Risikomythen in der IT-Sicherheit

Was wir falsch glauben – und was wirklich hilft.

Kai Jendrian · Oliver Oettinger
Secorvo Security Consulting GmbH

heise secIT 2026

Die Fleißarbeit bei diesem Handout wurde mit freundlicher Unterstützung von Anthropic Sonnet 4.6 geleistet.
Die Grafiken wurden mit freundlicher Unterstützung von Google Gemini 3.1 umgesetzt.

Inhalt

1. Einführung: All models are wrong – but some are useful 4
   1.1. Die Welt der Regulatorik und Standards 4
   1.2. Risikoanalyse unterstützt Entscheidungsfindung 4
   1.3. Value of Information 5
2. Mythos 1: „Wir reden alle über Risiken.“ 5
   2.1. Ein Workshop. Zehn Leute. 5
   2.2. Drei Frameworks. Drei Sprachen. Null Konsens. 6
   2.3. Schritt 1: Einigt euch auf eine Definition. 6
   2.4. Schritt 2: Den Rest für alle festlegen. 6
3. Fail: „Du hättest das wissen müssen.“ 7
   3.1. Awareness hat zwei Ziele. 7
   3.2. Das Swiss-Cheese-Modell. 7
   3.3. Just Culture. 8
4. Mythos 2: „Ich kann keine Wahrscheinlichkeit angeben – ich habe keine Daten.“ 8
   4.1. Bertrand Russell · 1929. 9
   4.2. Drei Adäquatheitskriterien. 9
   4.3. Sechs Interpretationen für einen Begriff. 9
   4.4. Klassische Wahrscheinlichkeit: Symmetrie als Fundament. 10
   4.5. Frequentistische Wahrscheinlichkeit: Wiederholung als Voraussetzung. 10
   4.6. Subjektive Wahrscheinlichkeit: Überzeugungsgrad als Ausgangspunkt. 11
   4.7. Bayes’ Theorem: Vom Prior zur belastbaren Abschätzung. 11
   4.8. Imprecise Probabilities: Intervall statt Scheinsicherheit. 12
5. Mythos 3: „One method fits all.“ 12
   5.1. Ein Admin. Ein Risiko. Ein falscher Weg. 13
   5.2. Drei Ebenen. Drei Aufgaben. 13
   5.3. Vergleich: Luftfahrt. 14
   5.4. Die taktische Lücke. 14
6. Mythos 4: „Die Welt ist eine Ampel – Rot-Gelb-Grün erklärt alles.“ 14
   6.1. Von Begriff zu Maß: Risiko als Konzept – und als Zahl. 15
   6.2. Warum Skalen wichtig sind. 15
   6.3. Stevens 1946: Die vier Skalenniveaus. 15
   6.4. Risikomatrizen sind Ordinalskalen – mit schlecht definierten Klassen. 16
   6.5. Warum arbeiten wir trotzdem mit Risikomatrizen? 17
   6.6. Das Skalenproblem im Detail: Extremwerte und schwere Verteilungsausläufer. . 18
   6.7. Es geht besser: Quantitative Methoden. 18
   6.8. Schätzen als Fertigkeit. 19
   6.9. Matrizen richtig einsetzen. 19
   6.10. Risikoappetit messbar formulieren. 20
7. Mythos 5: „GRC braucht GRC-Software.“ 20
   7.1. Das klassische Muster. 21
   7.2. Was wirklich gebraucht wird. 22
   7.3. Git als GRC-System. 22

7.4. Wann lohnt ein GRC-Tool? 23
8. Zusammenfassung 23
9. Literaturverzeichnis 24

Einführung: All models are wrong – but some are useful

George Edward Pelham Box formulierte 1976 in seiner Arbeit *Science and Statistics* einen der meistzitierten Sätze der angewandten Statistik: „All models are wrong – but some are useful." [Box 1976] Box beschrieb damit die grundlegende Spannung zwischen der Vereinfachung, die jedes Modell zwangsläufig darstellt, und dem pragmatischen Nutzen, den ein Modell trotzdem haben kann. Was für statistische Modelle gilt, gilt in gleichem Maße für das Risikomanagement: Kein Risikomodell bildet die Wirklichkeit vollständig ab. Die zentrale Frage ist daher nicht, ob ein Modell falsch ist – das ist jedes Modell per Definition – sondern ob es hinreichend nützlich ist, um bessere Entscheidungen zu ermöglichen.

Die Welt der Regulatorik und Standards

Regulatorische Rahmenwerke und Standards wie NIS-2, ISO 27001, BSI-Grundschutz und DORA fordern allesamt Risikomanagement. [NIS-2 2022; ISO 27001 2022; BSI 200-2 2017; DORA 2022] Sie beschreiben dabei das *Was* mit großer Konsequenz, lassen das *Wie* aber weitgehend offen. Spezialisierte Risikomanagement-Standards wie ISO 27005 und BSI-Standard 200-3 machen zwar einige methodische Vorschläge, zielen aber vor allem auf einen generischen Prozess ab: identifizieren, bewerten, behandeln, berichten. [ISO 27005 2022; BSI 200-3 2017] Ähnliches gilt für das NIST Risk Management Framework. [NIST RMF 2018] Die implizite Botschaft lautet: Es gibt die eine Methode – wende sie an. Die Praxis sieht anders aus. Die Anforderung steht präzise formuliert im Normtext; die Methodenwahl liegt vollständig beim Anwender. Standards setzen die Ziellinie. Den Weg dorthin muss man selbst kennen.

Risikoanalyse unterstützt Entscheidungsfindung

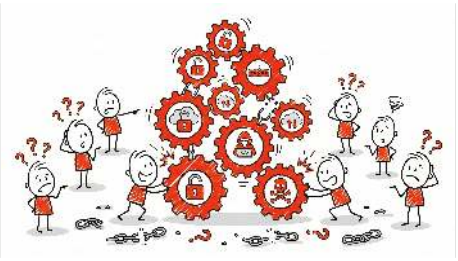
Eine Risikoanalyse ist kein Selbstzweck und kein Instrument der Compliance-Hygiene. Sie dient als strukturierte Eingabe für drei klar abgrenzbare Entscheidungsaufgaben. Erstens: Einen strukturierten Beitrag zu Entscheidungen zu liefern – innerhalb der Gewichtung, die das Management zwischen Risiko, wirtschaftlichen, strategischen und operativen Faktoren vorgibt. Im besten Fall lässt sich diese Gewichtung durch belastbare Zahlen noch beeinflussen. Zweitens: Maßnahmen zur Risikoreduktion systematisch zu bewerten, auszuwählen und umzusetzen. Drittens: Restrisiken bewusst, nachvollziehbar und doku-

mentiert zu akzeptieren. Ein Risikoregister, das keiner dieser drei Aufgaben dient, ist Compliance-Theater. [Howard 1966; Hubbard & Seiersen 2023]

Value of Information

Nicht jede Vertiefung einer Risikoanalyse ist ökonomisch sinnvoll. Das Konzept des Value of Information, das auf Arbeiten von Ronald Howard zurückgeht [Howard 1966], bietet einen einfachen Entscheidungstest: Kann eine präzisere Analyse die zu treffende Entscheidung noch verändern? Wenn ja, ist der zusätzliche Aufwand gerechtfertigt. Wenn nein, sind die Ressourcen besser anderweitig eingesetzt. Douglas Hubbard und Richard Seiersen demonstrieren in *How to Measure Anything in Cybersecurity Risk*, dass in der IT-Sicherheitspraxis viele Analysen weit über den Punkt hinausgetrieben werden, an dem zusätzliche Präzision noch irgendetwas am Ergebnis ändert. [Hubbard & Seiersen 2023] Das Ziel ist eine hinreichend gute Risikoanalyse – nicht eine perfekte.

Mythos 1: „Wir reden alle über Risiken."



Terminologische Verwirrung ist im Risikomanagement kein Randproblem – sie ist systematisch. Seit dem Turmbau zu Babel reden Menschen aneinander vorbei; im Risikomanagement hat man daraus eine Disziplin gemacht. Wer in einem typischen Risikomanagement-Workshop zehn erfahrene Fachleute zu ihrer Arbeitsdefinition von „Risiko" befragt, erhält bis zu zwölf verschiedene Antworten. Drei meinen die Eintrittswahrscheinlichkeit. Zwei meinen den erwarteten Schaden. Fünf meinen irgendetwas dazwischen oder etwas ganz anderes. Alle nicken. Keiner hinterfragt es. Alle reden aneinander vorbei. [Renn 2008]

Ein Workshop. Zehn Leute.

Das Szenario ist kein Gedankenexperiment, sondern gelebte Praxis in nahezu jeder Organisation, die mit dem Aufbau eines Risikomanagementsystems beginnt. Der Grund liegt in der Vieldeutigkeit des Begriffs selbst: „Risiko" hat in unterschiedlichen wissenschaftlichen Disziplinen, Normungstexten und Industriestandards unterschiedliche, teils inkompatible Bedeutungen. Wer im Workshop nicht explizit klärt, welches Verständnis verbindlich ist, produziert Ergebnisse, die nicht belastbar, nicht vergleichbar und im Zweifelsfall nicht auditierfähig sind. [ISO 31000 2018]

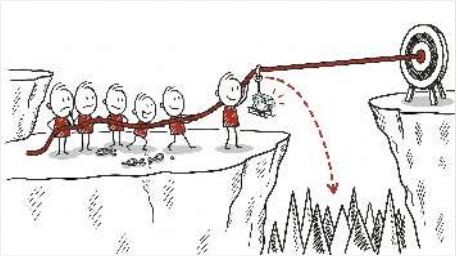
Was fehlt, ist ein transparentes Terminologiemanagement – und das ist keine akademische Forderung, sondern eine operative Notwendigkeit. Jede Organisation, die Risiken systema-

tisch steuern will, braucht einen explizit dokumentierten und kommunizierten Begriffsrahmen. Dieser Rahmen muss mit einer einzigen, aber zentralen Frage beginnen: Was meinen wir, wenn wir „Risiko“ sagen? Erst wenn diese Frage verbindlich beantwortet ist, lassen sich alle weiteren Begriffe – Bedrohung, Schwachstelle, Schadenshöhe, Akzeptanzschwelle – konsistent darauf aufbauen. [Renn 2008; ISO 31000 2018]

Drei Frameworks. Drei Sprachen. Null Konsens.

