenz für Baumdiagramm). So konnte auch bei einigen Teilnehmenden mit zwei präferierten Darstellungen bei einem der beiden Kriterien (Fixationsdauer oder Interview) die Entscheidung für eine Präferenz getroffen werden. Insgesamt konnte so bei 27 der 30 Teilnehmenden die intraindividuelle Präferenz festgelegt werden.

Tab. 2 Präferenzen der Teilnehmenden in der Pilotierungsstudie anhand der Fixationsdauer, Interviews und deren Übereinstimmung

	Präferenz für eine Darstellungsform				Zwei präferierte Darstellungsformen				Keine Präferenz	
	BD*	VFT**	Text	Gesamt	BD* und Text	BD* und VFT**	VFT** und Text	Gesamt	Gesamt	Gesamt
Fixationsdauer	11	9	1	22	0	8	0	8	1	1
Interviews	13	16	0	29	0	1	0	1	0	0
Intraindividuelle Präferenz (aus Fixationsdauer und Interviews)	12	15	0	27	0	0	0	0	0	3

* *BD* Baumdiagramm** *VFT* Vierfeldertafel

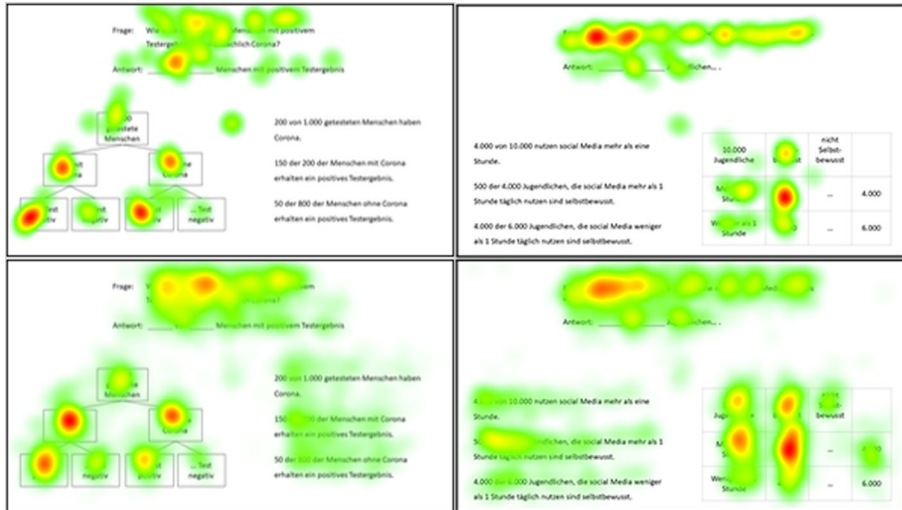


Abb. 3 Exemplarische Heatmaps zweier Schüler:innen zu zwei (verschiedenen) Aufgaben (Individual-ebene, *oben*) und exemplarische, über alle Schüler:innen aggregierte Heatmaps zu denselben zwei (verschiedenen) Aufgaben (Aufgabenebene, *unten*)

Die Übereinstimmung zwischen den beiden Kriterien „erste Präferenz anhand der Fixationsdauer“ und „erste Präferenz anhand der Interviewdaten“ kann mit einem Wert für Cohens Kappa von $\kappa=0,81$ als sehr gut eingestuft werden (Hinweis: Es handelt sich nicht um eine Interrater-Reliabilität, sondern eine Messmethoden-Übereinstimmung).

Insgesamt konnte in der Pilotierungsstudie ein klarer Zusammenhang zwischen der beim Eyetracking über die Fixationsdauer gemessenen und den im Kurzinterview angegebenen Präferenzen festgestellt werden, weshalb dieses Verfahren dann auch für die Hauptstudie übernommen wurde (vgl. Abschn. 4.4).

5 Ergebnisse

5.1 Forschungsfrage 1: Präferenz

Zur Beantwortung der ersten Forschungsfrage, welche der drei Darstellungsformen von den Schüler:innen zur Lösung der Bayesianischen Aufgaben präferiert wird, wurden im Folgenden einerseits Eyetracking-Daten (5.1.1) und andererseits Interviewaussagen ausgewertet (5.1.2), die abschließend beide integriert betrachtet wurden (5.1.3).

5.1.1 Analyse mittels Eyetracking

Bereits die Heatmaps einzelner Schüler:innen zeigen bei zwei Dritteln der individuell bearbeiteten Aufgaben (36 Schüler:innen · 12 Aufgaben = 432 Bearbeitungen)

Tab. 3 Präferenzen der Schüler:innen anhand der Fixationsdauer (Individualebene)

Präferenz für eine Darstellungsform				Zwei präferierte Darstellungsformen				Keine Präferenz
BD*	VFT**	Text	Gesamt	BD* und Text	BD* und VFT**	VFT** und Text	Gesamt	Gesamt
13	6	1	20	9	6	0	15	1

Für Details, wann eine bzw. zwei Darstellungsform(en) als präferiert gewertet wurde(n), vergleiche Abschnitt 4.4

* *BD* Baumdiagramm

** *VFT* Vierfeldertafel

Tab. 4 Präferenzen der Schüler:innen anhand ihrer Aussagen in den Interviews (Individualebene)

Eine eindeutig präferierte Darstellungsform				Zwei präferierte Darstellungsformen				Keine Präferenz
BD*	VFT**	Text	Gesamt	BD* und VFT**	BD* und Text	VFT** und Text	Gesamt	Gesamt
19	9	4	32	0	1	1	2	2

* *BD* Baumdiagramm

** *VFT* Vierfeldertafel

deutliche Unterschiede bezüglich der beiden jeweils gegebenen Darstellungsformen. Abb. 3 (oben) zeigt die Heatmaps zweier Schüler:innen bei zwei ausgewählten Aufgaben.

Fasst man die Heatmaps zu je einer Aufgabe über alle Schüler:innen hinweg zusammen (aggregierte Heatmaps, d. h. auf Aufgabenebene), so sind die Verhältnisse zwischen den Darstellungen in der Regel nicht mehr so eindeutig, jedoch ist auch hier eine sichtbare Tendenz zugunsten der jeweiligen Visualisierung gegenüber einer rein textuellen Präsentation der Informationen bemerkbar (vgl. Abb. 3 unten).

Im Folgenden werden die Präferenzen anhand des quantifizierbaren Verhältnisses der kumulierten Fixationszeiten der beiden Darstellungsformen operationalisiert und auf Individualebene betrachtet (vgl. Abschn. 4.4). Insgesamt lässt sich so bei 20 der 36 Schüler:innen eine eindeutige Präferenz zugunsten einer bestimmten Darstellungsform feststellen. 15 weitere Schüler:innen zeigen eine etwa gleichwertige Präferenz für zwei der drei Darstellungen, sodass zumindest eine Darstellung ausgeschlossen werden kann. In beiden Gruppen sticht dabei das Baumdiagramm prominent als die bevorzugte Darstellungsform heraus (vgl. Tab. 3).

5.1.2 Validierung mittels Interviews

Aus den Aussagen der Schüler:innen in den anschließenden Interviews ergaben sich die in Tab. 4 dargestellten (verbalisierten) Präferenzen.

Die meisten Schüler:innen gaben in den Interviews auf die Frage, welche der drei Darstellungen sie bevorzugen, wiederum das Baumdiagramm an, gefolgt von der Vierfeldertafel und schließlich dem Text; selten wurden auch zwei Darstellungen als Präferenz genannt. Auf Nachfrage nannten die Schüler:innen als Begründung

Tab. 5 Intraindividuelle Präferenz der Schüler:innen

Baumdiagramm	Vierfeldertafel	Text	Gesamt mit Präferenz	Keine Präferenz	Gesamt
17	5	2	24	12	36

für ihre Präferenz beispielsweise Argumente wie die (größere) Vertrautheit mit der entsprechenden Darstellungsform. Zudem wurde deren bessere Strukturierung (z. B. „Weil es übersichtlicher war“, „Weil das am besten von der Struktur her für mich aufgebaut war“) oder die Vollständigkeit der Informationen, die man einfach und nachvollziehbar habe ablesen können („Also ich fand das halt verständlicher“), als Argumente genannt.

5.1.3 *Passung der Eyetracking-Daten und der Aussagen in den Interviews*

Beim Vergleich von 5.1.1 und 5.1.2 zeigt sich auf Individualebene, dass bei insgesamt 19 der 36, also etwa der Hälfte der Schüler:innen die jeweils ermittelten Präferenzen übereinstimmen. Bei weiteren fünf Schüler:innen konnten anhand der Eyetracking-Daten zwei Präferenzen ermittelt werden, von denen auch eine der beiden als erste Präferenz beim Interview angegeben wurde. In neun Fällen konnte keine Übereinstimmung der beiden gemessenen Präferenzen gefunden werden und in insgesamt drei weiteren Fällen konnte keine Präferenz zugeordnet werden, da entweder beim Eyetracking oder beim Interview keine Präferenz festgelegt werden konnte.

Sowohl für die 19 Schüler:innen mit vollständiger Übereinstimmung als auch bei den fünf Schüler:innen mit zwei Präferenzen beim Eyetracking und einer Übereinstimmung im Kurzinterview wurde eine intraindividuelle Präferenz festgelegt (damit insgesamt also bei 24 Schüler:innen). Somit gaben zwei Drittel der Schüler:innen eine gleichbleibende Präferenz für eine der drei Darstellungsformen an, das verbleibende Drittel der Schüler:innen hingegen nicht (vgl. Tab. 5).

Die Übereinstimmung zwischen den beiden Kriterien „erste Präferenz anhand der Fixationsdauer“ und „erste Präferenz anhand der Interviewdaten“ kann mit einem Wert für Cohens Kappa von $\kappa = 0,40$ als moderat eingestuft werden (zur Bedeutung von Cohens Kappa vgl. Abschnitt 4.5).

5.2 **Forschungsfrage 2: Präferenzunterschiede in den Klassenstufen**

Zur Beantwortung der zweiten Forschungsfrage, ob die Präferenz zugunsten einer bestimmten Darstellungsform von der Klassenstufe (5./6. vs. 10. Klasse) abhängt, wurden im Folgenden wieder einerseits Eyetracking-Daten (5.2.1) und andererseits Interviewaussagen (5.2.2) ausgewertet, die abschließend integriert – und weiterhin auf Individualebene – betrachtet wurden (5.2.3).

Aufgrund der daraus resultierenden kleinen Stichprobe in jeder Kategorie (z. B. Klassenstufe 5/6 und gleichzeitig Präferenz Baumdiagramm) sind die Ergebnisse nur bedingt statistisch belastbar, sondern bieten eher interessante erste Einblicke.

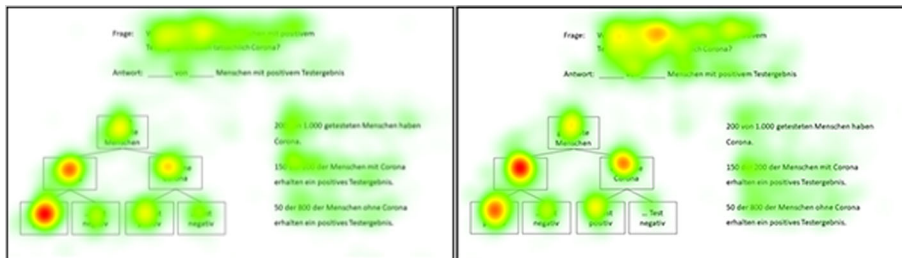


Abb. 4 Aggregierte Heatmaps über alle Schüler:innen der Klassenstufe 5/6 (*links*) und der Klassenstufe 10 (*rechts*) zur selben Aufgabe

5.2.1 Analyse mittels Eyetracking

Vergleiche zwischen (über die beiden Klassenstufen) aggregierten Heatmaps zu einzelnen Aufgaben können Gemeinsamkeiten und Unterschiede bereits gut illustrieren: Vergleicht man die aggregierten Heatmaps beider Klassenstufen zu einzelnen Aufgaben miteinander (für eine Beispielaufgabe vgl. Abb. 4), so zeigt sich visuell kaum ein Unterschied. Auf Individualebene hingegen zeigt sich, dass in Klassenstufe 5/6 deutlich häufiger beide gegebenen Darstellungen betrachtet wurden, während in Klassenstufe 10 häufiger nur eine der beiden Darstellungsformen herangezogen wurde (ohne Abbildung). Das Baumdiagramm ist dabei gefolgt von der Vierfeldertafel (vgl. auch Tab. 5) in beiden Klassenstufen die bevorzugte Visualisierung, jedoch wird in beiden Klassenstufen auch der Text immer wieder zur Ermittlung der Lösung herangezogen.

Anhand der Fixationszeiten wird der mittels Heatmaps vermittelte optische Eindruck tiefergehend quantitativ analysiert (analog zum in Abschn. 5.1 beschriebenen Vorgehen). Hierbei zeigt sich folgendes Bild (vgl. Tab. 6): Die 20 Schüler:innen mit eindeutiger Präferenz aus den Eyetracking-Daten (vgl. Tab. 3) verteilen sich differenziell auf die beiden Klassenstufen: In Klassenstufe 5/6 konnte bei nur sieben Schüler:innen eine eindeutige Präferenz festgelegt werden, in Klassenstufe 10 sind es hingegen 13 Schüler:innen. Deutlich häufiger traten in Klassenstufe 5/6 dafür zwei zugleich präferierte Darstellungsformen auf (11 vs. 4).

5.2.2 Validierung mittels Interviews

Die Auswertung der Interviews ergab, dass von den 18 Schüler:innen der 10. Klasse etwa zwei Drittel das Baumdiagramm und ein Drittel die Vierfeldertafel bevorzugten (vgl. Tab. 7); der Text wurde praktisch nie präferiert. Bei den Schüler:innen der 5./6. Klassenstufe war das Bild nicht ganz so eindeutig: Das Baumdiagramm wurde nur von etwa der Hälfte bevorzugt, gefolgt vom Text und der Vierfeldertafel. Insofern zeigte sich bei den älteren Schüler:innen eine deutliche Präferenz zugunsten der beiden Visualisierungen (Vierfeldertafel und Baumdiagramm). Im Gegensatz zu den jüngeren Schüler:innen zeigte sich hier insbesondere keine individuelle Präferenz mehr für den Text.

Tab. 6 Präferenzen der Schüler:innen der Klassenstufe 5/6 und 10 anhand der Fixationsdauer

Klassenstufe	Eindeutige Präferenz für eine Visualisierung				Zwei präferierte Visualisierungen				Keine Präferenz	Gesamt
	BD*	VFT**	Text	Gesamt	VFT** und Text	BD* und VFT**	VFT** und Text	Gesamt		
5/6	5	1	1	7	6	5	0	11	0	18
10	8	5	0	13	3	1	0	4	1	18
Gesamt	13	6	1	20	9	6	0	15	1	36

Für Details, wann eine bzw. zwei Darstellungsform(en) als präferiert gewertet wurde(n), vergleiche Abschnitt 4.4. In Tab. 3 wurde bereits die untere Zeile von Tab. 6 dargestellt

* BD Baumdiagramm

** VFT Vierfeldertafel

Tab. 7 Präferenzen der Schüler:innen anhand der Interviews

Klassenstufe	Eindeutige Präferenz für eine Visualisierung				Zwei präferierte Visualisierungen				Keine Präferenz	Gesamt
	BD*	VFT**	Text	Gesamt	BD* und VFT**	BD* und Text	VFT** und Text	Gesamt	Gesamt	
5/6	8	3	4	15	0	0	1	1	2	18
10	11	6	0	17	0	1	0	1	0	18
Gesamt	19	9	4	32	0	1	1	2	2	36

* BD Baumdiagramm

** VFT Vierfeldertafel

5.2.3 Passung der Eyetracking-Daten und der Aussagen in den Interviews

Im Folgenden wurde auch die Passung der Interviewaussagen zu den Eyetracking-Daten getrennt für die Klassenstufen betrachtet (vgl. Tab. 8). Bei der Analyse wurde deutlich, dass eine über beide Erhebungsmethoden konsistente Präferenz mit steigendem Alter der Schüler:innen ausgeprägter wird und zugunsten der beiden Visualisierungen ausfällt. Dabei war das Baumdiagramm mit 17 übereinstimmenden Präferenzen sowohl in den Eyetracking-Erhebungen als auch in den Interviews in beiden Klassenstufen die am meisten präferierte Darstellung.

Die Übereinstimmung zwischen den beiden Kriterien „erste Präferenz anhand der Fixationsdauer“ und „erste Präferenz anhand der Interviewdaten“ in Klassenstufe 5/6 fiel mit einem Cohens Kappa von $\kappa=0,28$ eher gering aus, während die Übereinstimmung in Klassenstufe 10 mit einem Cohens Kappa von $\kappa=0,52$ als moderat eingestuft werden kann (zur Bedeutung von Cohens Kappa vgl. Abschnitt 4.5).