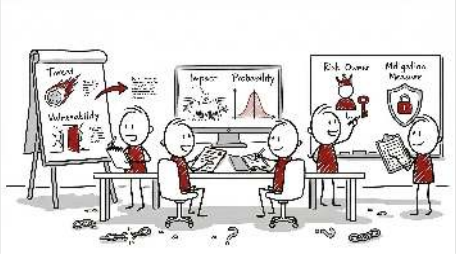
Drei wichtige Risiko-Frameworks der IT-Sicherheit sprechen grundlegend verschiedene Sprachen. ISO 31000:2018 definiert Risiko als die Auswirkung von Ungewissheit auf Ziele – ein zielorientierter, kontextabhängiger Ansatz, der positive wie negative Abweichungen einschließt. [ISO 31000 2018] NIST SP 800-30 operiert mit dem Produkt aus Bedrohung, Schwachstelle und Schadensauswirkung – ein technisch-analytischer Ansatz mit klarem Infosec-Bezug. [NIST SP 800-30 2012] FAIR (Factor Analysis of Information Risk) quantifiziert Risiko als erwarteten Jahresverlust, ausgedrückt als Produkt aus Verlustereignisfrequenz und Verlusthöhe. [Jones & Freund 2014; Jones & Freund 2026] Alle drei Ansätze sind in ihrem jeweiligen Kontext methodisch vertretbar. Werden sie undeklariert im selben Raum verwendet, ist das Ergebnis toxisch: Scheindebatten über Bewertungsergebnisse, die schlicht nicht auf derselben Grundlage basieren.

Schritt 1: Einigt euch auf eine Definition.



Als organisationsweit verbindliche Basisdefinition empfiehlt sich die Formulierung aus ISO 31000:2018: *Risiko ist die Auswirkung von Ungewissheit auf Ziele.* [ISO 31000 2018] Diese Definition ist zielorientiert statt rein technisch, gilt für Organisationen aller Größen und Branchen und ist offen genug, um sowohl negative als auch positive Abweichungen zu umfassen. Entscheidend ist, dass diese Definition *vor* dem ersten Risiko-Workshop für alle Beteiligten verbindlich festgelegt wird – nicht als nachträgliche Korrektur eines bereits laufenden Prozesses.

Schritt 2: Den Rest für alle festlegen.



Auf die Kernbegriffsdefinition muss zwingend die Einigung auf alle abgeleiteten Termini folgen: Bedrohung, Schwachstelle, Schadenshöhe, Eintrittswahrscheinlichkeit, Risikoakzeptanz, Restrisiko, Kontrollziel, Maßnahme und Risikoeigner. Diese Terminologie muss dokumentiert, in die Sprache der Organisation integriert und für alle verbindlich kommuniziert werden. Der Aufwand für ein konsistentes Glossar verhindert Jahre inkonsistenter Risikoregister. Wer diesen Schritt überspringen will, muss beim Auditor den Kopf hinhalten. [Renn 2008; ISO 31000 2018]

Fail: „Du hättest das wissen müssen.“

Nach einem erfolgreichen Phishing-Vorfall gibt es typischerweise zwei organisatorische Reflexe. Der erste: Pflicht-Awareness-Training für alle Mitarbeitenden. Der zweite: persönliche Konsequenzen für die Person, die geklickt hat. Beide Reaktionen gehen am Problem vorbei – aber aus unterschiedlichen Gründen.

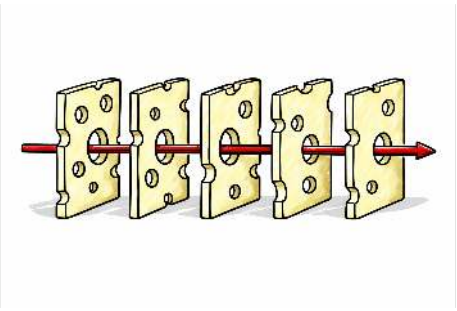
Das Pflicht-Training für alle ist eine zusätzliche Belastung für alle Mitarbeitenden, die nichts geklickt haben. Vor allem adressiert es das falsche Problem. Social Engineering funktioniert durch systematische Ausnutzung kognitiver Mechanismen – Zeitdruck, Autorität, soziale Konformität – die bei jedem Menschen wirken, unabhängig von Wissenstand und Erfahrung. Wer einmal versteht, wie Phishing tatsächlich funktioniert, hört auf zu fragen, warum jemand geklickt hat, und fängt an zu fragen, warum die systemischen Schutzmaßnahmen versagt haben. [Hadnagy 2011]

Die persönlichen Konsequenzen für die handelnde Person sind das zweite und gefährlichere Problem. Wer nach einem Klick mit negativen Konsequenzen rechnen muss, wird den Klick nicht melden. Wer nicht meldet, verhindert die Incident Response. Wer die Incident Response verhindert, maximiert den Schaden. Die implizite Botschaft *Du hättest das wissen müssen* entlastet das System und bestraft die Person – und macht die Organisation dabei unsicherer. [Dekker 2007]

Awareness hat zwei Ziele.

Awareness-Programme verfolgen typischerweise nur das erste von zwei notwendigen Zielen: *Nicht klicken* – also Phishing erkennen und richtig handeln. Das zweite Ziel ist mindestens ebenso kritisch: *Sofort melden*, wenn doch geklickt wurde – und zwar ohne Angst vor Konsequenzen. Ein Awareness-Programm, das ausschließlich auf Prävention setzt, vernachlässigt die Reaktionsfähigkeit der Organisation im Schadenfall. Die Reaktion auf einen Sicherheitsvorfall entscheidet über den tatsächlichen Schaden – nicht der Vorfall selbst.

Das Swiss-Cheese-Modell.



James Reason beschrieb 1990 das Swiss-Cheese-Modell als konzeptionellen Rahmen für das Verständnis komplexer Systemversagen. [Reason 1990] Mehrere gestaffelte Schutzschichten mit unterschiedlichen Lücken verhindern gemeinsam, dass ein Fehler zum Schaden wird. Auf einen Phishing-Angriff übertragen: Vor dem Klick wirken Spam-Filter, URL-Scanner und Awareness-Training. Nach dem Klick schützen EDR-Systeme vor der Ausführung von Schadsoftware. Eine funktionierende Meldekultur startet die Incident Response. SIEM und Monitoring erkennen auffälliges Verhalten im Netz.

Der Mensch ist eine Schutzschicht – eine wichtige –, aber nicht die letzte. Wer ihn so behandelt, als wäre er die einzige, baut eine defensive Architektur auf einem einzelnen Punkt auf. Das Swiss-Cheese-Modell macht damit auch eine Aussage über Risikoreduktion: Jede zusätzliche Schutzschicht reduziert die Wahrscheinlichkeit, dass alle Lücken gleichzeitig ausgerichtet sind. Risiko wird nicht durch einen einzelnen perfekten Kontrollmechanismus minimiert, sondern durch die Kombination mehrerer unabhängiger, imperfekter Schichten.

Just Culture.

Sidney Dekker prägte den Begriff der Just Culture als Gegenentwurf zur Schuld- und Strafkultur in sicherheitskritischen Systemen. [Dekker 2007] In der Luftfahrt gilt das Prinzip seit Jahrzehnten: Wer einen Fehler meldet, ist geschützt. Systeme und Prozesse werden bewertet, nicht Personen. Das Ziel ist organisatorisches Lernen. Die EASA hat dieses Prinzip in ihre Safety Management-Vorgaben verankert. [EASA 2014] IT-Security steht diesem Prinzip häufig diametral entgegen. Ein Klick auf einen Phishing-Link gilt als persönliches Versagen, das Konsequenzen nach sich zieht – also schweigt man lieber. Awareness wird besser, wenn Melden sicher ist. Haltung und Vertrauen sind dabei keine weichen Faktoren, sondern messbare risikoreduzierende Größen: Eine Organisation, in der Mitarbeitende darauf vertrauen, dass Melden keine negativen Konsequenzen hat, erfährt Vorfälle früher, reagiert schneller und begrenzt den Schaden wirksamer als eine Organisation, die auf Kontrolle und Konsequenz setzt.

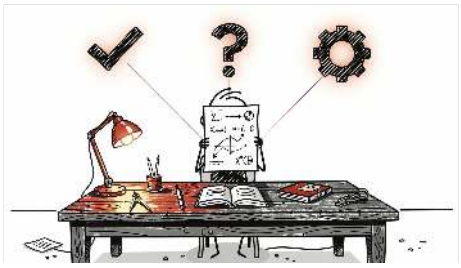
Mythos 2: „Ich kann keine Wahrscheinlichkeit angeben – ich habe keine Daten.“

„Ich kann keine Wahrscheinlichkeit angeben – ich habe keine Daten.“ Dieser Satz fällt in nahezu jedem ISMS-Workshop, irgendwann. Er klingt methodisch vorsichtig. Er ist es nicht. Das dahinterliegende Missverständnis ist, dass Wahrscheinlichkeit Massendaten voraussetzt – also eine frequentistische Interpretation: Wahrscheinlichkeit als relative Häufigkeit beobachteter Ereignisse. Wer keine Statistik über eingetretene Ransomware-Angriffe im eigenen Unternehmen hat, kann sich angeblich nicht äußern. Tatsächlich tut dieselbe Person es trotzdem – nur ohne es zuzugeben. Denn wer ein Risiko als „hoch“, „mittel“ oder „niedrig“ einstuft, gibt eine Wahrscheinlichkeitsaussage ab. Eine grobe, implizite, undokumentierte – aber eine. Eine qualitative Risikobewertung ist keine Alternative zur Quantifizierung. Sie ist eine besonders grobe Quantifizierung.

Bertrand Russell · 1929.