5.3 Forschungsfrage 3: Präferenz und Performanz

Um die dritte Forschungsfrage zu beantworten, ob sich die Lösungsraten in Abhängigkeit von der präferierten Darstellungsform (Baumdiagramm vs. Vierfeldertafel

Tab. 8 Intraindividuelle Präferenzen der Schüler:innen, differenziert nach Jahrgangsstufe

Klassenstufe	Baumdiagramm	Vierfeldertafel	Text	Gesamt mit Präferenz	Keine Präferenz	Gesamt
5/6	7	1	2	10	8	18
10	10	4	0	14	4	18
Gesamt	17	5	2	24	12	36

vs. Text) und der Klassenstufe unterscheiden, wurden Zusammenhänge zwischen der Korrektheit bei der Beantwortung der 12 Aufgaben und der jeweils präferierten Darstellungsform analysiert. Dazu wurden im Folgenden zunächst die Performanzen aller 36 Schüler:innen dargestellt (5.3.1); für die Zusammenhänge zwischen Präferenz und Performanz wurden dann nur noch diejenigen 24 Schüler:innen betrachtet, bei denen eine eindeutige Präferenz festgestellt werden konnte (vgl. Abschn. 5.1.3). Die Auswertungen fanden dabei nicht mehr auf Individualebene, sondern auf der Ebene einzelner Aufgaben statt (d.h., es gibt zunächst 12·36 und dann noch 12·24 Untersuchungseinheiten).

5.3.1 Überblick zur Performanz

Die insgesamt 36 teilnehmenden Schüler:innen lösten im Schnitt $M(SD) = 3,89(3,80)$ der 12 Aufgaben (Min=0; Max=12) korrekt, also etwa ein Drittel der ihnen gestellten Aufgaben (gesamt: 32 %, vgl. Tab. 9). Dabei ist erwartungsgemäß ein Unterschied zwischen den Klassenstufen festzustellen: Die 18 teilnehmenden Schüler:innen der 10. Klassenstufe lösten im Schnitt $M(SD) = 5,50(3,35)$ und damit etwa 46 % der 12 Aufgaben (Min=0; Max=10) korrekt. Bei den 18 Schüler:innen der 5./6. Klassenstufe waren es hingegen nur $M(SD) = 2,28(3,61)$ und damit 19 % der 12 Aufgaben, die richtig gelöst wurden (Min=0; Max=12). Insgesamt zeigte sich also ein großer Effekt in der Performanz zugunsten der Schüler:innen der 10. Klassenstufe, die mehr als doppelt so viele Aufgaben korrekt lösen konnten wie ihre 4–5 Jahre jüngeren Mitschüler:innen (Cohens $d = 0,93$).

Betrachtet man die Lösungsraten – noch ohne Berücksichtigung der Präferenzen – getrennt nach der Kombination der zur Wahl gestellten Darstellungen (1. Baumdiagramm vs. Vierfeldertafel, 2. Baumdiagramm vs. Text, 3. Vierfeldertafel vs. Text), so zeigte sich sowohl bei Schüler:innen der Klassenstufe 5/6 als auch der Klassenstufe 10 deskriptiv eine leicht positive Tendenz zugunsten solcher Aufgaben, bei denen eine Vierfeldertafel als Darstellungsform zur Wahl stand.

Tab. 9 Lösungsraten der Aufgaben

Klassenstufe	Baumdiagramm & Vierfeldertafel (%)	Baumdiagramm & Text (%)	Vierfeldertafel & Text (%)	Gesamt (%)
5/6	18	17	21	19
10	44	40	49	46
Gesamt	31	28	35	32

Tab. 10 Performanz (Anteil korrekt gelöster Aufgaben) in Abhängigkeit von den Präferenzen

Intraindividuelle Präferenz* (Anzahl Schüler:innen)	Performanz (Anzahl richtig gelöster Aufgaben)			Performanz ohne bestimmte Präferenz (%)	Gesamt (%)
	Performanz nach Präferenz bei einzelnen Aufgaben	Baumdiagramm (%)	Vierfeldertafel (%)	Text (%)	
Baumdiagramm (N= 17)	37,0	52,8	19,0	21,8	33,8
Vierfeldertafel (N= 5**)	36,8	70,6	0	50,0	55,9
Text (N= 2)	0	0	0	0	0
Keine Präferenz (N= 12**)	27,5	45,9	14,3	9,8	25,9
Gesamt (N= 36**)	31,8	55,6	13,0	17,5	32,3

* Anhand Eyetracking-Daten und Kurzinterviews festgelegt (vgl. Abschn. 5.1.3 und 5.2.3)

** Von jedem/jeder Schüler:in gehen 12 Aufgaben – insgesamt 430 Bearbeitungen bei 432 (= 36 · 12) Aufgaben – in die Auswertung ein. Abweichungen erklären sich durch fehlende Eyetrackingdaten (n= 2) bei einzelnen Aufgaben (je eine Aufgabe bei einer/m Schüler:in mit der Präferenz Vierfeldertafel und keiner Präferenz)

5.3.2 Präferenz und Performanz

Im Folgenden wurden nur die 24 der 36 Schüler:innen mit eindeutiger Präferenz betrachtet (vgl. Abschn. 5.1.3 und 5.2.3). Schüler:innen ohne konsistente Präferenz in beiden Erhebungsformaten wurden in den Tabellen nur der Vollständigkeit halber aufgelistet.

Tab. 10 gibt eine Übersicht über die Performanz in Kombination mit der jeweiligen Präferenz. Die erste Spalte gibt die intraindividuelle Präferenz wieder, die mittels Eyetracking- und Interview-Daten festgelegt wurde. Die Performanz ist sowohl im Hinblick auf die Unterteilung der Präferenzen auf Aufgabenebene (zweite bis fünfte Spalte) als auch insgesamt (sechste Spalte) dargestellt.

Mit den Daten der 24 Schüler:innen, bei denen eine konsistente Präferenz zugunsten einer bestimmten Darstellungsform festgestellt werden konnte, wurde trotz der geringen Stichprobengröße ergänzend eine Varianzanalyse für einen ersten Überblick durchgeführt¹. Eine zweifaktorielle ANOVA mit den beiden Faktoren der intraindividuellen Präferenz und der Präferenz bei den einzelnen Aufgaben zeigte keine signifikanten Interaktionen ($F(6; 275)=0,64$; $p=0,702$). Ebenso zeigten sich keine signifikanten Effekte für intraindividuelle Präferenz ($F(2; 275)=2,91$; $p=0,056$) und für die Präferenz bei den einzelnen Aufgaben ($F(3; 275)=1,11$; $p=0,345$).

¹ Weiterhin sind Normalverteilung und Varianzhomogenität der Daten nicht erfüllt und die Unabhängigkeit der Beobachtungen ist aufgrund der aufgabenweisen Auswertung nicht gegeben. Die ANOVA ist zwar robust gegenüber einer Verletzung der Normalverteilung und der Varianzhomogenität (Field 2009), nicht aber gegenüber einer Verletzung der Unabhängigkeit. Die inferenzstatistischen Ergebnisse sind nur als ergänzende Informationen zu verstehen.

Für detailliertere Analysen wurde für die beiden Faktoren im Anschluss jeweils eine einfaktorielle ANOVA durchgeführt, wobei weiterhin die Aufgabenebene im Fokus stand. Zunächst wurde die ANOVA mit dem Faktor Präferenz bei einzelnen Aufgaben durchgeführt. Die Performanz² war mit der Vierfeldertafel am höchsten ($M_{VFT}(SD)=0,61(0,49)$), gefolgt vom Baumdiagramm ($M_{BD}(SD)=0,34(0,47)$), keiner Präferenz ($M_{kp}(SD)=0,23(0,42)$)³ und schließlich dem Text ($M_T(SD)=0,13(0,34)$). Die ANOVA zeigt einen statistisch signifikanten Unterschied zwischen den Gruppen ($F(3; 283)=11,66; p<0,001$) mit einem mittleren bis großen Effekt (partielles $\eta^2=0,11$; Döring 2023). Der Games-Howell post-hoc-Test zeigte signifikante Unterschiede in der Performanz zwischen allen Präferenzen: Vierfeldertafel vs. Baumdiagramm ($\Delta M=0,27; p=0,002$), Vierfeldertafel vs. Text ($\Delta M=0,48; p<0,001$) und Baumdiagramm vs. Text ($\Delta M=0,21; p=0,027$). Darüber hinaus zeigte sich bei der Vierfeldertafel (nicht aber beim Baumdiagramm und Text) auch ein signifikanter Unterschied gegenüber keiner Präferenz: Vierfeldertafel vs. keine Präferenz ($\Delta M=0,38; p<0,001$), Baumdiagramm vs. keine Präferenz ($\Delta M=0,11; p=0,378$) und Text vs. keine Präferenz ($\Delta M=-0,10; p=0,591$).

Analog wurde eine ANOVA für die intraindividuellen Präferenzen durchgeführt, hierbei verlagerte sich die Analyseebene von den Einzelaufgaben auf die Performanz der Schüler:innen über alle zwölf Aufgaben hinweg. Dadurch wurde systematisch die zum Messzeitpunkt vorliegende intraindividuelle Präferenz herangezogen – auch bei Aufgaben, bei denen diese von der aufgabenbezogenen Präferenz abweicht. Die Performanz über alle 12 Aufgaben hinweg war mit der Vierfeldertafel am höchsten ($M_{VFT}(SD)=6,60(3,44)$), gefolgt vom Baumdiagramm ($M_{BD}(SD)=4,06(3,60)$) und schließlich dem Text ($M_T(SD)=0,00(0,00)$)³. Die ANOVA zeigte keinen statistisch signifikanten Unterschied zwischen den Gruppen ($F(2; 21)=2,66; p=0,093$).

Insgesamt zeigte sich anhand der Ergebnisse der ANOVA sowie des Post-hoc-Tests (vgl. auch Tab. 10), dass die Performanz bei einer Präferenz der Vierfeldertafel höher (als bei den anderen Darstellungen) und bei einer Präferenz des Baumdiagramms diese wiederum höher als bei der Präferenz des Textes ausfiel. Diese Ergebnisse wurden im Folgenden noch einmal getrennt nach Klassenstufen untersucht (auch wenn dies nur noch mit qualitativem Charakter möglich ist). Zunächst wurden die 18 Schüler:innen der Klassenstufen 5 und 6 detaillierter in den Blick genommen.

Klassenstufe 5/6 Insgesamt gab es hier 10 von 18 Schüler:innen, denen mittels Eyetracking und Interviews eine klare, hinsichtlich beider Erhebungsformate übereinstimmende Präferenz (7× Baumdiagramm, 1× Vierfeldertafel, 2× Text) zugeordnet werden konnte (vgl. Abschn. 5.1.3 und 5.2.3). Tab. 11 gibt einen Überblick zur Performanz im Hinblick auf die unterschiedlichen Präferenzen. Die Performanz

² Zur Bestimmung der Performanz je Visualisierung wurde die Anzahl korrekter Lösungen durch die Anzahl derjenigen Bearbeitungen geteilt, bei denen diese Visualisierung bevorzugt wurde – unabhängig von der allgemeinen Präferenz einer Person (ohne Tabelle).

³ Bei einer aufgabenweisen Betrachtung taucht auch die Kategorie „keine Präferenz“ auf, obwohl nur Personen in die Analyse integriert waren, die eine eindeutige intraindividuelle Präferenz zeigten. Werden nur intraindividuelle Präferenzen von Personen betrachtet, bei denen sich eine eindeutige Präferenz ergab, kommt die Kategorie „keine Präferenz“ nicht vor.

Tab. 11 Performanz (Anteil korrekt gelöster Aufgaben) in Abhängigkeit von den Präferenzen in Klassenstufe 5/6

Intraindividuelle Präferenz* (Anzahl Schüler:innen)	Performanz (Anzahl richtig gelöster Aufgaben)				Gesamt (%)
	Perfomanz nach Präferenz bei Baumdiagramm (%)	Perfomanz nach Präferenz bei Vierfeldertafel (%)	Text (%)	Performanz ohne bestimmte Präferenz (%)	
Baumdiagramm ($N=7$)	32,4	53,3	0	11,1	26,2
Vierfeldertafel ($N=1$)	0	25	0	0	8,3
Text ($N=2$)	0	0	0	0	0
Keine Präferenz ($N=8$)	17,6	36,4	11,1	9,7	19,1
Gesamt ($N=18$)	20,2	40,5	3,7	9,5	19,0

* Anhand Eyetracking-Daten und Kurzinterviews festgelegt. Von jedem/jeder Schüler:in gehen 12 Aufgaben in die Auswertung ein, so dass es sich insgesamt um 216 (= 18 · 12) Aufgaben handelt

ist sowohl bezüglich der Art der Präferenz auf Aufgabenebene als auch insgesamt dargestellt (analog zu Tab. 10).

Eine zweifaktorielle ANOVA mit den beiden Faktoren der intraindividuellen Präferenz und der Präferenz bei den einzelnen Aufgaben zeigte keine signifikanten Interaktionen ($F(6; 108)=0,587$; $p=0,740$). Ebenso zeigten sich keine signifikanten Effekte für intraindividuelle Präferenz ($F(2; 108)=2,504$; $p=0,086$) und für die Präferenz bei den einzelnen Aufgaben ($F(3; 108)=0,723$; $p=0,541$).

Für detailliertere Analysen wurde für die beiden Faktoren im Anschluss jeweils eine einfaktorielle ANOVA durchgeführt. Zunächst wurde die ANOVA mit dem Faktor Präferenz bei einzelnen Aufgaben durchgeführt. Die Performanz war mit der Vierfeldertafel am höchsten ($M_{VFT}(SD)=0,45(0,51)$), gefolgt vom Baumdiagramm ($M_{BD}(SD)=0,22(0,42)$), keiner Präferenz ($M_{kp}(SD)=0,09(0,30)$)³ und schließlich dem Text ($M_T(SD)=0,00(0,00)$). Die ANOVA zeigte einen statistisch signifikanten Unterschied zwischen den Gruppen ($F(3; 116)=5,58$; $p=0,001$) mit einem großen Effekt (partiell $\eta^2=0,13$). Der Games-Howell post-hoc-Test zeigte signifikante Unterschiede in der Performanz zwischen Vierfeldertafel vs. Text ($\Delta M=0,45$; $p=0,004$) und Baumdiagramm vs. Text ($\Delta M=0,22$; $p=0,003$), jedoch keinen Unterschied zwischen Vierfeldertafel vs. Baumdiagramm ($\Delta M=0,23$; $p=0,298$). Darüber hinaus zeigte sich bei der Vierfeldertafel (nicht aber beim Baumdiagramm und Text) auch ein signifikanter Unterschied gegenüber keiner Präferenz: Vierfeldertafel vs. keine Präferenz ($\Delta M=0,36$; $p=0,040$), Baumdiagramm vs. keine Präferenz ($\Delta M=0,13$; $p=0,386$) und Text vs. keine Präferenz ($\Delta M=-0,094$; $p=0,297$).

Analog wurde eine ANOVA für die intraindividuellen Präferenzen durchgeführt, hierbei verlagerte sich die Analyseebene wieder von den Einzelaufgaben auf die Performanz der Schüler:innen über alle zwölf Aufgaben hinweg. Dadurch wird systematisch die zum Messzeitpunkt vorliegende intraindividuelle Präferenz herangezogen – auch bei Aufgaben, bei denen diese von der aufgabenbezogenen Präferenz abweicht. Die Performanz über alle 12 Aufgaben hinweg war mit dem Baumdiagramm am höchsten ($M_{VFT}(SD)=3,14(3,94)$), gefolgt von der Vierfeldertafel ($M_{BD}(SD)=1,00(—)$) und schließlich dem Text ($M_T(SD)=0,00(0,00)$)³. Die ANOVA zeigte keinen statistisch signifikanten Unterschied zwischen den Gruppen ($F(2; 71)=0,650$; $p=0,551$), was natürlich aufgrund der geringen Stichprobengröße nicht überbewertet werden darf.

Insgesamt zeigte sich in der Klassenstufe 5/6, dass die Performanz bei einer Präferenz mit einer der beiden Visualisierungen (Baumdiagramm und Vierfeldertafel) höher war im Vergleich zur Präferenz des Texts. Das bedeutet, dass Schüler:innen, die eine der beiden Visualisierungen als Informationsquelle zur Aufgabenbearbeitung bevorzugten, besser abschnitten als Schüler:innen, die den Text bevorzugten oder – zumindest im Vergleich zur Wahl der Vierfeldertafel – die keine klare Präferenz hatten.