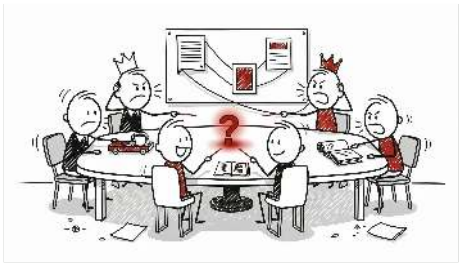
Russells Bonmot „Probability is the most important concept in modern science, especially as nobody has the slightest notion what it means." hat bis heute Bestand, weil er auf ein strukturelles Problem zeigt: Der Begriff „Wahrscheinlichkeit" wird in Wissenschaft und Praxis mit stark unterschiedlichen Bedeutungen belegt. Was in einem Kontext eine frequentistische Häufigkeit ist, ist in einem anderen ein Überzeugungsgrad, in einem dritten eine physikalische Eigenschaft von Systemen. Wer im Risikomanagement mit Wahrscheinlichkeitsangaben arbeitet, ohne sich über das zugrundeliegende Konzept im Klaren zu sein, produziert Zahlen, die formal konsistent, inhaltlich aber bedeutungslos sind. Das war 1929 ein Problem. Es ist heute eines. [Salmon 1966]

Drei Adäquatheitskriterien.



Wesley Salmon arbeitete 1966 drei Kriterien heraus, anhand derer eine Wahrscheinlichkeitsdefinition beurteilt werden kann. [Salmon 1966] Erstens die Admissibilität: Erfüllt die Definition die Kolmogorov-Axiome, also Nichtnegativität, Normierung und Additivität? Zweitens die Bestimmbarkeit: Kann der Wert der Wahrscheinlichkeit prinzipiell ermittelt werden? Drittens die Anwendbarkeit: Taugt die Wahrscheinlichkeitsangabe als Handlungsgrundlage? Für das IT-Sicherheits-Risikomanagement ist das dritte Kriterium das entscheidende – und es ist dasjenige, das die gebräuchlichsten Konzepte oft nicht erfüllen.

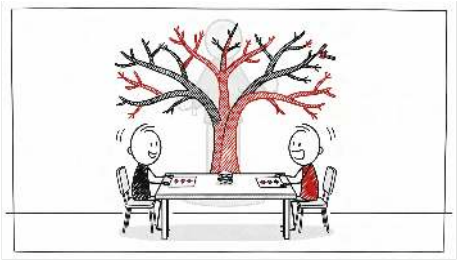
Sechs Interpretationen für einen Begriff.



Die Wissenschaftsphilosophie unterscheidet sechs etablierte Interpretationen des Wahrscheinlichkeitsbegriffs. [Salmon 1966; Hájek 2012] Die *klassische* Interpretation nach Laplace definiert Wahrscheinlichkeit als Verhältnis günstiger zu gleichmöglichen Fällen – sie setzt eine symmetrische, abzählbare Grundgesamtheit voraus und funktioniert für Würfel und Münzen, nicht für Bedrohungsszenarien. Die *frequentistische* Interpretation nach von Mises definiert Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit in einer unendlich langen Versuchsreihe – sie setzt Wiederholbarkeit voraus und scheitert an Einzelfallereignissen. Die *logisch-evidenzielle* Interpretation nach Carnap behandelt Wahrscheinlichkeit als objektive logische Beziehung zwischen verfügbarer Evidenz und einer Hypothese – sie ist beobachterunabhängig, aber schwer zu operationalisieren. Die *Best-System*-Interpretation nach Lewis versteht Wahrscheinlichkeit als Teil des besten wissenschaftlichen Systems zur Beschreibung der Welt – erkenntnistheoretisch interessant, praktisch kaum handhabbar. Die *Propensitäts*-Interpretation nach Popper beschreibt Wahrscheinlichkeit als objektive Tendenz eines physikalischen Systems – sinnvoll für Zerfallsraten, nicht für

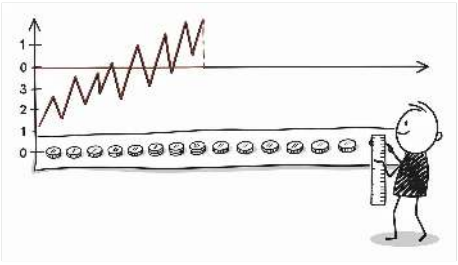
menschliches Angriffsverhalten. Die *subjektive* oder *bayesianische* Interpretation nach de Finetti und Savage fasst Wahrscheinlichkeit als Überzeugungsgrad eines rationalen Akteurs auf – als das beste informierte Urteil unter explizit gemachten Annahmen. Für IT-Sicherheitsrisiken, die typischerweise Einzelfallereignisse ohne belastbare Wiederholungsdaten sind, kommt praktisch nur diese letzte Interpretation in Frage – auch wenn sich das zunächst unbequem anfühlt.

Klassische Wahrscheinlichkeit: Symmetrie als Fundament.



Die klassische Wahrscheinlichkeit nach Laplace definiert die Wahrscheinlichkeit eines Ereignisses als das Verhältnis der Anzahl günstiger Fälle zur Gesamtanzahl gleichmöglicher Fälle. Dieses Konzept ist mathematisch elegant und praktisch anwendbar, solange die Menge der möglichen Elementarereignisse abzählbar und symmetrisch verteilt ist – wie bei einem fairen Würfel oder einer idealen Münze. Im IT-Sicherheits-Risikomanagement liegen diese Bedingungen fast nie vor. Bei einem Cyberangriff gibt es keine abzählbare Menge symmetrischer Elementarereignisse. Das Laplace-Modell ist für Spielkasinos konzipiert, nicht für Bedrohungsszenarien. [von Mises 1957]

Frequentistische Wahrscheinlichkeit: Wiederholung als Voraussetzung.

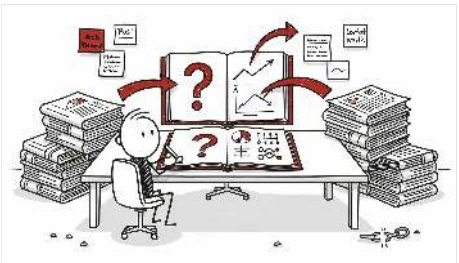


Das mathematische Fundament des Frequentismus legte Jakob Bernoulli mit dem Gesetz der großen Zahlen (*Ars Conjectandi*, 1713): Relative Häufigkeiten konvergieren mit wachsender Versuchsanzahl gegen einen stabilen Grenzwert. Zur erkenntnistheoretischen Interpretation ausgebaut hat dieses Ergebnis erst Richard von Mises, der Wahrscheinlichkeit formal als den Grenzwert der relativen Häufigkeit in einer unendlich langen Versuchsreihe unter gleichen Bedingungen definierte. [von Mises 1957] Das Konzept hat seinen berechtigten Platz überall dort, wo belastbare Wiederholungsdaten vorliegen: Versicherungsstatistiken, Hardware-Ausfallraten, klinische Studien. Der Aussage, dass 23 Prozent aller ungeschulten Mitarbeiter auf Phishing-Mails klicken, liegt eine große Stichprobe vergleichbarer Fälle zugrunde und ist frequentistisch valide. Die Frage jedoch, mit welcher Wahrscheinlichkeit *dieses* Unternehmen *dieses Jahr* von *dieser* APT-Gruppe angegriffen wird, ist eine Frage nach einem Einzelfall – und damit frequentistisch prinzipiell nicht beantwortbar. [Hubbard & Seiersen 2023]

Subjektive Wahrscheinlichkeit: Überzeugungsgrad als Ausgangspunkt.

Bruno de Finetti und Leonard Savage entwickelten in der Mitte des 20. Jahrhunderts die subjektive Interpretation der Wahrscheinlichkeit als Überzeugungsgrad eines rationalen Akteurs. [de Finetti 1974; Savage 1954] Operationalisierbar ist dieser Überzeugungsgrad über das Wettparadigma: Die Wahrscheinlichkeit, die eine Person einem Ereignis zuschreibt, entspricht dem maximalen Wetteinsatz, den sie bereit wäre zu zahlen – für eine Auszahlung von 1 im Eintrittsfall. Das Dutch-Book-Argument zeigt, dass widersprüchliche Wahrscheinlichkeitsurteile systematisch ausgenutzt werden können; rationale Akteure sind daher gezwungen, ihre Einschätzungen untereinander kohärent zu halten. Subjektiv bedeutet dabei nicht: Bauchgefühl ohne Grundlage. Es bedeutet: das beste verfügbare, informierte Urteil unter explizit gemachten Annahmen.

Bayes' Theorem: Vom Prior zur belastbaren Abschätzung.

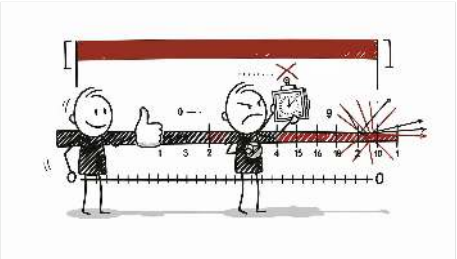


Das bayesianische Aktualisierungsverfahren bietet den methodischen Rahmen, subjektive Startschätzungen schrittweise zu belastbaren Einschätzungen zu verfeinern. Die Grundformel lautet:

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

Dabei ist $P(H)$ der Prior – der Überzeugungsgrad für Hypothese H vor Berücksichtigung neuer Evidenz. $P(E \mid H)$ ist die Likelihood – die Wahrscheinlichkeit, die Evidenz E zu beobachten, wenn H gilt. $P(E)$ ist die Gesamtwahrscheinlichkeit der Evidenz über alle Hypothesen. $P(H \mid E)$ ist der Posterior – der revidierte Überzeugungsgrad nach Berücksichtigung der Evidenz. Ausgangspunkt ist der Prior: „APT-Angriffe auf unsere Branche sind selten, ich schätze 5 Prozent." Neue Evidenz – ein Branchenreport mit drei betroffenen Unternehmen von zehn, ein EDR-Alarm der letzten Woche – fließt als Likelihood in die Berechnung ein. Das Ergebnis ist der Posterior: eine revidierte Einschätzung von über 35 Prozent. [Jaynes 2003; Hubbard & Seiersen 2023] Dieser Prozess ist explizit, nachvollziehbar und revidierbar. Er kombiniert Bauchgefühl und Evidenz – und das ist genau das, was Risikomanagement unter Unsicherheit leisten muss.

Imprecise Probabilities: Intervall statt Scheinsicherheit.