Klassenstufe 10 In Klassenstufe 10 zeigte sich bei 14 von insgesamt 18 Schüler:innen eine eindeutige Präferenz zugunsten einer Darstellungsform (vgl. Abschn. 5.1.3 und 5.2.3). 10 Schüler:innen bevorzugten das Baumdiagramm und 4 Schüler:innen die Vierfeldertafel, aber niemand den Text. Tab. 12 gibt einen Überblick

Tab. 12 Performanz (Anteil korrekt gelöster Aufgaben) in Abhängigkeit von den Präferenzen in Klassenstufe 10

Intraindividuelle Präferenz* (Anzahl Schüler:innen)	Performanz (Anzahl richtig gelöster Aufgaben)				Gesamt (%)
	Baumdiagramm (%)	Vierfeldertafel (%)	Text (%)	Performanz ohne bestimmte Präferenz (%)	
Baumdiagramm ($N = 10$)	39,7	52,4	30,8	32,1	39,2
Vierfeldertafel ($N = 4^{**}$)	50	76,7	0	100	68,1
Text ($N = 0$)	–	–	–	–	–
Keine Präferenz ($N = 4^{**}$)	47,1	60,0	20,0	10,0	40,4
Gesamt ($N = 18^{**}$)	42,7	62,3	26,3	30	45,8

* Anhand Eyetracking-Daten und Kurzinterviews festgelegt

** Von jedem/jeder Schüler:in gehen 12 Aufgaben in die Auswertung ein. Insgesamt 214 Aufgaben. Von 216 (= $18 \cdot 12$) abweichende Werte erklären sich durch fehlende Eyetrackingdaten ($n=2$) bei einzelnen Aufgaben (je eine Aufgabe bei einer/m Schüler:in mit der Präferenz Vierfeldertafel und keiner Präferenz)

zur Performanz im Hinblick auf die unterschiedlichen Präferenzen (analog zu Tab. 10 und 11).

Auch mit den 14 Schüler:innen aus Klassenstufe 10, bei denen mittels Eye-tracking und Interview eine eindeutige Präferenz festgelegt werden konnte, wurde eine Varianzanalyse durchgeführt. Eine zweifaktorielle ANOVA mit den beiden Faktoren der intraindividuellen Präferenz und der Präferenz bei den einzelnen Aufgaben zeigte keine signifikanten Interaktionen ($F(3; 159)=1,14; p=0,334$). Ebenso zeigten sich keine signifikanten Effekte für intraindividuelle Präferenz ($F(1; 159)=1,25; p=0,266$) und für die Präferenz bei den einzelnen Aufgaben ($F(3; 159)=2,28; p=0,081$).

Die anschließende einfaktorielle ANOVA mit dem Faktor Präferenz bei einzelnen Aufgaben ergab, dass die Performanz mit der Vierfeldertafel am höchsten war ($M_{VFT}(SD)=0,67(0,48)$), gefolgt vom Baumdiagramm ($M_{BD}(SD)=0,42(0,50)$), keiner Präferenz ($M_{kp}(SD)=0,37(0,49)$)³ und schließlich dem Text ($M_T(SD)=0,29(0,47)$). Die ANOVA zeigte einen statistisch signifikanten Unterschied zwischen den Gruppen ($F(3; 163)=4,18; p=0,007$) mit einem mittleren Effekt (partielles $\eta^2=0,07$). Der Games-Howell post-hoc-Test zeigte einen signifikanten Unterschied in der Performanz zwischen Vierfeldertafel vs. Baumdiagramm ($\Delta M=0,25; p=0,029$), jedoch keine signifikanten Unterschiede zwischen Vierfeldertafel vs. Text ($\Delta M=0,38; p=0,062$) sowie Baumdiagramm vs. Text ($\Delta M=0,13; p=0,780$). Darüber hinaus zeigte sich bei der Vierfeldertafel (nicht aber beim Baumdiagramm und Text) auch ein signifikanter Unterschied gegenüber keiner Präferenz: Vierfeldertafel vs. keine Präferenz ($\Delta M=0,300; p=0,045$), Baumdiagramm vs. keine Präferenz ($\Delta M=0,050; p=0,966$) und Text vs. keine Präferenz ($\Delta M=-0,081; p=0,952$).

Analog wurde eine ANOVA für die intraindividuellen Präferenzen durchgeführt, hierbei verlagerte sich die Analyseebene wieder von den Einzelaufgaben auf die Performanz der Schüler:innen über alle zwölf Aufgaben hinweg. Die Performanz über alle 12 Aufgaben hinweg war mit der Vierfeldertafel am höchsten ($M_{VFT}(SD)=8,00(1,63)$), gefolgt vom Baumdiagramm ($M_{BD}(SD)=4,70(3,40)$). Die ANOVA zeigte keinen statistisch signifikanten Unterschied zwischen den Gruppen ($F(1; 12)=3,33; p=0,093$).

Ähnlich zu den Auswertungen für die Schüler:innen der Klassenstufe 5/6 zeigten sich auch in Klassenstufe 10 Performanzunterschiede je nach Präferenz einer bestimmten Darstellungsform. Die Performanz war bei den Schüler:innen höher, die die Vierfeldertafel gegenüber dem Baumdiagramm präferierten oder keine Präferenz bei einer Aufgabe zeigten. Im Hinblick auf die Präferenz Text zeigten sich keine signifikanten Unterschiede, weder gegenüber der Vierfeldertafel noch dem Baumdiagramm.

6 Diskussion

Ziel der vorliegenden Studie war es, bei im Häufigkeitsformat gestellten Bayesischen Aufgaben (individuelle) Präferenzen von Schüler:innen bezüglich verschiedener Informationsdarstellungen (Baumdiagramm vs. Vierfeldertafel vs. Text) zu identifizieren. Weiterhin sollten diesbezügliche Klassenstufenunterschiede (5./6. vs.

10. Klasse) beleuchtet werden und dabei auch der Frage nachgegangen werden, ob mit den jeweils individuell präferierten Darstellungen auch die beste Performanz erzielt wird.

Im Folgenden werden die Ergebnisse der Studie im Hinblick auf diese Ziele zusammengefasst, interpretiert und in den Kontext der bisherigen Forschung eingeordnet (Abschn. 6.1–6.3), Limitationen diskutiert sowie Implikationen für Wissenschaft und Praxis beleuchtet (Abschn. 6.4).

6.1 Präferenz

Methodisch wurden zwei verschiedene Zugänge zur Präferenz von Schüler:innen umgesetzt: Ein Zugang mittels Eyetracking, bei dem sowohl ein optischer Eindruck mittels Heatmaps als auch eine Auswertung mithilfe von kumulierten Fixationszeiten berichtet wurden, und ein Zugang mittels Kurzinterviews, bei denen die drei Darstellungen durch die Schüler:innen in eine präferierte Reihenfolge zu bringen waren. Beide Zugänge zeigen, dass das Baumdiagramm als Darstellungsform gegenüber der Vierfeldertafel und dem Text bevorzugt wird. Begründet wurde dies in den Interviews vor allem mit der Vertrautheit und der Strukturierung bzw. besseren Übersichtlichkeit der Visualisierung. Bezüglich der Vierfeldertafel ist über alle Schüler:innen hinweg keine so klare Präferenz gegenüber einer textuellen Informationsdarstellung zu sehen. Besonders interessant ist, dass einige wenige Schüler:innen auch den Text präferierten, dies jedoch (selbst intraindividuell) nicht durchgängig bei allen Aufgaben. Insgesamt bestätigen die vorliegenden Ergebnisse zu den Präferenzen von Schüler:innen bei Bayesianischen Aufgaben die bisherigen Befunde im Bereich bedingter Wahrscheinlichkeiten, bei denen ebenfalls das Baumdiagramm, gefolgt von der Vierfeldertafel, als Präferenz erkennbar war (Rach 2018).

Obwohl sich auch die Vierfeldertafel als Visualisierung durch eine strukturierte Darbietung der relevanten Informationen auszeichnet (nested style), wird das Baumdiagramm deutlich präferiert. Die Aussagen der Schüler:innen in den Interviews legen eine intuitive Bevorzugung einer sequenziellen Darstellung von Informationen nahe, wie sie im Besonderen in einem branch style verwirklicht ist (vgl. Abschn. 2.2.2). Umgekehrt spricht die fehlende Vertrautheit mit der Vierfeldertafel (sie ist in Hessen nicht fest im Lehrplan verankert) im Kontext Bayesianischer Aufgaben oder bedingter Wahrscheinlichkeiten sowie die fehlende sequenzielle Darbietung der Informationen gegen die Vierfeldertafel.

6.2 Präferenz und Klassenstufe

Betrachtet man die Präferenzen der Schüler:innen aufgeteilt nach den Klassenstufen (5/6 vs. 10; Forschungsfrage 2), so ergibt sich zunächst ein ähnliches Bild wie in der Gesamtbetrachtung: Das Baumdiagramm war sowohl in der Unterstufe (Klassenstufe 5/6) als auch in der Mittelstufe (Klassenstufe 10) bei beiden untersuchten Zugängen (Heatmaps und Fixationsdauer mittels Eyetracking sowie Kurzinterviews) die präferierte Informationsdarstellung. Während jedoch bei den Schüler:innen der Klassenstufe 10 die Vierfeldertafel – nach dem Baumdiagramm – die präferierte Darstellungsform war, zeigte sich bei den Schüler:innen der Klassenstufe 5/6 vor allem,

das häufig zwei Darstellungsformen *parallel* genutzt werden, darunter insbesondere auch der Text. Da die Schüler:innen zur Beantwortung von Forschungsfrage 2 in zusätzliche Subgruppen unterteilt wurden und die resultierende Stichprobengröße entsprechend gering ausfiel, sind die berichteten Ergebnisse unter diesem methodischen Aspekt mit Zurückhaltung zu interpretieren. Zur Erhöhung der statistischen Aussagekraft und zur Absicherung der Befunde wäre eine Folgestudie mit einer größeren Stichprobe wünschenswert.

Wie bereits Rach (2018) vermutet, werden nicht so bekannte und vertraute Visualisierungen weniger präferiert als solche, mit denen der Umgang bereits geläufig ist. Dies bestätigt sich zumindest teilweise in der vorliegenden Studie: Das Baumdiagramm wurde von den Schüler:innen beider Klassenstufen bevorzugt. Bei denjenigen der Unterstufe zeigte sich hier jedoch deutlich die Tendenz, *beide* gegebenen Visualisierungen zu nutzen, während die Schüler:innen aus Klassenstufe 10 bereits eine deutlichere Präferenz für *eine* der gegebenen Darstellungsformen zeigten (bevorzugt Baumdiagramm).

Ein Grund für die bevorzugte Nutzung von zwei Darstellungsformen parallel in Klassenstufe 5/6 könnte in einer gewissen Unsicherheit liegen, die beim Umgang mit den Visualisierungen noch vorliegt. Das Baumdiagramm wird erst in Klassenstufe 8 systematisch eingeführt, sodass es bei Schüler:innen der Klassenstufe 5/6 wenig bis gar nicht bekannt ist, wenngleich die Vierfeldertafel oder optisch ähnliche Visualisierungen (wie bspw. Verknüpfungstabellen) bereits in der Grundschule behandelt werden (aber natürlich noch nicht im Zusammenhang mit bedingten Wahrscheinlichkeiten). Eine Vermutung ist daher, dass die Schüler:innen der Unterstufe noch stärker dazu tendieren, die Informationen aus den verschiedenen Informationsdarstellungen zu vergleichen.

6.3 Präferenz und Performanz

Die Ergebnisse reihen sich bezüglich der Performanz in die aktuelle Forschung zur Wirksamkeit von Visualisierungen im Hinblick auf Bayesianisches Denken ein (Yamagishi 2003; Binder et al. 2015, 2020). Durch die vorliegende Studie konnte darüber hinaus gezeigt werden, dass Vierfeldertafel und Baumdiagramm nicht nur im Hinblick auf die Performanz einen Mehrwert bieten, sondern gleichzeitig auch von den Schüler:innen gegenüber textuellen Darstellungen bevorzugt zum Lösen Bayesianischer Aufgaben verwendet wurden.

Die Ergebnisse deuten darauf hin, dass mit der Vierfeldertafel eine höhere Performanz erreicht wird als mit dem Baumdiagramm oder dem Text. Insgesamt zeigten Schüler:innen aus Klassenstufe 10 eine höhere Performanz als diejenigen aus Klassenstufe 5/6, wobei selbst die 10. Klässler:innen im Durchschnitt nur knapp die Hälfte aller Aufgaben richtig lösten.

Hinsichtlich des Zusammenhangs einer bestimmten Informationsdarstellung mit der Lösungsrate lassen sich erkennbare Unterschiede zwischen Baumdiagramm, Vierfeldertafel und Text feststellen. In der jeweiligen Gegenüberstellung von zwei gegebenen Informationsdarstellungen (Baumdiagramm vs. Vierfeldertafel, Baumdiagramm vs. Text, Vierfeldertafel vs. Text) zeigte sich, dass die Verwendung der Vierfeldertafel zu einer signifikant höheren Lösungsrate führt als die Verwendung

des Baumdiagramms oder des Textes. Diese Ergebnisse bestätigen sowohl die bereits erwähnten Vorteile der Vierfeldertafel im Hinblick auf Prinzipien des multimedialen Lernens wie dem Spatial-Contiguity-Prinzip (Mayer 2009; vgl. Abschn. 2.2.2) als auch die Ergebnisse vorheriger Studien, die zeigten, dass eine Vierfeldertafel in Kombination mit natürlichen Häufigkeiten die Lösungsrate gegenüber einer reinen Textdarstellung signifikant steigert (Binder et al. 2015). Es zeigen sich also insgesamt widersprüchliche Ergebnisse in Bezug auf die beiden Visualisierungen: Während das Baumdiagramm bevorzugt wird, lassen sich mit der Vierfeldertafel höhere Lösungsraten erzielen.

Ebenso bleibt die Frage offen, warum sich Schüler:innen in ihrer Präferenz vor allem für das Baumdiagramm entscheiden, während die Vierfeldertafel insgesamt eine höhere Lösungsrate verspricht. Eine mögliche Ursache ist wieder die Vertrautheit im Umgang mit den Visualisierungen. Das Baumdiagramm wird in Hessen im Bildungsplan der 8. Klasse aufgeführt und entsprechend unterrichtet, während die Vierfeldertafel in den unteren Klassenstufen nicht explizit auftaucht und daher (in der Regel) nicht so ausführlich besprochen wird. Interessant wäre entsprechend eine Untersuchung mit Schüler:innen, in der die Unterschiede, die durch die Bekanntheit entstehen, keine so große Rolle spielen. Denkbar wäre beispielsweise eine Interventionsstudie mit einer ausführlichen Einführung zu allen verwendeten Darstellungsformen. In der vorliegenden Studie wurden die Visualisierungen Baumdiagramm und Vierfeldertafel zwar vorgestellt, jedoch nicht ausführlich besprochen oder zuvor explizit angewendet.

6.4 Limitationen und Implikationen

Die Ergebnisse der Studie gliedern sich in den aktuellen Forschungsstand zu Visualisierungen bei Bayesianischen Aufgaben ein und bieten einen Einblick in die Präferenzen von Schüler:innen im Rahmen der Propädeutik bedingter Wahrscheinlichkeiten. Eine Limitation ist dabei die Stichprobengröße, zumal die Stichprobe lediglich aus einem einzigen Bundesland bzw. aus einer einzigen Schule stammt. Dadurch, dass Eyetracking eine aufwändige Methode ist, bei der die Teilnehmer:innen jeweils einzeln befragt werden müssen, ist das Problem einer vergleichsweise geringen Stichprobe jedoch in der Regel nicht zu umgehen (z. B. Abt et al. 2024).

Insbesondere für die Forschungsfragen 2 und 3, bei der die Stichprobe in weitere Gruppen unterteilt wird, ist die Berücksichtigung dieser Limitation bei der Interpretation der Ergebnisse relevant. Hinzu kommt, dass einzelne Präferenzkategorien nur sehr gering besetzt sind. Während sich über alle Schüler:innen hinweg mehrere Teilnehmende mit eindeutiger Präferenz für Baumdiagramm oder Vierfeldertafel finden, zeigen lediglich zwei Schüler:innen eine Präferenz für die Textdarstellung. Auf Ebene der Klassenstufen reduziert sich diese Zahl teilweise weiter; in der Klassenstufe 5/6 präferiert lediglich eine Person die Vierfeldertafel (was keine Generalisierung zulässt). Entsprechend lassen sich diese Ergebnisse nicht verallgemeinern. In der Klassenstufe 10 wird die Textdarstellung zudem nicht als intraindividuelle Präferenz angegeben, sodass dieser Fall in den Analysen nicht berücksichtigt werden konnte. Dennoch zeigt die Studie deutliche Tendenzen sowohl in Bezug auf die Präferenz als auch auf die Performanz, die mit dem momentanen Forschungs-

stand konform sind. Zukünftige Studien mit größeren Stichproben könnten hier die beobachteten Tendenzen – insbesondere im Hinblick auf die zugunsten der Vierfeldertafel erkennbaren Performanzeffekte – weiter absichern.