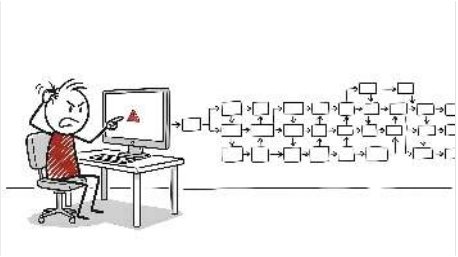
Zwei identische Wahrscheinlichkeitswerte können epistemisch völlig unterschiedliches Gewicht tragen. Eine perfekt getestete Münze hat P(Kopf) = 0,5 auf der Grundlage tausender kontrollierter Würfe – ein schmaler, robuster Schätzwert. Eine auf der Straße gefundene Münze hat ebenfalls P(Kopf) = 0,5, weil schlicht keine Information vorliegt – eine Schätzung aus purer Ignoranz. Der numerische Wert ist identisch; das Wissen dahinter ist es nicht. Wer beide gleich behandelt, macht einen epistemischen Fehler. Peter Walley entwickelte in seiner Theorie der Imprecise Probabilities einen formalen Rahmen für genau diese Situationen: Statt einer Punktschätzung P(A) = 0,12 erlaubt das Konzept Intervallangaben der Form $P(A) \in [0{,}05;\ 0{,}30]$ – den Bereich, über den vernünftige Fachleute uneinig sein dürfen, ohne dass einer von ihnen falsch liegt. [Walley 1991] Im Risikomanagement bedeutet das: „10 Prozent" auf Basis einer fundierten Analyse mit mehreren Evidenzquellen ist etwas grundlegend anderes als „10 Prozent" ohne jede Grundlage – auch wenn die Zahl dieselbe ist. Wer für den gesamten plausibelsten Bereich plant, braucht keine falsche Präzision.

Mythos 3: „One method fits all."

„Wer ist für das Risikomanagement zuständig?" Die Antwort lautet: Kommt drauf an. Risiken in der Informationssicherheit werden immer in Bezug auf ein soziotechnisches System bewertet – eine spezifische Kombination aus Menschen, Prozessen und Technologie in einem bestimmten Kontext. Was für ein kleines, spezifisches soziotechnisches System ein inakzeptables Risiko darstellt, kann für ein größeres, umfassenderes soziotechnisches System durchaus tolerierbar sein: Ein einzelner Server mit bekannter Schwachstelle ist für den zuständigen Administrator ein dringendes Problem; für das Gesamtunternehmen, das über Kompensationsmaßnahmen und Priorisierungsentscheidungen verfügt, möglicherweise eines von hundert. Das Ebenen-Modell überträgt diesen pragmatischen Umgang mit Risiken auf Organisationen: Es macht explizit, auf welcher Systemebene ein Risiko bewertet wird, wer dafür zuständig ist und welche Methode angemessen ist. Jede Ebene hat dabei ihren eigenen Prozess und eigene Dokumentationsanforderungen: Die operative Ebene arbeitet mit Checklisten, Ticketsystemen und direkten Behebungsmaßnahmen. Die taktische Ebene aggregiert, priorisiert und dokumentiert in Risikoregistern mit definierten Behandlungsfristen. Die strategische Ebene entscheidet und dokumentiert in Risikoberichten, die Leitungsebene und ggf. Aufsichtsgremien adressieren. Risiken können und müssen zwischen Ebenen eskalieren – nämlich dann, wenn ein Risiko die Ressourcen, Kompetenzen oder Entscheidungsbefugnisse der aktuellen Ebene übersteigt. Ein Administrator, der eine Schwachstelle nicht ohne Budgetfreigabe schließen kann, eskaliert auf die taktische Ebene. Ein IT-Leiter, der ein strategisches Reputationsrisiko

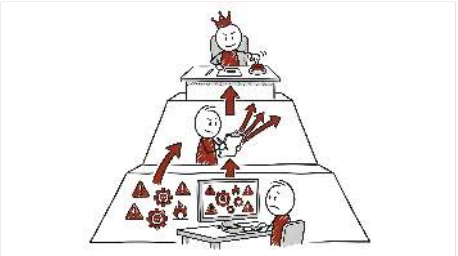
erkennt, eskaliert auf die Managementebene. Funktioniert diese Eskalation nicht, bleiben Risiken auf der falschen Ebene stecken – und werden dort entweder ignoriert oder falsch behandelt. [ISO 31000 2018; ICAO SMM 2018; Hollnagel 2014]

Ein Admin. Ein Risiko. Ein falscher Weg.



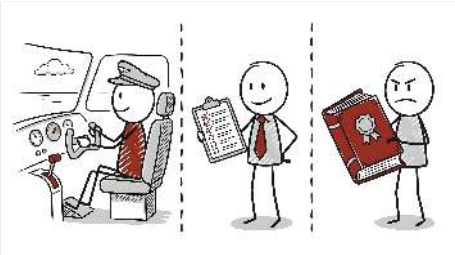
Ein konkretes Beispiel verdeutlicht die Dysfunktion: Ein Systemadministrator identifiziert eine kritische Fehlkonfiguration – technisch behebbar innerhalb von zehn Minuten, ohne Systemausfall, ohne nennenswertes Aufwandsrisiko. Der gelebte Prozess lässt jedoch keine Abkürzung zu: Sicherheitsrelevante Feststellungen werden ins ISMS überführt. Also entsteht ein Ticket. Das Risiko wird formal beschrieben, klassifiziert und nach Eintrittswahrscheinlichkeit sowie Schadenpotenzial bewertet. Ein Genehmigungsworkflow wird angestoßen, Verantwortliche werden notifiziert. Die Feststellung wandert ins Risikoregister – wo sie fortan neben Risiken mit Millionenhaftung und Reputationsfolgen wartet. Das nächste planmäßige Review findet in drei Monaten statt. Die Fehlkonfiguration bleibt in dieser Zeit offen. Das ISMS ist nicht das Problem – es erfüllt exakt den Zweck, für den es konzipiert wurde. Das Problem ist der kategorische Einsatz eines strategischen Steuerungsinstruments für ein operatives Einzelrisiko. Methode und Risikotyp passen nicht zusammen; das Ergebnis ist kein Sicherheitsgewinn, sondern organisierte Verzögerung. [ISO 31000 2018]

Drei Ebenen. Drei Aufgaben.



Risikomanagement vollzieht sich oft auf strukturell verschiedenen Ebenen. Auf der operativen Ebene erkennt und behebt der Systemadministrator konkrete, tägliche Risiken direkt – mit einem Zeithorizont von Stunden bis Tagen. Auf der taktischen Ebene bündelt, priorisiert und eskaliert der IT-Leiter operative Risiken, die das Individuum nicht alleine trägt – mit einem Zeithorizont von Wochen bis Monaten. Auf der strategischen Ebene entscheidet und verantwortet das Management große, seltene Risiken mit Relevanz für Haftung, Reputation und Unternehmensziele – mit einem Zeithorizont von Quartalen bis Jahren. Drei Ebenen sind dabei die praktische Untergrenze – in sehr großen oder stark regulierten Organisationen entstehen weitere Zwischenschichten, etwa eine dedizierte Compliance-Ebene oder ein separates Programm-Management. Entscheidend ist nicht die genaue Zahl, sondern das Prinzip: Jede Ebene braucht ihre eigene Methode, ihre eigene Sprache und ihre eigene klare Verantwortung – und zwischen den Ebenen braucht es pragmatische Eskalationswege, die im Alltag tatsächlich funktionieren. [ISO 31000 2018; Hollnagel 2014]

Vergleich: Luftfahrt.

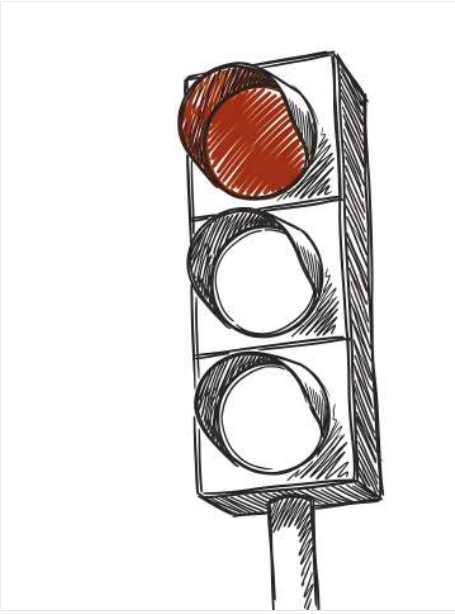


Die Luftfahrt strukturiert ihr Sicherheits- und Risikomanagement seit Jahrzehnten nach exakt diesem dreistufigen Prinzip und hat es in internationale Normen gegossen. Der Pilot im Cockpit betreibt operatives Risikomanagement in Echtzeit: Wetter, Treibstoff, Ausweichoptionen. Das Safety Management System der Airline aggregiert Betriebserfahrungen zu taktischen Mustern: Wartungsrhythmen, Schulungsplanung, Prozessverbesserungen. EASA, LBA und ICAO setzen den systemischen Rahmen durch strategische Regulierung und Unfallanalyse. [ICAO SMM 2018; EASA 2014] Der Pilot entscheidet nicht über Vorschriften. Die EASA fliegt nicht. Jede Ebene hat ihre Aufgabe – und nur ihre.

Die taktische Lücke.

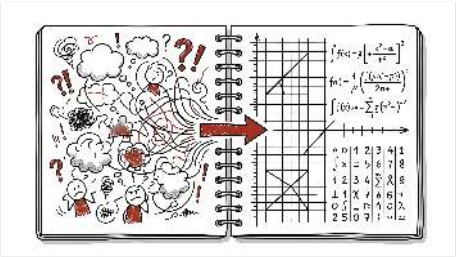
Das häufigste und folgenreichste Versagen in der Praxis ist das strukturelle Fehlen der taktischen Ebene. Operative Risiken stapeln sich beim Systemadministrator, bis er sie nicht mehr alleine trägt. Ohne einen definierten Eskalationsweg verschwinden sie im Tagesgeschäft. Das strategische ISMS bleibt unberührt und formal sauber, weil die Probleme nie dort ankommen. Ohne taktische Ebene landet alles entweder beim Admin – der überfordert ist – oder nirgendwo. Jedes für sich wäre das einzelne operative Risiko handhabbar. Zusammen und ohne Eskalationsstruktur wird die Summe unbeherrschbar. [Hollnagel 2014]

Mythos 4: „Die Welt ist eine Ampel – Rot-Gelb-Grün erklärt alles.“



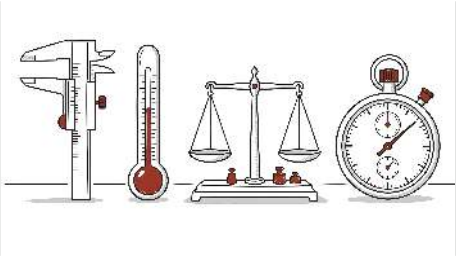
Risikomatrizen und Ampeldarstellungen sind in nahezu jeder Organisation anzutreffen: im Board-Reporting, im Auditbericht, im ISMS-Dashboard. Sie erscheinen intuitiv verständlich, kommunizierbar und normenkonform. Das Problem liegt tiefer als ein gelegentlicher Missbrauch – es ist methodisch strukturell. Die 5×5-Risikomatrix mit den Werten 1 bis 5 verwendet eine Ordinalskala, ohne dass dies den Beteiligten bewusst ist. Und auf Ordinalskalen sind die meisten Operationen, die routinemäßig angewendet werden, mathematisch schlicht unzulässig. [Cox 2008; Stevens 1946]

Von Begriff zu Maß: Risiko als Konzept – und als Zahl.