Speziell in Bezug zu Forschungsfrage 3, bei der je 12 individuelle Aufgabenbearbeitungen von 24 Schüler:innen untersucht wurden, besteht eine weitere Limitation darin, dass diese Untersuchungseinheiten nicht unabhängig voneinander sind, da die Daten innerhalb der betrachteten Schüler:innen geschachtelt sind (vgl. Fußnote 1).

Eine weitere Limitation betrifft das – zu einem gewissen Grad arbiträre – Präferenzkriterium. Was ein geeigneter Richtwert für die kumulierte Blickdauer ist bzw. ab welchem eine Darstellung als präferiert gegenüber einer anderen einzustufen ist, kann nicht normativ beantwortet werden. Eine Pilotierungsstudie (vgl. Abschn. 4.5) sowie empirische Vergleiche verschiedener Kriterien (z. B. 60 %, 70 %, 75 % und 80 %) ließen für die vorliegende Studie einen Cut-Off-Wert von 70 % als geeignet erscheinen.

Die Methode des Eyetracking bietet eine Vielzahl an Auswertungsmöglichkeiten, von denen hier vor allem zwei genutzt wurden: Heatmaps und Fixationsdauer. Hierbei stellen die Heatmaps nur einen optischen Eindruck dar, der nicht ohne Weiteres quantifizierbar ist. Diese Eindrücke sind daher nur als visueller Einblick zu bewerten und nicht belastbar zu generalisieren. Daher wurden in dieser Studie die optischen Eindrücke der Heatmaps sowohl mit quantifizierbarer Fixationsdauer als auch mit anschließenden Interviews kombiniert, um Ergebnisse wechselseitig validieren zu können.

Der Artikel verwirklicht einen Zugang zu Präferenzen bei (häufigkeitsbasierten) Bayesianischen Aufgaben, der mit einer Kombination aus Kurzinterviews und Eyetracking gewisse Limitationen der bisherigen Forschung (z. B. lediglich Notizen oder Selbstauskünfte, Acevedo Nistal et al. 2012) vermeiden konnte. Denkbar wären als Ergänzung zum Beispiel tiefergehende Analysen mittels GazePlots (Visualisierung beim Eyetracking bzgl. Reihenfolge und Dauer der Fixationen), eine Analyse der Präferenzen im Hinblick auf Variationen in Abhängigkeit der Aufgabenstellung in Bezug auf die exakte Formulierung oder die interne Konsistenz der Präferenz. So ergeben sich zahlreiche weitere Forschungsdesiderate für zukünftige Untersuchungen.

Weiterhin muss die Frage offen bleiben, ob die Präferenz der Vierfeldertafel zu besseren Ergebnissen führt oder ob bessere Schüler:innen die Vierfeldertafel bevorzugen (Frage der Kausalität). Für die Schule ist es im Sinne einer Methodenpluralität sicher sinnvoll, beide Visualisierungen nicht nur (unkommentiert) zu verwenden, sondern auch deren Nutzung und Bedeutung zu erläutern. Eine Implikation für die Praxis ist also, verschiedene Visualisierungen (insbesondere auch die Vierfeldertafel) früh in den Unterricht zu integrieren und so möglichst bald einen Grundstein für Risk Literacy zu legen (Eichler und Vogel 2015; Gigerenzer 2020). Im Gegensatz zur bekannten Schwierigkeit, Bayesianische Schlussfolgerungen im Wahrscheinlichkeitsformat zu ziehen (McDowell und Jacobs 2017), ist es auf der Basis natürlicher Häufigkeiten jedenfalls möglich, bereits mit Schüler:innen der Klassenstufe 5/6 erste Grundlagen für solche Denkweisen zu legen.

Unsere Empfehlung und Implikation für den Unterricht ist dabei, zunächst mit dem bereits von jungen Schüler:innen präferierten Baumdiagramm Interesse zu we-

cken und die Bayesianische Situation anschließend mit der (verständnisfördernden) Vierfeldertafel zu vertiefen. Ein alternativer Weg könnte auch aus Verbesserungen bzw. Weiterentwicklungen von Baumdiagrammen bestehen, so dass diese nicht nur attraktiv für den Betrachter, sondern auch effektiver für die Lösungsfindung werden (vgl. z.B. Doppelbäume oder Netzdiagramme; Binder et al. 2023).

Funding Open Access funding enabled and organized by Projekt DEAL.

Datenverfügbarkeit Die Rohdaten, auf denen die Schlussfolgerungen dieses Artikels beruhen, werden von den Autoren ohne ungerechtfertigte Einschränkungen zur Verfügung gestellt.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Abt, M., Leuders, T., Loibl, K., Strohmaier, A. R., van Dooren, W., & Reinhold, F. (2024). How can eye-tracking data be used to understand cognitive processes when comparing data sets with box plots? *Frontiers in Education*, 9, 1425663. <https://doi.org/10.3389/feduc.2024.1425663>.
- Acevedo Nistal, A., van Dooren, W., & Verschaffel, L. (2012). What counts as a flexible representational choice? An evaluation of students' representational choices to solve linear function problems. *Instructional Science*, 40(6), 999–1019. <https://doi.org/10.1007/s11251-011-9199-9>.
- Ackermann, L., & Domhardt, M. (2015). Vorhersagbarkeit von Produktpräferenzen durch Eyetracking. In S. Diefenbach, N. Henze & M. Pielot (Hrsg.), *Mensch und Computer 2015 – Tagungsband*. Mensch & Computer – Tagungsbände / Proceedings, (Bd. 2015, S. 395–398). Bd: De Gruyter. <https://doi.org/10.1515/9783110443929-061>.
- Aven, T. (2024). Risk literacy: Foundational issues and its connection to risk science. *Risk Analysis*, 44(5), 1011–1020. <https://doi.org/10.1111/risa.14223>.
- 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*. <https://doi.org/10.3389/fpsyg.2015.01186>.
- Binder, K., Krauss, S., & Steib, N. (2020). Bedingte Wahrscheinlichkeiten und Schnittwahrscheinlichkeiten gleichzeitig visualisieren: Das Häufigkeitsnetz. *Stochastik in der Schule*, 40(2), 2–14.
- Binder, K., Steib, N., & Krauss, S. (2023). Von Baumdiagrammen über Doppelbäume zu Häufigkeitsnetzen – kognitive Überlastung oder didaktische Unterstützung? *Journal für Mathematik-Didaktik*, 44(2), 471–503. <https://doi.org/10.1007/s13138-022-00215-9>.
- Böcherer-Linder, K., Eichler, A., & Vogel, M. (2018). Die Formel von Bayes: Kognitionspsychologische Grundlagen und empirische Untersuchungen zur Bestimmung von Teilmenge-Grundmenge-Beziehungen. *Journal für Mathematik-Didaktik*, 39(1), 127–146. <https://doi.org/10.1007/s13138-018-0128-1>.
- Bruckmaier, G., Binder, K., Krauss, S., & Kufner, H.-M. (2019). An eye-tracking study of statistical reasoning with tree diagrams and 2×2 tables. *Frontiers in Psychology*, 10, 632. <https://doi.org/10.3389/fpsyg.2019.00632>.
- Bruckmaier, G., Krauss, S., Binder, K., Hilbert, S., & Brunner, M. (2021). Tversky and Kahneman's cognitive illusions: Who can solve them, and why? *Frontiers in Psychology*, 12, 584689. <https://doi.org/10.3389/fpsyg.2021.584689>.