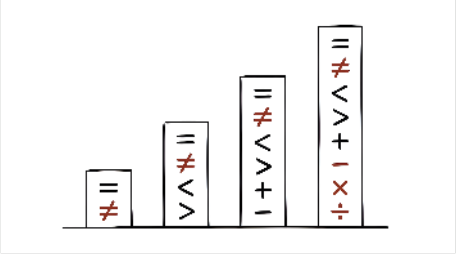
Risiko ist zunächst ein Begriff, der eine Eigenschaft von Entscheidungssituationen beschreibt: die Unsicherheit über mögliche Schäden oder Zielerreichungen. Als Begriff allein ist es jedoch nicht operationalisierbar. Sobald eine Norm fordert, Risiken zu identifizieren, zu bewerten und zu priorisieren, ist damit implizit gefordert, sie messbar zu machen – also in eine Zahl oder Kategorie zu überführen. Jede Messung basiert auf einem Modell, das festlegt, welche Informationen erfasst werden und welche Operationen auf den Messergebnissen erlaubt sind. Wer Risiko bewertet, misst – ob er es weiß oder nicht. Dabei ist Messen mehr als das Erzeugen einer Zahl: Es ist das systematische Reduzieren von Unsicherheit durch Information. Douglas Hubbard definiert Messung präzise als „a quantitatively expressed reduction of uncertainty based on one or more observations" – nicht als exakte Bestimmung eines wahren Werts, sondern als Annäherung, die Entscheidungen verbessert. Ein Risikowert muss also nicht perfekt sein, um nützlich zu sein; er muss nur die verbleibende Unsicherheit sichtbarer machen als der Zustand ohne Messung. Daraus folgt eine produktive Konsequenz: Auch grobe, explizit unsichere Schätzungen sind legitime Messungen – solange ihre Grenzen transparent gemacht werden. Die Alternative, auf Messung zu verzichten, reduziert Unsicherheit nicht; sie versteckt sie lediglich in impliziten Annahmen. [Stevens 1946; Hubbard 2010]

Warum Skalen wichtig sind.



Stanley Smith Stevens legte 1946 in *On the Theory of Scales of Measurement* die theoretische Grundlage für die Klassifikation von Messskalen, die bis heute Bestand hat. [Stevens 1946] Messungen übersetzen Beobachtungen der Realität in Modelle, die Schlüsse erlauben. Das Modell ist nie die Realität selbst; es erlaubt lediglich bestimmte Schlussfolgerungen – und zwar nur jene, die das Skalenniveau auch trägt. Wer eine Operation anwendet, die das zugrundeliegende Skalenmodell nicht unterstützt, rechnet formal, aber fachlich falsch. Das Ergebnis sieht aus wie eine Antwort, ist aber keine. [Stevens 1946]

Stevens 1946: Die vier Skalenniveaus.



Stevens unterschied vier Skalenniveaus mit zunehmenden Aussagemöglichkeiten. [Stevens 1946]

Nominalskalen erfassen nur Kategorien – Gleichheit oder Unterschied, nicht mehr. Erlaubte Operationen: Häufigkeitsauszählung, Modus. Beispiel: Angriffsvektoren (Netzwerk, lokal, physisch). Die Zuweisung von Zahlen (1 = Netzwerk, 2 = lokal) ist reine Codierung; „2 ist doppelt so viel wie 1" ist sinnlos.

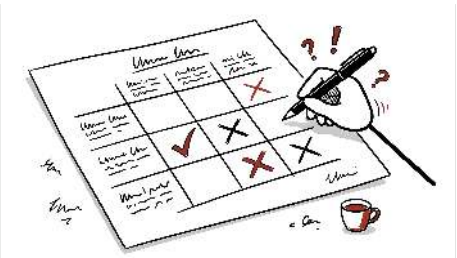
Ordinalskalen erlauben zusätzlich eine Reihenfolge, machen aber keine Aussage über die Größe der Abstände. Beispiel: Schulnoten, Reifegrade (CMMI Level 1–5), die Eintrittswahrscheinlichkeit „selten / möglich / wahrscheinlich“ in einer Risikomatrix, codiert als 1, 2, 3. Hier liegt der entscheidende Trugschluss: Die Zahlen sehen aus wie mathematische Objekte – sie sind es aber nicht. Der Abstand zwischen „selten“ und „möglich“ ist nicht notwendigerweise gleich dem Abstand zwischen „möglich“ und „wahrscheinlich“. Die Zahl 2 steht lediglich dafür, dass etwas mehr ist als 1 – wie viel mehr, ist nicht definiert. Addition und Multiplikation sind deshalb unzulässig: „möglich + wahrscheinlich“ ergibt keine interpretierbare Größe.

Intervallskalen erlauben Addition und Subtraktion, haben aber keinen natürlichen Nullpunkt – der Nullpunkt ist konventionell gesetzt. Beispiel: Temperatur in Grad Celsius. 0 °C bedeutet nicht „keine Temperatur“, und 20 °C ist nicht doppelt so warm wie 10 °C. Mittelwerte sind erlaubt, Verhältnisaussagen nicht.

Verhältnisskalen erlauben alle Rechenoperationen einschließlich Multiplikation und Division und besitzen einen absoluten Nullpunkt. Beispiel: Geldbetrag, Zeit, Wahrscheinlichkeit als echter Prozentwert. Nur hier ist „doppelt so hoch“ eine sinnvolle Aussage.

Die Konsequenz für das Risikomanagement ist unmittelbar: Die Formel Risiko = Eintrittswahrscheinlichkeit × Schadenshöhe ist nur dann mathematisch sinnvoll, wenn beide Faktoren auf Verhältnisskalen gemessen werden. Die Werte 1 bis 5 einer Risikomatrix sind Ordinalzahlen – sie erfüllen diese Bedingung nicht. Wer sie multipliziert, betreibt Arithmetik mit Etiketten.

Risikomatrizen sind Ordinalskalen – mit schlecht definierten Klassen.



In einer 5×5-Risikomatrix sind die Werte 1 bis 5 Ränge, keine Maßzahlen. Der Abstand zwischen dem Rang 2 und dem Rang 3 ist nicht notwendigerweise gleich groß wie der Abstand zwischen Rang 4 und Rang 5. Zusätzlich sind die Klassengrenzen in der Praxis selten präzise definiert: Ab wann ist ein Risiko „hoch“ statt „mittel“? Die ehrliche Antwort lautet meistens: Es kommt auf den Kontext, das Szenario und den Angreifer an. Die Matrix hat dafür kein Feld. Trotzdem werden routinemäßig Durchschnitte über mehrere Bewerter gebildet, Eintrittswahrscheinlichkeit und Schadenshöhe multipliziert und Risikogesamtsummen über Kategorien aggregiert – alles Operationen, die Ordinalskalen nicht erlauben. [Cox 2008; Stevens 1946]

Drei konkrete Fehlertypen illustrieren das:

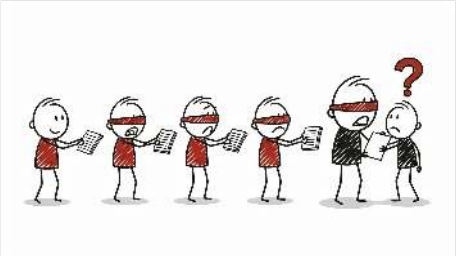
Informationsverlust durch Multiplikation. Ein Angriff mit sehr hoher Eintrittswahrscheinlichkeit und niedrigem Schaden (5 × 1 = 5) erhält denselben Risikowert wie ein seltener, aber katastrophaler Angriff mit niedrigem Schaden (1 × 5 = 5). Die Präventionsstrategie, der Handlungsdruck und die Gegenmaßnahmen sind völlig verschieden – die Matrix macht sie ununterscheidbar. Entscheidungen, die auf dieser Gleichsetzung basieren, sind strukturell oft falsch.

Grenzwert-Artefakte. Ein Risiko, das intern mit 2,9 bewertet wird, fällt in die Kategorie „mittel“ und landet im nachrangigen Backlog. Dasselbe Risiko, minimal anders eingeschätzt

mit 3,0, ist „hoch“ und löst einen Eskalationsprozess aus. Der Unterschied liegt nicht im Risiko selbst, sondern in einer willkürlich gezogenen Klassengrenze auf einer Ordinalskala. Cox zeigte, dass solche Grenzeffekte systematisch zu Fehlpriorisierungen führen – nicht als Ausnahme, sondern als inhärentes Matrixmerkmal.

Durchschnittsbildung über Bewerter. Drei Experten bewerten dieselbe Bedrohung mit 2, 3 und 5. Der rechnerische Durchschnitt ist 3,33 – „mittel“. Tatsächlich aber haben die drei Experten das Risiko fundamental unterschiedlich eingeschätzt: einer als gering, einer als mittel, einer als sehr hoch. Die Streuung ist die eigentliche Information. Der Durchschnitt vernichtet sie. Auf Ordinalskalen ist der Mittelwert kein statistisch zulässiges Maß; der Median wäre das korrekte Lagemaß – aber auch er verbirgt die Streuung. Nur die vollständige Verteilung (2 / 3 / 5) macht sichtbar, dass kein Konsens besteht. Mittelwerte über Ordinalskalen sind statistisch schlicht nicht zulässig.