- Büchter, T., Steib, N., Böcherer-Linder, K., Eichler, A., Krauss, S., Binder, K., & Vogel, M. (2022a). Designing visualisations for Bayesian problems according to multimedia principles. *Education Sciences*, 12(11), 739. <https://doi.org/10.3390/educsci12110739>.
- Büchter, T., Eichler, A., Steib, N., Binder, K., Böcherer-Linder, K., Krauss, S., & Vogel, M. (2022b). How to train novices in Bayesian reasoning. *Mathematics*. <https://doi.org/10.5283/EPUB.52374>.
- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in cognitive sciences*, 10(7), 287–291. <https://doi.org/10.1016/j.tics.2006.05.007>.
- Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *WIREs Cognitive Science*, 1(6), 811–823. <https://doi.org/10.1002/wcs.79>.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). *PsycTESTS Dataset*. <https://doi.org/10.1037/t45862-000>.
- Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88(4), 715–730. <https://doi.org/10.1037/0022-0663.88.4.715>.
- Cui, L., Lo, S., & Liu, Z. (2023). The Use of Visualizations to Improve Bayesian Reasoning: A Literature Review. *Vision*, 7(1), 17. <https://doi.org/10.3390/vision7010017>
- Döring, N. (2023). Bestimmung von Teststärke, Effektgröße und optimalem Stichprobenumfang. In N. Döring (Hrsg.), *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6. Aufl. S. 789–846). Springer. https://doi.org/10.1007/978-3-662-64762-2_14.
- Eichler, A., & Vogel, M. (2015). Teaching risk in school. *The Mathematics Enthusiast*, 12(1–3), 168–183. <https://doi.org/10.54870/1551-3440.1342>.
- Fiedler, K., & von Sydow, M. (2015). Heuristics and biases: Beyond Tversky and Kahneman's (1974) judgment under uncertainty. In M. W. Eysenck & D. Groome (Hrsg.), *Cognitive psychology: Revisiting the classic studies* (S. 146–161). SAGE.
- Field, A. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)* (3. Aufl.). SAGE. <https://permalink.obvsg.at/AC07539716>
- Figueiras, A. (2018). A review of visualization assessment in terms of user performance and experience. In *22nd International Conference Information Visualisation (IV)* (S. 145–152). <https://doi.org/10.1109/iv.2018.00035>.
- Gigerenzer, G. (2012). Risk literacy. In J. Brockman (Hrsg.), *This will make you smarter* (S. 259–261). Harper Collins.
- Gigerenzer, G. (2015). *Risk savvy: How to make good decisions*. Penguin Books.
- Gigerenzer, G. (2020). *Risiko: Wie man die richtigen Entscheidungen trifft* (3. Aufl.). Pantheon.
- Gigerenzer, G. (2024). The rationality wars: A personal reflection. *Behavioural Public Policy*. <https://doi.org/10.1017/bpp.2024.51>.
- 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>.
- Gilovich, T., Griffin, D., & Kahneman, D. (Hrsg.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511808098>.
- Hahn, C. (1997). *Conjoint- und discrete Choice-Analyse als Verfahren zur Abbildung von Präferenzstrukturen und Produktauswahlentscheidungen: Ein theoretischer und computergestützter empirischer Vergleich*. Lit-Verlag. <https://permalink.obvsg.at/AC01986510>
- Hardy, I., Decristan, J., & Klieme, E. (2019). *Adaptive teaching in research on learning and instruction*. <https://doi.org/10.25656/01:18004>.
- Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12), 517–523. <https://doi.org/10.1016/j.tics.2009.09.004>.
- Hessisches Kultusministerium (2010). *Lehrplan Mathematik: Gymnasialer Bildungsgang (Jahrgangsstufen 5G bis 9G)*. <https://kultus.hessen.de/sites/kultus.hessen.de/files/2021-06/g8-mathematik.pdf>
- Jones, I., & Davies, B. (2024). Comparative judgement in education research. *International Journal of Research & Method in Education*, 47(2), 170–181. <https://doi.org/10.1080/1743727X.2023.2242273>.
- Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: Can comparative judgement help? *Educational Studies in Mathematics*, 89(3), 337–355. <https://doi.org/10.1007/s10649-015-9607-1>.
- 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](https://doi.org/10.1016/0010-0285(72)90016-3).
- Khan, A., Breslav, S., & Hornbæk, K. (2018). Interactive instruction in Bayesian inference. *Human-Computer Interaction*, 33(3), 207–233. <https://doi.org/10.1080/07370024.2016.1203264>.

- Kohler, M. P. (1996). Risk-taking behavior: A cognitive approach. *Psychological Reports*, 78(2), 489–490. <https://doi.org/10.2466/pr0.1996.78.2.489>.
- Krauss, S., Weber, P., Binder, K., & Bruckmaier, G. (2020). Natürliche Häufigkeiten als numerische Darstellungsart von Anteilen und Unsicherheit – Forschungsdesiderate und einige Antworten. *Journal für Mathematik-Didaktik*, 41, 485–521. <https://doi.org/10.1007/s13138-019-00156-w>.
- Lumer, C. (2022). Präferenzen, Nutzen und ihre Aggregation. In C. Hiebaum (Hrsg.), *Handbuch Gemeinwohl* (1. Aufl. S. 177–193). Springer VS. https://doi.org/10.1007/978-3-658-21085-4_12.
- Lusardi, A. (2015). Risk literacy. *Italian Economic Journal*, 1(1), 5–23. <https://doi.org/10.1007/s40797-015-0011-x>.
- Malone, S., Altmeyer, K., Vogel, M., & Brinken, R. (2020). Homogeneous and heterogeneous multiple representations in equation-solving problems: An eye-tracking study. *Journal of Computer Assisted Learning*, 36(6), 781–798. <https://doi.org/10.1111/jcal.12426>.
- Mayer, R. E. (2009). *Multimedia learning* (2. Aufl.). Cambridge University Press.
- 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>.
- Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological bulletin*, 134(2), 270–300. <https://doi.org/10.1037/0033-2909.134.2.270>.
- Rach, S. (2018). Visualisierungen bedingter Wahrscheinlichkeiten – Präferenzen von Schülerinnen und Schülern. *mathematica didactica*, 41(1), 1–18. <https://doi.org/10.18716/OJS/MD/2018.1371>.
- Saket, B., Endert, A., & Stasko, J. (2016). Beyond usability and performance. In M. Sedlmair (Hrsg.), *BE-LIV 2016: Proceedings : Beyond time and errors: Novel evaluation methods for visualization: sixth bi-annual workshop (Baltimore, MD, USA, October 24, 2016)*. ACM International Conference Proceedings Series. (S. 133–142). The Association for Computing Machinery. <https://doi.org/10.1145/2993901.2993903>.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In S. N. Spencer (Hrsg.), *Proceedings of the Eye Tracking Research & Applications Symposium 2000* (S. 71–78). ACM. <https://doi.org/10.1145/355017.355028>.
- Schulze, C., & Hertwig, R. (2021). A description–experience gap in statistical intuitions: Of smart babies, risk-savvy chimps, intuitive statisticians, and stupid grown-ups. *Cognition*, 210, 104580. <https://doi.org/10.1016/j.cognition.2020.104580>.
- Spiegelhalter, D. (2014). *Using expected frequencies when teaching probability*. <https://understandinguncertainty.org/using-expected-frequencies-when-teaching-probability>
- Spiegelhalter, D., & Gage, J. (2014). What can educators learn from real-world communication of risk and uncertainty? In K. Makar, B. de Sousa & R. Gould (Hrsg.), *Sustainability in statistics education: Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)*. International Statistical Institute.
- Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science*, 333(6048), 1393–1400. <https://doi.org/10.1126/science.1191181>.
- Steib, N., Büchter, T., Eichler, A., Binder, K., Krauss, S., Böcherer-Linder, K., Vogel, M., & Hilbert, S. (2025). How to teach Bayesian reasoning: An empirical study comparing four different probability training courses. *Learning and Instruction*, 95, 102032. <https://doi.org/10.1016/j.learninstruc.2024.102032>.
- Steiner, M. (2007). *Nachfragerorientierte Präferenzmessung: Bestimmung zielgruppenspezifischer Eigenschaftssets auf Basis von Kundenbedürfnissen* (1. Aufl.). Dt. Univ.-Verl.. <https://doi.org/10.1007/978-3-8350-9593-9>.
- Till, C. (2014). Fostering risk literacy in elementary school. *Mathematics Education*, 9(2), 83–96.
- Tobii Pro AB (2014). *Tobii Pro lab (Version 1.16232461 (x64) [Computer Software]*. Tobii Pro AB.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9).
- Yamagishi, K. (2003). Facilitating normative judgments of conditional probability: Frequency or nested sets? *Experimental Psychology*, 50(2), 97–106. <https://doi.org/10.1026/1618-3169.50.2.97>.
- Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, 98(3), 287–308. <https://doi.org/10.1016/j.cognition.2004.12.003>.

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.