Warum arbeiten wir trotzdem mit Risikomatrizen?



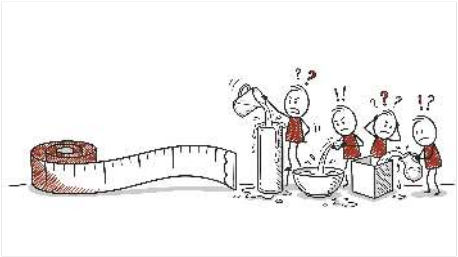
Die Gründe für die Verbreitung von Risikomatrizen trotz ihrer methodischen Schwächen sind gut dokumentiert – und sie sind, für sich genommen, durchaus nachvollziehbar. [Hubbard 2010; Cox 2008]

Das Management versteht Ampelfarben. Rot bedeutet: handeln. Grün bedeutet: entspannen. Eine Wahrscheinlichkeitsverteilung mit Konfidenzintervallen verlangt Abstraktionsvermögen, das in einem 30-Minuten-Steuerkreis schlicht nicht vorauszusetzen ist. Kommunikative Einfachheit ist kein Fehler – sie ist ein legitimes Ziel. Das Problem entsteht, wenn die vereinfachte Darstellung mit der Messung selbst verwechselt wird.

Auditoren und Zertifizierungsstellen akzeptieren Risikobewertungen in Matrixform routinemäßig, ohne das Skalenniveau zu hinterfragen. ISO 27001 fordert eine Risikobewertung – nicht zwingend eine methodisch valide. Wer eine ausgefüllte 5×5-Matrix vorlegt, besteht das Audit. Der Compliance-Druck belohnt Dokumentation, nicht Korrektheit.

Dazu kommt institutionelle Trägheit: Die Matrix steckt in den Vorlagen, in den Schulungsunterlagen, in den ISMS-Tools. Wer eine Alternative vorschlägt, muss zunächst erklären, warum das Bisherige falsch war – eine politisch unattraktive Position. Und schließlich, als ehrlichster Grund: Viele Praktikerinnen und Praktiker wissen nicht, dass robustere Methoden existieren, praxistauglich sind und keinen Doktortitel voraussetzen. Hubbard und Seiersen haben das überzeugend gezeigt. Erst wenn dieser Wissensstand sich ändert, werden die anderen Gründe verhandelbar.

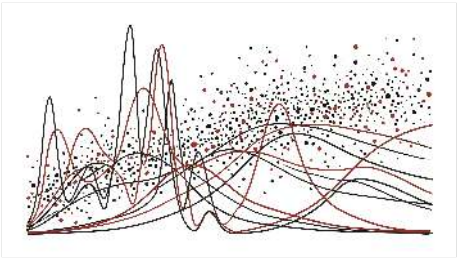
Das Skalenproblem im Detail: Extremwerte und schwere Verteilungsausläufer.



Ein besonders folgenreicher Aspekt des Skalenproblems betrifft die Verteilung von Schadenshöhen. Viele IT-Sicherheitsrisiken folgen keiner Normalverteilung, sondern Potenzgesetzen (Power Laws): Wenige, seltene Ereignisse verursachen den weitaus größten Teil des Gesamtschadens. Ransomware-Angriffe auf kritische Infrastruktur, Supply-Chain-Kompromittierungen oder Datenschutzvorfälle mit regulatorischen Folgen sind nicht „etwas größer" als ein durchschnittlicher Vorfall – sie liegen in einer anderen Größenordnung. In der Statistik spricht man von *heavy tails* – Verteilungen mit schweren Ausläufern: Die Kurve fällt nach rechts sehr langsam ab, Extremereignisse sind deutlich häufiger als eine Normalverteilung erwarten ließe.

Eine 5×5-Risikomatrix komprimiert diesen Ausläuferbereich in ein einziges Feld. Der Wert „5" bei der Schadenshöhe deckt alles ab: vom teuren Systemausfall bis zum existenzbedrohenden Totalschaden. Der Abstand zwischen einem Schaden von 100.000 Euro und einem von 100 Millionen Euro ist auf der Ordinalskala null – beide sind „5". Die Matrix ist damit strukturell blind für genau den Teil der Risikolandschaft, der die größte Entscheidungsrelevanz hat. Quantitative Methoden wie Monte-Carlo-Simulationen oder FAIR modellieren Schadenshöhen explizit als Verteilungen und machen den Ausläuferbereich sichtbar – statt ihn in einer Kategorie zu verbergen. [Hubbard & Seiersen 2023; Cox 2008]

Es geht besser: Quantitative Methoden.



Robustere Alternativen zur Ordinalskalen-basierten Risikobewertung existieren, sind praxistauglich und setzen keine akademische Statistikausbildung voraus.

Kalibriertes Schätzen. Hubbard und Seiersen zeigen in *How to Measure Anything in Cybersecurity Risk*, dass Expertenwissen systematisch in Wahrscheinlichkeitsaussagen überführt werden kann – auch ohne Massendaten. [Hubbard & Seiersen 2023] Statt „Eintrittswahrscheinlichkeit: hoch" lautet die Aussage: „Wir schätzen mit 80 % Konfidenz, dass dieses Ereignis innerhalb der nächsten 12 Monate einmal auftritt." Diese Formulierung ist überprüfbar, kommunizierbar und verbesserbar – ein ordinales Label ist es nicht.

Monte-Carlo-Simulation. Statt einzelner Punktschätzungen werden Eingangsgrößen als Verteilungen definiert: Die Schadenshöhe eines Ransomware-Angriffs liegt mit 90 % Wahrscheinlichkeit zwischen 200.000 und 4 Millionen Euro. Aus diesen Verteilungen werden Zehntausende Szenarien simuliert; das Ergebnis ist selbst eine Verteilung – mit Median, Konfidenzintervallen und sichtbarem Ausläuferbereich. [Hubbard 2010] Diese Methode ist nicht nur theoretisch robust, sondern auch praktisch umsetzbar: Es gibt zahlreiche Software-Tools, die Monte-Carlo-Simulationen mit benutzerfreundlichen Oberflächen ermöglichen.

FAIR (Factor Analysis of Information Risk). FAIR zerlegt Risiko in zwei Hauptkomponenten: die Häufigkeit eines Verlustereignisses (Loss Event Frequency) und die wahrscheinliche Schadenshöhe (Probable Loss Magnitude) – beide als Verteilungen modelliert. [Jones & Freund 2014; Jones & Freund 2026] Ein konkretes FAIR-Modell für einen Phishing-Angriff auf eine Finanzabteilung würde etwa Bedrohungshäufigkeit, Resistenz der Kontrollen und direkte sowie indirekte Schadenskomponenten (Forensik, Benachrichtigungspflichten, Reputation) separat schätzen und aggregieren. Das Ergebnis ist ein annualisierter Erwartungswert in Euro – vergleichbar, priorisierbar, kommunizierbar.

SHELF (Sheffield Elicitation Framework). Wo keine Daten vorliegen, sind Expertenurteile die primäre Informationsquelle – aber ungefilterte Expertenurteile sind anfällig für Ankereffekte, Überzeugungsstärke und Gruppendenken. SHELF strukturiert die Erhebung: Experten geben individuelle Einschätzungen ab, bevor sie diskutieren; ihre Aussagen werden als Wahrscheinlichkeitsverteilungen formalisiert und anschließend mathematisch aggregiert. [O'Hagan et al. 2006] Das Verfahren macht Meinungsverschiedenheiten sichtbar, statt sie im Durchschnitt zu verbergen.

Quantifizieren heißt nicht: perfekte Daten haben. Es heißt: Unsicherheit explizit machen statt implizit zu verstecken.

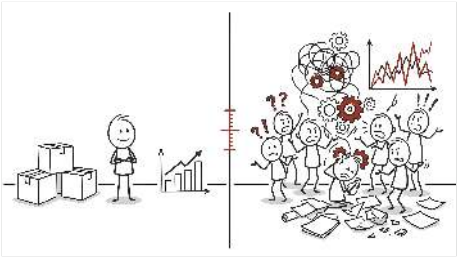
Schätzen als Fertigkeit.

Kalibriertes Schätzen ist eine lernbare Fertigkeit, keine Charaktereigenschaft. Wer behauptet, zu 70 Prozent sicher zu sein, sollte in etwa 70 Prozent der Fälle richtig liegen. Philip Tetlock und Dan Gardner zeigen in *Superforecasting*, dass Personen mit Training messbar besser kalibrierte Vorhersagen treffen als Experten ohne Kalibrierungspraxis. [Tetlock & Gardner 2015] Hubbards Kalibrierungsübungen mit 90-Prozent-Konfidenzintervallen belegen, dass die meisten Menschen ohne Training systematisch überkonfident sind – und dass sich das in wenigen Stunden gezielt korrigieren lässt. [Hubbard 2010] Bauchgefühl ist der legitime Ausgangspunkt, nicht das Ergebnis.

Matrizen richtig einsetzen.

Risikomatrizen sind keine schlechten Werkzeuge – sie werden häufig nur falsch eingesetzt. Die entscheidende Reihenfolge: zuerst quantifizieren, dann vereinfacht kommunizieren. Eine 5×5-Ampeldarstellung, die auf einer FAIR-Analyse oder einer Monte-Carlo-Simulation basiert, ist ein legitimes Kommunikationswerkzeug. Eine 5×5-Ampeldarstellung, die als Ersatz für eine strukturierte Analyse dient, ist Folklore. Was hinter der Ampel liegen muss, sind Wahrscheinlichkeitsverteilungen oder kalibrierte Intervallschätzungen mit expliziten, nachvollziehbaren Annahmen und Evidenzquellen. Die Ampel darf das Ergebnis zeigen – nicht das Fundament ersetzen. [Cox 2008; Hubbard & Seiersen 2023]

Risikoappetit messbar formulieren.



Risikoappetit ist in vielen Organisationen eine Sammlung wohlklingender, aber operativ wertloser Aussagen. „Wir tolerieren mittlere Risiken bei gleichzeitig geringem Risikoappetit im regulatorischen Umfeld“ – solche Formulierungen stehen in ISMS-Dokumenten, werden im Audit abgehakt und entfalten keinerlei Steuerungswirkung. Was ist mittel? Für wen? Ab welchem Schwellenwert? In welchem Zeitraum? Auf welcher Skala gemessen? Diese Fragen bleiben systematisch unbeantwortet. [Jones & Freund 2014; Jones & Freund 2026; NIST SP 800-30 2012]

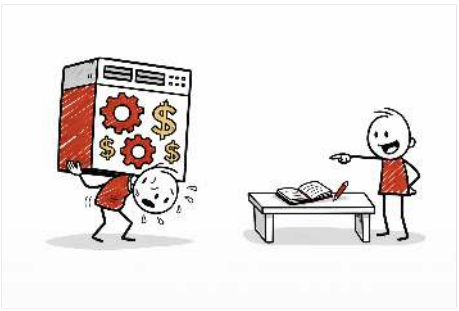
Operativ formulierter Risikoappetit benennt dagegen konkrete Schwellenwerte, die Entscheidungen auslösen und Verantwortung zuweisen. Zwei Beispiele aus der Praxis:

„Einzelrisiken mit einem annualisierten Erwartungsverlust unter 200.000 Euro werden auf operativer Ebene entschieden und dokumentiert. Risiken zwischen 200.000 und 2 Millionen Euro erfordern Freigabe durch den Bereichsleiter. Risiken über 2 Millionen Euro oder mit mehr als zehn Prozent Wahrscheinlichkeit auf eine regulatorische Sanktion eskalieren an den Vorstand.“ Diese Struktur ist anwendbar: Man weiß, wer entscheidet, wann eskaliert wird und was akzeptiert ist – ohne Interpretationsspielraum.

„Kein produktives System bleibt länger als 72 Stunden mit einem bekannten CVSS-Score ≥ 9.0 ungepatcht.“ Das ist kein Risikoappetit in Ampelfarben – das ist eine operative Vorgabe mit messbarem Erfüllungskriterium, die sich im nächsten Audit oder im nächsten Incident überprüfen lässt.

Der Unterschied ist grundlegend: Qualitative Risikoappetit-Aussagen beschreiben eine Haltung. Quantitative Schwellenwerte erzeugen Verantwortung. Nur letztere erlauben es, Risikoentscheidungen nachträglich zu prüfen – und damit auch, aus ihnen zu lernen. Appetit in Farben ist keine Steuerungsgröße. Es ist Dekoration.

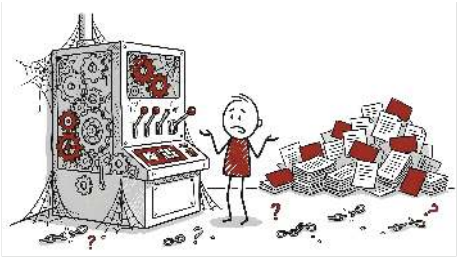
Mythos 5: „GRC braucht GRC-Software.“



GRC-Tools kosten oft fünfstellige Beträge pro Jahr oder mehr. Onboarding, Schulung, Beratung kommen hinzu. Und trotzdem sind in vielen Organisationen die Risiken nicht im Griff, das Risikoregister nicht aktuell und der nächste Audit eine Zitterpartie. Das Problem ist nicht das Tool. Es ist das Muster, das regelmäßig dahintersteht: Tool gekauft, Prozesse nicht definiert, Tool ungenutzt. Software kann keine Prozesse ersetzen, die nicht existieren. Sie kann ausschließlich Prozesse unterstützen, die bereits funktionieren. 50.000 Euro im Jahr kaufen kein Risikomanagement – sie kaufen ein leeres System, wenn der Prozess fehlt. [Racz et al. 2010]

Ein zweites, subtileres Problem tritt häufig hinzu: GRC-Tools diktieren eine Methodik. Sie sind auf ein bestimmtes Risikobewertungsmodell ausgelegt – meist eine 5×5-Matrix mit vordefinierten Kategorien, festem Workflow und starrem Berichtsformat. Wer das Tool kauft, kauft implizit auch diese Methodik, ob sie zur Organisation passt oder nicht. Ein mittelständisches Produktionsunternehmen mit überschaubarer IT-Landschaft hat andere Anforderungen als ein regulierter Finanzdienstleister mit hundert Anwendungen und drei Jurisdiktionen. Das Tool unterscheidet das nicht. Es bietet Felder an – und die Organisation lernt schnell, diese Felder zu befüllen, statt die richtigen Fragen zu stellen. Das Ergebnis ist ein gepflegtes System, das die Wirklichkeit nicht mehr abbildet, sondern ersetzt.

Das klassische Muster.



Das Beschaffungsmuster folgt in der Praxis einer vorhersehbaren Dramaturgie. [Racz et al. 2010]

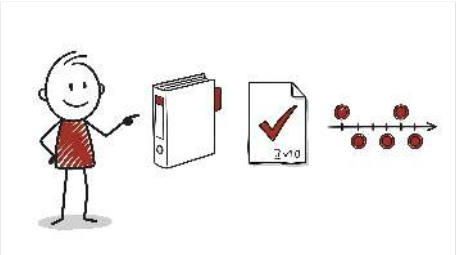
Schritt 1: Die Beschaffung. Nach einem Audit mit ernüchterndem Ergebnis oder auf Druck der Geschäftsführung wird ein Enterprise-GRC-Tool evaluiert. Die Demos überzeugen: Dashboards, Heatmaps, automatische Erinnerungen, Audit-Trails, ISO-27001-Mapping auf Knopfdruck. Der Vertrag wird unterschrieben – inklusive Onboarding-Paket und Beratungsleistung für die initiale Befüllung.

Schritt 2: Die Implementierung. Im Onboarding stellt sich heraus, dass grundlegende Prozessfragen nie geklärt wurden. Wer pflegt das Risikoregister – die IT, der CISO, die Fachabteilungen? Wer ist Risikoeigner für einen Geschäftsprozess, der über drei Abteilungen verläuft? Wer gibt Restrisiken frei, und nach welchen Kriterien? Wie häufig werden Risiken überprüft? Was passiert, wenn ein Risikoeigner das Unternehmen verlässt? Der Berater befüllt das Tool mit den Informationen, die er bekommt. Was er bekommt, ist unvollständig, inkonsistent und teilweise veraltet. Das Tool wird befüllt – nicht das Risikomanagement.

Schritt 3: Der Alltag. Sechs Monate nach Go-Live sind 80 Prozent der Einträge im Risikoregister unverändert. Die automatischen Erinnerungen zur quartalsweisen Überprüfung werden ignoriert oder mit „keine Änderungen“ quittiert. Das Dashboard zeigt grüne Ampeln – weil niemand rote einträgt, nicht weil keine roten existieren. Die Lizenzkosten laufen weiter.

Schritt 4: Der nächste Audit. Der Auditor fragt nach dem letzten Review-Zyklus. Einträge wurden aktualisiert – das stimmt formal. Ob die Einträge die tatsächliche Risikolage widerspiegeln, ist eine andere Frage. Oft lautet die ehrliche Antwort: nein. Das Tool ist nicht das Problem. Der fehlende Prozess ist es – und der fehlende Wille, Risikoeigner tatsächlich zur Verantwortung zu ziehen.

Was wirklich gebraucht wird.



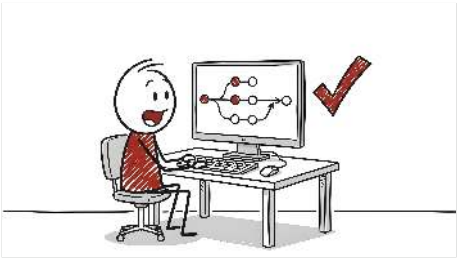
Jede GRC-Lösung – ob spezialisierte Software oder zusammengesetzte Werkzeugkette – muss drei Kernanforderungen erfüllen. [ISO 27001 2022; BSI 200-2 2017]

Risikoregister. Das Register ist das operative Herzstück. Es erfasst nicht nur den Risikotitel, sondern für jedes Risiko: den verantwortlichen Risikoinhaber (eine Person, keine Abteilung), die gewählte Behandlungsstrategie (Akzeptieren, Mitigieren, Transferieren, Vermeiden), den aktuellen Behandlungsstatus, das verbleibende Restrisiko nach Maßnahmen – und dokumentiert, wer das Restrisiko zu welchem Zeitpunkt freigegeben hat. Letzteres ist entscheidend: Ohne nachvollziehbare Freigabe ist das Register eine Sammlung von Einschätzungen ohne Verantwortungszuordnung. Ein Risikoinhaber, der nicht namentlich und datiert freizeichnet, ist kein Risikoinhaber.

Dokumentenmanagement. Richtlinien und Policies haben einen Lebenszyklus. Eine Passwort-Policy, die seit 2019 unverändert im Intranet steht, ist kein Steuerungsdokument – sie ist ein Relikt. Funktionsfähiges Dokumentenmanagement stellt sicher, dass jedes Dokument eine versionierte Historie hat, Änderungen eine Vier-Augen-Freigabe durchlaufen, die aktuell gültige Version eindeutig identifizierbar ist und veraltete Versionen archiviert statt gelöscht werden. Im Audit interessiert nicht nur der aktuelle Stand, sondern auch, was zum Zeitpunkt eines Vorfalls gültig war.

Audit Trail. Jede Änderung an einem Risikoeintrag, jede Freigabe, jede Eskalation und jede Behandlungsentscheidung muss mit Zeitstempel, Benutzer und Begründung protokolliert sein – unveränderlich und reproduzierbar. Das ist keine Formalität: Im Schadensfall, bei einer behördlichen Prüfung oder im Streitfall zwischen Versicherer und Versicherungsnehmer entscheidet der Audit Trail, ob eine Organisation nachweisen kann, dass sie mit angemessener Sorgfalt gehandelt hat. Ein System ohne lückenlosen Audit Trail ist für Nachweiszwecke wertlos. Keine dieser drei Anforderungen ist von vornherein an spezialisierte GRC-Software gebunden. Das Werkzeug ist sekundär. Die Disziplin ist primär.

Git als GRC-System.



Ein gepflegtes Git-Repository mit einem integrierten Ticketsystem erfüllt alle drei Kernanforderungen – vollständig, nachweisbar und ohne fünfstellige Jahreslizenz. Das ist kein Behelfskonstrukt. Es ist ein durchdachter Ansatz, der die Stärken von Werkzeugen nutzt, die Entwicklungs- und Betriebsteams ohnehin beherrschen. [Racz et al. 2010]

Audit Trail: die Commit-History. Jedes Git-Commit enthält Autor, Zeitstempel und Änderungsinhalt – kryptografisch verkettet und nachträglich nicht manipulierbar. Wer welche Policy wann in welcher Version freigegeben hat, ist lückenlos rekonstruierbar. Kein GRC-Tool liefert einen robusteren Audit Trail als ein sauber geführtes Repository.

Vier-Augen-Prinzip: Pull Requests. Jede Änderung an einer Richtlinie, einem Kontrollrahmen oder einem Risikoeintrag durchläuft einen Pull Request: Review durch eine zweite Person, dokumentierte Diskussion, explizite Genehmigung vor dem Merge. Das ist nicht

nur Compliance – das ist kollaborative Qualitätssicherung. Der gesamte Freigabeprozess ist im Repository sichtbar, suchbar und exportierbar.

Risikoregister: Issues und Projects. Ein strukturiertes Issue-System mit definierten Labels (Risikoinhaber, Behandlungsstrategie, Status, Restrisiko), Meilensteinen und Verantwortlichkeiten bildet ein vollwertiges Risikoregister ab. Filter- und Exportfunktionen liefern die Übersichtsformate, die im Steuerkreis oder Audit gefragt werden. Automatische Erinnerungen bei Review-Fälligkeit lassen sich mit wenigen Zeilen CI/CD-Konfiguration umsetzen.

Dokumentenmanagement: Markdown und statische Seiten. Policies und Richtlinien als Markdown-Dateien im Repository versioniert – mit GitHub Pages, MkDocs oder Docusaurus veröffentlicht – erfüllen alle Anforderungen an Zugänglichkeit, Versionierung und Aktualität. Die jeweils gültige Version ist die auf dem Hauptbranch. Veraltete Versionen sind in der History abrufbar, nicht verloren.

Wer nicht mit Git arbeiten möchte oder kann, findet denselben Funktionsumfang in anderen Werkzeugkombinationen: Confluence mit strukturierten Seitenvorlagen und Jira für das Risikoregister, Notion mit Datenbanken und Genehmigungsworkflows, oder Microsoft 365 mit SharePoint-Versionierung und Planner und vielen weiteren Lösungen. Die Plattform ist nicht entscheidend. Entscheidend ist, dass die Prozesse existieren, gelebt werden und Spuren hinterlassen. Ein gepflegtes Git-Repository leistet mehr Compliance als ein ungenutztes GRC-Tool – und es skaliert mit der Organisation.

Wann lohnt ein GRC-Tool?

Spezialisierte GRC-Tools haben berechtigte Stärken, die unter bestimmten Bedingungen ihren Einsatz rechtfertigen. Sie eignen sich insbesondere dann, wenn eine Organisation ausreichend IT-affin ist, um ein weiteres Werkzeug tatsächlich zu nutzen und in bestehende Prozesse zu integrieren – fehlende Akzeptanz macht auch einfache Werkzeuge wertlos. Wenn spezifische Stärken der GRC-Software tatsächlich benötigt werden: Multi-Framework-Mapping über mehrere Normen hinweg, regulatorische Berichtsformate, strukturierte Prozessführung für weniger erfahrene Nutzer. Und wenn alle eigenen Anforderungen vollständig bekannt und im Tool abbildbar sind – *bevor* die Kaufentscheidung fällt. Das Problem definiert die Lösung – nicht andersherum. [Racz et al. 2010; ISO 27001 2022]

Zusammenfassung

Alle Mythen teilen eine gemeinsame Struktur: Sie vereinfachen komplexe Sachverhalte so weit, dass die Vereinfachung selbst zum Problem wird. Risikomanagement ist kein Selbstzweck und keine Compliance-Übung. Es ist Entscheidungsvorbereitung unter Unsicherheit – mit dem Ziel, besser zu entscheiden, nicht vollständiger zu dokumentieren.

Der erste Mythos der Terminologie zeigt, dass Kommunikation über Risiken nur dann funktioniert, wenn alle Beteiligten dasselbe meinen, wenn sie denselben Begriff verwenden. Ein

gemeinsames Glossar ist keine Bürokratie – es ist die Voraussetzung für jede sinnvolle Risikodiskussion. Der Fail der Phishing-Reaktion demonstriert, dass Sicherheitskultur wichtiger ist als Sicherheitstraining. Awareness funktioniert nur, wo Melden sicher ist. Der zweite Mythos der Wahrscheinlichkeit enttarnt die Fehlende-Daten-Ausrede als Kategorienfehler: Fehlende Massendaten sind kein Argument gegen Wahrscheinlichkeitsaussagen – das falsche Konzept schon. Der dritte Mythos der Risikohierarchie verdeutlicht, dass jedes Risiko auf die organisatorische Ebene gehört, auf der es beherrschbar ist – operativ, taktisch oder strategisch. Der vierte Mythos der Ampelwelt zeigt, dass Ordinalskalen nicht multipliziert werden dürfen und dass Quantifizierung kein Luxusproblem ist, sondern die einzige methodisch ehrliche Alternative. Der fünfte Mythos der GRC-Software erinnert daran, dass Prozesse Werkzeuge ermöglichen – nicht Werkzeuge Prozesse erzeugen.

Literaturverzeichnis

Box 1976 — Box, G.E.P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791–799.

BSI 200-2 2017 — Bundesamt für Sicherheit in der Informationstechnik (2017). *BSI-Standard 200-2: IT-Grundschutz-Vorgehensweise* (Version 1.0). BSI. https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/BSI-Standards/BSI-Standard-200-2-IT-Grundschutz-Methodik/bsi-standard-200-2-it-grundschutz-methodik_node.html

BSI 200-3 2017 — Bundesamt für Sicherheit in der Informationstechnik (2017). *BSI-Standard 200-3: Risikoanalyse auf der Basis von IT-Grundschutz* (Version 1.0). BSI. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/BSI_Standards/standard_200_3.pdf

Cox 2008 — Cox, L.A. (2008). What's wrong with risk matrices? *Risk Analysis*, 28(2), 497–512.

de Finetti 1974 — de Finetti, B. (1974). *Theory of Probability* (2 Bde.). Wiley.

Dekker 2007 — Dekker, S. (2007). *Just Culture: Balancing Safety and Accountability*. Ashgate.

DORA 2022 — Europäisches Parlament und Rat (2022). Verordnung (EU) 2022/2554 (Digital Operational Resilience Act). Amtsblatt der EU. https://eur-lex.europa.eu/eli/reg/2022/2554/oj

EASA 2014 — European Aviation Safety Agency (2014). *Just Culture Policy*. EASA.

Hadnagy 2011 — Hadnagy, C. (2011). *Social Engineering: The Art of Human Hacking*. Wiley.

Hájek 2012 — Hájek, A. (2012). Interpretations of probability. In E. Zalta (Hrsg.), *Stanford Encyclopedia of Philosophy*. Stanford University.

Hollnagel 2014 — Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management*. Ashgate.

Howard 1966 — Howard, R.A. (1966). Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1), 22–26.

Hubbard 2010 — Hubbard, D.W. (2010). *How to Measure Anything: Finding the Value of Intangibles in Business* (2. Aufl.). Wiley.

Hubbard & Seiersen 2023 — Hubbard, D.W. & Seiersen, R. (2023). *How to Measure Anything in Cybersecurity Risk* (2. Aufl.). Wiley.

ICAO SMM 2018 — International Civil Aviation Organization (2018). *Safety Management Manual* (SMM), Doc 9859 (4. Aufl.). ICAO.

ISO 27001 2022 — International Organization for Standardization (2022). *ISO/IEC 27001:2022 – Information security, cybersecurity and privacy protection*. ISO. https://www.iso.org/standard/82875.html

ISO 27005 2022 — International Organization for Standardization (2022). *ISO/IEC 27005:2022 – Information security, cybersecurity and privacy protection – Guidance on managing information security risks*. ISO. https://www.iso.org/standard/80585.html

ISO 31000 2018 — International Organization for Standardization (2018). *ISO 31000:2018 – Risk management: Guidelines*. ISO.

Jaynes 2003 — Jaynes, E.T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jones & Freund 2014 — Jones, J. & Freund, M. (2014). *Measuring and Managing Information Risk: A FAIR Approach*. Butterworth-Heinemann.

Jones & Freund 2026 — Jones, J. & Freund, M. (2026). *Measuring and Managing Information Risk: A FAIR Approach* (2. Aufl.). Elsevier.

NIS-2 2022 — Europäisches Parlament und Rat (2022). Richtlinie (EU) 2022/2555 (NIS2-Richtlinie). Amtsblatt der EU. https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng

NIST RMF 2018 — National Institute of Standards and Technology (2018). *SP 800-37 Rev. 2: Risk Management Framework for Information Systems and Organizations*. NIST. https://csrc.nist.gov/projects/risk-management

NIST SP 800-30 2012 — National Institute of Standards and Technology (2012). *SP 800-30 Rev. 1: Guide for Conducting Risk Assessments*. NIST.

O'Hagan et al. 2006 — O'Hagan, A. et al. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley.

Racz et al. 2010 — Racz, N., Weippl, E. & Seufert, A. (2010). A frame of reference for research of integrated governance, risk & compliance (GRC). In *Business Information Systems Workshops*. Springer.

Reason 1990 — Reason, J. (1990). *Human Error*. Cambridge University Press.

Renn 2008 — Renn, O. (2008). *Risk Governance: Coping with Uncertainty in a Complex World*. Earthscan.

Russell 1929 — Russell, B. (1929). Zitiert nach: Bell, E.T. (1945). *The Development of Mathematics*. McGraw-Hill, S. 587.

Salmon 1966 — Salmon, W.C. (1966). *The Foundations of Scientific Inference*. University of Pittsburgh Press.

Savage 1954 — Savage, L.J. (1954). *The Foundations of Statistics*. Wiley.

Stevens 1946 — Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.

Tetlock & Gardner 2015 — Tetlock, P. & Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. Crown.

von Mises 1957 — von Mises, R. (1957). *Probability, Statistics and Truth* (2. engl. Aufl.). Dover Publications.

Walley 1991 — Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall.