Rationalität in der Praxis

Definitionen, Herausforderungen, Optimierungsstrategien

Marko Kovic
marko@kovic.ch
+41 76 335 06 17

Dezember 2022

Abstract

Dieses Grundlagenpapier bietet einen Überblick über das Thema (Ir-)Rationalität und dessen praktische Bedeutung. Das Dokument besteht aus drei Teilen. Zunächst wird das Konzept Rationalität philosophisch eingeordnet. Danach werden die praktischen Probleme, welche zu irrationalen Denken und Entscheiden führen, beschrieben. Zuletzt werden die Ansätze des Nudging, Debiasing und Red Teaming vorgestellt, mit denen Irrationalität in der Praxis angegangen werden kann.

Copyright

Das vorliegende Dokument dürfen Sie für *nichtkommerzielle* Zwecke frei teilen, unter Beachtung der üblichen Zitierregeln. Wenn Sie dieses Dokument für *kommerzielle* Zwecke egal welcher Art nutzen möchten, nehmen Sie bitte Kontakt mit mir auf.

Kontakt

✉ marko@kovic.ch

☎ +41 76 335 06 17

🏠 kovic.ch

📍 Lommisweg 12, 8048 Zürich

Inhaltsverzeichnis

I	Auf den Spuren von Rationalität	4
1	Einleitung: Warum Rationalität wichtig ist	5
2	Begriffsklärung	6
2.1	Annäherung ex negativo: Was Rationalität *nicht* ist	6
2.2	Epistemische Rationalität	7
2.3	Instrumentelle Rationalität	8
2.4	Epistemische und instrumentelle Rationalität: Ein Widerspruch? .	8
2.5	Zusammenführung der zwei Dimensionen	9
3	Exkurs 1: Wahrscheinlichkeit, Risiko, Ungewissheit	10
4	Exkurs 2: Logik	11
4.1	Arten des logischen Schliessens	12
4.2	Logische Fehlschlüsse	14
4.3	Denkfallen bei Abduktion	17
II	Irrationalität	19
5	Was Menschen irrational macht	20
5.1	Die besondere Bedeutung kognitiver Biases	20
5.2	Nützliche Heuristiken und gefährliche Biases	20
6	Einige wichtige kognitive Biases	22
6.1	Wahrscheinlichkeit	22
6.2	Informationsverarbeitung	23
6.3	Ursache und Wirkung	24
6.4	Gruppendynamiken	25
6.5	Dinge, Menschen und Organisationen einschätzen	26
6.6	Nutzen und Risiken	27
6.7	Bestehende Überzeugungen	29
III	Rationalität stärken	31

7	Drei Ansätze für bessere Entscheidungen	32
7.1	Nudging	32
7.2	Debiasing	33
7.3	Red Teaming	34
7.4	Nudging, Debiasing und Red Teaming im Vergleich	35
8	Generelle Nudging-Techniken	37
9	Debiasing-Techniken	41
10	Red Teaming-Techniken	45
11	Zusammenfassung	51

Tabellenverzeichnis

1	Wissensmatrix und Unknown Unknowns.	5
2	Vergleich von Nudging, Debiasing, Red Teaming	36
3	Exposure-Nudges.	37
4	Choice Architecture-Nudges.	38
5	Signalling-Nudges.	39
6	Information Design-Nudges.	39
7	Motivation Hacking-Nudges.	40

Abbildungsverzeichnis

1	Schematische Darstellung einer logischen Schlussfolgerung. . . .	12
---	--	----

Teil I

Auf den Spuren von Rationalität

1 Einleitung: Warum Rationalität wichtig ist

Der Begriff «Rationalität» weckt in unseren Köpfen Bilder und Assoziationen, die zusammen einen vagen Umriss, eine Silhouette einer Idee ergeben; eine grobe Vorstellung, was es mit Rationalität auf sich hat (Wir denken beispielsweise intuitiv alle, dass rational zu sein gut ist und, dass wir selber rational sind.). Was Rationalität aber *genau* bedeutet, sowie, wie es denn wirklich um Rationalität in bestimmten Entscheidungen und Entscheidungskontexten steht, ist aber gar nicht so einfach zu beantworten.

Das verwundert nicht besonders, denn das Thema Rationalität ist eigentlich nie ganz explizit auf unserer kollektiven Agenda: Rationalität steht nicht auf dem Lehrplan in der Schule, Rationalität wird in der Berufsausbildung oder im Studium nicht aufgegriffen, und Rationalität wird auch im Berufsleben im Grunde nie aktiv behandelt. Es gibt aber einen gewichtigen Grund, warum wir uns genauer mit Rationalität befassen sollten: Menschen denken und entscheiden in der Praxis systematisch *irrational*.

Die menschliche Irrationalität in der Praxis ist ein doppeltes Problem. *Zum einen* treffen Individuen, Teams und ganze Organisationen aufgrund irrationaler Denk- und Entscheidungsmuster systematisch and anhaltend (katastrophal) schlechte Entscheidungen. *Zum anderen*, und mindestens ebenso wichtig, wissen die Betroffenen aber gar nicht, dass es überhaupt ein Problem geben könnte. Rationalität und Irrationalität ist weitgehend *terra incognita*, und der Schaden, den Irrationalität anrichtet, ist darum ein *Unknown Unknown*: Ein Problem, von dem wir nicht einmal wissen, dass es es geben könnte. Die Logik der Unknown Unknowns ist in Tabelle 1 dargestellt¹.

Tabelle 1: Wissensmatrix und Unknown Unknowns.

		Weiss ich es?	
		Ja	Nein
Weiss ich, ob ich es weiss?	Ja	Known Known	Known Unknown
	Nein	Unknown Known	Unknown Unknown

Irrationalität ist in der Praxis also ein doppelter blinder Fleck: Wir wissen nicht, dass wir irrational denken und entscheiden, wenn wir es tun, und wir wissen nicht, dass wir es nicht wissen. Das macht das Themenfeld Rationalität zu einer doppelt wichtigen Herausforderung, denn nicht nur richtet Irratio-

¹Die Bezeichnung Unknown Unknowns wurde 2002 vom damaligen umstrittenen US-Verteidigungsminister Donald Rumsfeld geprägt, der dieses analytische Schema im Rahmen einer Pressekonferenz verwendet hatte[1, 2].

nalität tagtäglich grossen Schaden an — niemand realisiert, dass Irrationalität diesen Schaden anrichtet.

In Teil II ist die Natur von Irrationalität genauer umrissen, und in Teil III werden drei Ansätze vorgestellt, um gegen Irrationalität vorzugehen und dabei Rationalität zu stärken. Zuvor aber ist es aber nötig, das Konzept Rationalität auf dem sprichwörtlichen Sezierteller im nachfolgenden Abschnitt 2 genauer zu durchleuchten.

2 Begriffsklärung

2.1 Annäherung ex negativo: Was Rationalität **nicht** ist

Rationalität ist ein gleichermassen vager wie bedeutungsschwangerer Begriff. Bevor beschrieben wird, was mit Rationalität im Rahmen des vorliegenden Dokumentes gemeint ist, lohnt es sich, einige Missverständnisse rund um Rationalität zu klären. Mit Rationalität werden nämlich gerne Ideen und Vorstellungen verbunden, die nicht ganz zutreffend oder sogar komplett falsch sind. Die wichtigsten dieser Missverständnisse sind die nachfolgenden:

- *Rationalität ist nicht Emotionslosigkeit.* Die Figur des Spock aus Star Trek wurde über die Jahrzehnte zu einem Sinnbild für Rationalität: Kalt, distanziert und emotionslos meistert Spock dank seiner überlegenen Rationalität viele schwierige Situationen. Die Vorstellung, dass rational zu sein bedeutet, à la Spock unsere Menschlichkeit und Emotionalität und Intuition komplett abzulegen, ist aber nicht nur falsch, sondern illusorisch (Wir können rationaler denken und entscheiden und trotzdem ganz «normale» Menschen bleiben. Wir können nicht nicht Mensch sein; Spock aus Star Trek können wir nur spielen).
- *Rationalität ist nicht Intelligenz.* Rational zu sein kann missverstanden werden als intelligent sein. Rationalität und Intelligenz sind aber grundsätzlich zwei verschiedene Konzepte[3, 4]. Bildlich gesprochen ist Intelligenz die Rechenpower unseres Denkapparates, während Rationalität die Qualität der Berechnungen ist.
- *Rationalität ist nicht Besserwissertum.* Rational zu sein bedeutet nicht, besonders viel Faktenwissen zu haben. Wer sehr viel Faktenwissen hat und eine Art lebendiges Wikipedia ist, hat wohl einen Vorteil in Quizsendungen, ist aber darum nicht besonders rational.

- *Rationalität ist nicht gruppenspezifisch.* In Alltagssituationen werden einzelne Gruppen bisweilen als weniger rational als andere angesehen. Ein typisches Beispiel ist der Glaube, dass Frauen weniger rational als Männer seien. Das ist nicht nur nicht zutreffend, sondern ein Ausdruck irrationaler Denkmuster (In diesem Fall schlägt der Stereotyping-Bias zu [5]; mehr dazu in Teil II). Rationalität und Irrationalität sind gleichverteilt.

Diese typischen Missverständnisse geben an, was Rationalität *nicht* ist. Doch was ist Rationalität nun in einem positiven Sinn? Rationalität umfasst zwei Dimensionen: Die *epistemische Rationalität* und die *instrumentelle Rationalität*. Diese sind in den nachfolgenden Unterabschnitten beschrieben.

2.2 Epistemische Rationalität

Das griechische Wort «epistēmē» bedeutet so viel wie Wissen oder Erkenntnis. Das eingedeutschte Adjektiv «epistemisch» bedeutet entsprechend, dass etwas auf Wissen bezogen ist. Epistemische Rationalität ist also eine Dimension von Rationalität, die mit Wissen zu tun hat. Doch was ist Wissen?

Im Alltagsverständnis verstehen wir Wissen am ehesten als so etwas wie Faktenwissen: Die Erde ist Rund (das ist eine Tatsache), und ich weiss, dass die Erde rund ist. Damit haben wir aber noch nicht definiert, was Wissen eigentlich ist, sondern lediglich zirkulär beschrieben, dass Wissen etwas ist, was wir wissen. Aus philosophischer Sicht ist es sinnvoll, zunächst eine Unterscheidung zwischen der Realität an sich und unserem Denken über die Realität zu machen. In der Erkenntnistheorie (auch Epistemologie genannt, also die Lehre des Wissens) sind darum unsere *subjektiven Überzeugungen* (oder *Glaubenssätze*; auf englisch «Beliefs» genannt) Dreh- und Angelpunkt. Genauer: Die Frage, wann unsere Überzeugungen als Wissen gelten können.

Es gibt unterschiedliche Denkrichtungen in der Erkenntnistheorie, aber weitgehend besteht Konsens darüber, dass subjektive Überzeugungen *begründet* sein müssen, um als Wissen gelten zu können [6]. Eine Person, die ihre Überzeugungen über die Welt auf diese Art handhabt, gilt tendenziell als epistemisch rational. Eine Person, die epistemisch rational denkt, ist mit anderen Worten also *eine Person, die gute Gründe hat, um zu glauben, was sie glaubt*².

²Bei der epistemischen Rationalität steht nicht im Vordergrund, ob das, was wir glauben, auch tatsächlich wahr ist. Zum einen können wir irrationalerweise rein zufällig etwas Wahres glauben, aber keine guten Gründe für den Glauben haben. Zum Beispiel kann ich glauben, dass die Erde rund ist, weil ich glaube, dass die runde Erde hohl ist und im Inneren Kobolde leben. Das ist ein Fall des sogenannten «epistemischen Glücks» [7]. Zum anderen können wir

2.3 Instrumentelle Rationalität

Instrumentelle Rationalität hat einen anderen Fokus als epistemische Rationalität. Bei der instrumentellen Rationalität geht es um die praktische Frage, *ob eine Person so handelt, dass sie ihre Ziele erreicht* [8]. Wenn ich nach New York will, aber einen Flug nach Tokyo buche, handle ich instrumentell irrational. Wenn ich nach New York will und einen Flug nach New York buche, handle ich instrumentell rational.

Instrumentelle Rationalität ist die Grundlage der berühmten Rational Choice-Theorie. Die Rational Choice-Theorie postuliert, dass Entscheidungen dann rational sind, wenn wir damit unseren Zielen *nutzenmaximierend* näher kommen. In der Realität sind viele Entscheidungen nicht total binär, bei denen wir unsere Ziele entweder komplett erreichen oder komplett verfehlen. Stattdessen haben wir oft unterschiedliche Entscheidungsoptionen und müssen jene wählen, mit denen wir den Nutzen, den wir anstreben, maximieren können. Das bedeutet in anderen Worten, dass wir so entscheiden müssen, dass wir unser jeweiliges Ziel so gut wie möglich erreichen. Wenn ich zum Beispiel nach New York will und zwischen mehreren Flügen auswählen kann, ordne ich die Optionen rationalerweise gemäss meinen zusätzlichen Präferenzen, etwa dem Preis des Fluges. Wenn Ticket A 800 Franken kostet, Ticket B 750 Franken und Ticket C 700 Franken (bei jeweils gleichem Komfort und Service für den Flug), dann bin ich instrumentell rational, wenn ich Ticket C besser finde als Ticket B; wenn ich Ticket B besser finde als Ticket A; sowie, wenn ich Ticket C besser finde als Ticket A.

2.4 Epistemische und instrumentelle Rationalität: Ein Widerspruch?

In den vorherigen Abschnitten wurden zwei Dimensionen von Rationalität vorgestellt: Epistemische und instrumentelle Rationalität. Nun drängt sich aber die Frage auf, wie genau diese zwei Dimensionen miteinander verbunden sind. Oder handelt es sich womöglich um zwei grundsätzlich unterschiedliche Definitionen von Rationalität, die gar nichts miteinander zu tun haben?

Diese Frage ist Gegenstand anhaltender philosophischer Debatten. Eine Sichtweise, die viele Anhängerinnen und Anhänger hat, ist die Auffassung, dass

sehr gute Gründe haben, etwas zu glauben, was aber nicht wahr ist. Zum Beispiel wurde Isaacs Newtons klassische Mechanik durch Albert Einsteins Relativitätstheorie widerlegt, aber bis zur Formulierung der Relativitätstheorie gab es sehr gute Gründe, an die Newton'sche Mechanik zu glauben.

epistemische Rationalität selber auch nur eine Form instrumenteller Rationalität darstellt, oder zumindest eng mit dieser verwoben ist[9]. Zum Beispiel können wir epistemische Rationalität so interpretieren, dass das Ziel des epistemisch rational Seins ist, möglichst genaue und möglichst präzise Vorstellungen über die Realität zu haben. Wir können epistemische Rationalität aber auch als Mittel zum Zweck erachten, um instrumentell rational sein zu können. Diese Interpretation bedeutet, dass epistemische Rationalität die Grundlage für erfolgreiche instrumentelle Rationalität ist. Wenn ich zum Beispiel einen Regentanz aufführe, weil ich glaube, dass Regentanz das beste Mittel ist, um Regen herbeizuführen, handle ich an und für sich instrumentell rational. Die epistemische Grundlage dieser Handlung ist aber hochgradig irrational, denn es gibt keine guten Gründe, zu glauben, dass Regentanz wirklich zu Regen führt.

2.5 Zusammenführung: Eine alltagstaugliche Definition von Rationalität

Die Diskussion der Definition von Rationalität in den vorangehenden Unterabschnitten ist nur oberflächlich, aber sie fasst einige wichtige Anhaltspunkte für unser Verständnis von Rationalität zusammen. *Erstens* bedeutet Rationalität nicht unbedingt das, was wir im Alltagsverständnis damit verbinden. *Zweitens* umfasst Rationalität zwei wichtige Dimensionen; die epistemische Rationalität und die instrumentelle Rationalität. *Drittens* sind epistemische und instrumentelle Rationalität nicht komplett unterschiedliche Konzepte, sondern zwei sprichwörtliche Seiten der Rationalitäts-Medaille. Wenn wir diese Anhaltspunkte zu einer kompakten Definition von Rationalität zusammenfassen, kann diese folgendermassen lauten:

- *Rationalität ist die Fähigkeit, Ziele sinnvoll zu verfolgen.*

Diese Definition ist stark angelegt an das instrumentelle Verständnis von Rationalität, aber das kleine Adjektiv «sinnvoll» soll auf die epistemische Dimension von Rationalität verweisen: Jede Entscheidung fusst auf einem epistemischen Fundament — ist das Fundament schlecht (also epistemisch irrational), führt die Entscheidung bestenfalls nirgendwo hin und richtet schlimmstenfalls Schaden an.

3 Exkurs 1: Wahrscheinlichkeit, Risiko, Ungewissheit

Ein Konzept, das bei Rationalität eine wichtige Rolle spielt, ist *Wahrscheinlichkeit*. Es gibt Entscheidungen, bei denen Gewissheit über das Ergebnis einer Entscheidung besteht. Wenn ich zum Beispiel beschliesse, einen Ball fallen zu lassen, dann wird der Ball garantiert auf den Boden fallen³. Bei den meisten Entscheidungen, die wir treffen, gibt es aber keine Gewissheiten.

Bei einem grossen Teil von Entscheidungen, die wir treffen, wissen wir zwar, was die Ergebnisse sein können, aber die Ergebnisse sind nicht gewiss, sondern haben jeweils eine bestimmte Wahrscheinlichkeit. Wenn ich zum Beispiel einen Würfel würfle, weiss ich, was die Ergebnisse sein können, und ich kann jedem Ergebnis eine klare Wahrscheinlichkeit zuteilen⁴. Oder wenn ich in ein Flugzeug steige, weiss ich, welche Ergebnisse möglich sind (das Flugzeug stürzt entweder ab oder es stürzt nicht ab), und ich kann die statistische Wahrscheinlichkeit der Ergebnisse sehr genau berechnen⁵. Solche Entscheidungssituationen werden bisweilen als *Risiko* beschrieben: Alle möglichen Ergebnisse und ihre Wahrscheinlichkeiten sind bekannt, und wir können explizit und rational kalkulierend entscheiden, was für ein Risiko wir eingehen wollen.

Es gibt aber auch viele Entscheidungssituationen, in denen nicht alle Ergebnisse und nicht alle Wahrscheinlichkeiten der Ergebnisse bekannt sind. Wenn ich zum Beispiel in ein Unternehmen investiere und Aktien kaufe, weiss ich weder, was in Zukunft alles passieren könnte, noch sind die Wahrscheinlichkeiten für die möglichen zukünftigen Ereignisse bekannt. Vielleicht wird in drei Jahren eine revolutionäre neue Technologie erfunden, welche das Unternehmen, in das ich investiert habe, obsolet macht. Oder wenn ich ein neues Restaurant ausprobieren will, weiss ich nicht, was mir in dem neuen Restaurant wie gut schmeckt. Das neue Restaurant könnte beispielsweise thailändisches Essen anbieten, und weil ich noch nie thailändisches Essen probiert habe, betrete ich

³Unter der vereinfachten Annahme, dass keine besonderen Bedingungen gegeben sind, wie etwa Schwerelosigkeit im Weltraum.

⁴Eine Wahrscheinlichkeit wie in diesem Beispiel, bei der alle Ergebnisse und alle Wahrscheinlichkeiten der Ergebnisse *a priori*, also vor dem Experiment, bekannt sind, ist die sogenannte «klassische» Interpretation von Wahrscheinlichkeit. Klassische Wahrscheinlichkeit ist schlicht das Verhältnis eines gewünschten Ergebnisses zu den möglichen Ergebnissen.

⁵Diese Form der Wahrscheinlichkeit ist als «frequentistische» Wahrscheinlichkeit bekannt. Frequentistische Wahrscheinlichkeit ist schlicht die relative Häufigkeit eines Ergebnisses in vielen zufälligen Experimenten. Wenn Hans zum Beispiel schätzen will, wie wahrscheinlich es ist, dass er morgen zu spät zur Arbeit kommt, kann er schlicht abzählen, an wie vielen Tagen im vergangenen Jahr er pünktlich und an wie vielen er verspätet war.

komplettes Neuland. Bei solchen Entscheidungssituationen ist also *Ungewissheit* (auf Englisch «Uncertainty») gegeben: Wir haben nicht alle Informationen, was die rationale Entscheidungsfindung erschwert. Das dürfte, realistischerweise, bei den meisten Entscheidungssituationen im Alltag der Fall sein.

Ungewissheit ist eine Herausforderung, aber sie muss aber keine komplett ungreifbare Grösse sein, die uns einen Strich durch die Rechnung macht. Im Rahmen der sogenannten bayesianischen Erkenntnistheorie⁶ [10, 11] wird Ungewissheit nämlich als *subjektive Wahrscheinlichkeit* quantifiziert [12, 13]. Subjektive Wahrscheinlichkeit bedeutet, dass wir Dinge, die wir glauben, nicht nur als Entweder-Oder-Überzeugung glauben, sondern die konkrete Stärke der jeweiligen Überzeugung als Wahrscheinlichkeit zum Ausdruck bringen. Wenn ich zum Beispiel glaube, dass ich in Firma A anstatt in Firma B investieren sollte, dann kann ich und sollte ich rationalerweise diese Überzeugung als Wahrscheinlichkeit quantifizieren: Wie stark glaube ich, dass eine Investition in Firma A die richtige Entscheidung ist? Diese Quantifizierung einer Überzeugung als subjektive Wahrscheinlichkeit umfasst sowohl Risiko als auch Ungewissheit.

Eine subjektive Wahrscheinlichkeit als Ausdruck der Stärke einer Überzeugung darf aber natürlich nicht willkürlich sein. Jede subjektive Wahrscheinlichkeit muss epistemisch rational, also *gut begründet* sein. Wenn wir subjektive Wahrscheinlichkeiten auf diese Art einsetzen, lässt sich Ungewissheit im Alltag bändigen: Wir haben nicht alle relevanten Informationen (beispielsweise, weil sie noch nicht existieren), aber wir können darüber nachdenken und begründen, dass und wie Ungewissheit eine Rolle spielen könnte. Das erlaubt uns, in Situationen, in denen Ungewissheit eine Rolle spielt, von dieser nicht total überrumpelt zu werden.

4 Exkurs 2: Logik

Logik, die Lehre der richtigen Schlussfolgerungen, spielt bei Rationalität grundsätzlich eine wichtige Rolle. Logik ist gewissermassen das Fundament, welches es uns erlaubt, rational zu sein. Oder, in einer anderen Metapher ausgedrückt: Wenn Rationalität die Fähigkeit ist, von Stadt A nach Stadt B zu fahren, dann ist Logik das Auto, mit dem wir fahren. Wenn das Auto kaputt ist, kommen wir nicht vom Fleck (oder erleiden einen gefährlichen Unfall). Wenn das Auto funk-

⁶Das Adjektiv bayesianisch ist abgeleitet von Thomas Bayes, einem englischen Statistiker und Pfarrer aus dem 18. Jahrhundert, dessen Arbeit zu bedingten Wahrscheinlichkeiten bis heute grossen Einfluss hat.

tioniert, ist zwar noch nicht garantiert, dass wir wirklich am Zielort ankommen, aber die notwendigen Vorbedingungen, um es zu versuchen, sind gegeben.

Im Kontext von Logik ist es einerseits wichtig, die zentralen Arten des logischen Schliessens zu kennen, sowie andererseits, mit der Problematik der logischen Fehlschlüsse vertraut zu sein.

4.1 Arten des logischen Schliessens

Eine logische Schlussfolgerung kann schematisch in drei Teile zerlegt werden: Eine erste Prämisse als allgemeine Regel, eine zweite Prämisse als spezifische Gegebenheit, sowie eine Schlussfolgerung als notwendige Konsequenz gegeben der Prämissen. Etwas praxisorientierter kann diese Zusammensetzung auch als Regel (erste Prämisse), Ursache (zweite Prämisse) und Wirkung (Schlussfolgerung) beschrieben werden. Das Verhältnis dieser drei Komponenten zueinander ist in Abbildung 1 schematisch dargestellt.

Abbildung 1: Schematische Darstellung einer logischen Schlussfolgerung.

Die Regel ist die erste Prämisse. Die Ursache ist die zweite Prämisse. Die Wirkung ist die Schlussfolgerung.

Im klassischen Sinn bedeutet eine logische Schlussfolgerung ausschliesslich, von einer Regel (erste Prämisse) und einer Ursache (zweite Prämisse) auf eine Wirkung (Schlussfolgerung) zu schliessen. Das ist die sogenannte *deduktive Logik*. Es ist aber auch möglich, von einer Ursache und Wirkung auf die Regel zu schliessen, sowie, von einer Wirkung und Regel auf die Ursache zu schlussfolgern. Diese Formen des logischen Schliessens werden *induktive* und *abduktive* Schlussfolgerungen genannt [14]. Die drei Arten der des logischen Schliessens unterscheiden sich stark voneinander.

Deduktion ist in engerem logischen Sinn die einzig zulässige Form des Schliessens. Bei einer deduktiven Schlussfolgerung haben wir eine Regel (erste Prämisse) und eine Ursache (zweite Prämisse). Daraus leiten wir die Wirkung (Schluss-

folgerung) ab. Deduktive Schlussfolgerungen sind *immer wahr*: Wenn wir eine Regel und eine Ursache haben, ist die Wirkung logisch immer zwingend wahr. Wenn zum Beispiel die Regel (erste Prämisse) «*Wenn es regnet, ist der Boden nass*» lautet, und die Ursache (zweite Prämisse) «*Es regnet*» ist, dann ist die Wirkung (Schlussfolgerung) «*Der Boden ist nass*». Diese Schlussfolgerung ist in logischer Hinsicht zwangsläufig, kategorisch und immer wahr, denn die Prämissen machen diese Schlussfolgerung unumgänglich. Diese Eigenschaft der «absoluten» Wahrheit (die Schlussfolgerung ist entweder absolut wahr oder absolut falsch) haben ausschliesslich deduktive Schlussfolgerungen.

Induktion ist eine Form des logischen Schliessens, bei der wir die Ursache (zweite Prämisse) und die Wirkung beobachten und daraus auf die zugrundeliegende Regel (erste Prämisse) schliessen. Induktive Schlussfolgerungen machen wir tagtäglich, ohne es wirklich zu merken. Wenn wir uns zum Beispiel als Kind die Finger an der Herdplatte verbrennen, machen wir eine induktive Schlussfolgerung: Hand auf Herdplatte (Ursache) tut weh (Wirkung), was bedeutet, dass es allgemein weh tut, wenn ich die Hand auf die Herdplatte setze (Regel). Induktive Schlussfolgerungen sind aber nicht immer und nicht kategorisch wahr. Mit induktiven Schlüssen sammeln wir lediglich Evidenz, dass eine Regel mit einer bestimmten Wahrscheinlichkeit wahr sein könnte, aber wir könnten uns, empirisch gesehen, auch täuschen. Induktive Schlussfolgerungen spielen nicht zuletzt in der Wissenschaft eine grosse Rolle: Immer, wenn in der Wissenschaft Daten gesammelt werden, um mögliche Zusammenhänge und Kausalitäten zu messen, findet induktives Schliessen statt.

Abduktion ist jene Form des logischen Schliessens, bei der wir eine Wirkung beobachten und eine oder mehrere mögliche Regeln als mögliche Erklärung in Betracht ziehen, um daraus zu schlussfolgern, was die Ursache sein könnte. Das Ziel einer abduktiven Schlussfolgerung ist also, jene Regel auszuwählen, welche die beobachtete Wirkung am besten erklärt, damit daraus ein Rückschluss auf die Ursache gemacht werden kann⁷. Ein klassisches Beispiel für Abduktion sind medizinische Diagnosen. Eine Ärztin wird mit Symptomen konfrontiert (die Wirkung), überlegt sich mögliche Regeln und Mechanismen, wie die Symptome zustande kommen können (Regel), und schlussfolgert dann anhand der am besten passenden Regel, was die Ursache für die Symptome ist, was also die Krankheit oder das Gebrechen ist, welches die Symptome verursacht. Ein anderes klassisches Beispiel für abduktives Schliessen ist die fiktive Figur des Sher-

⁷Im Englischen wird Abduktion auch «*Inference to the best explanation*» genannt, also «Schlussfolgerung auf die beste Erklärung», was diese Form des logischen Schliessens sehr treffend beschreibt.

lock Holmes. In seinen Abenteuern untersucht Sherlock Holmes immer wieder Tatorte und schlussfolgert blitzschnell, was passiert sein muss, damit der Tatort so wurde, wie er ist (Holmes rekonstruiert die passende Regel und die passende Ursache, welche zur Wirkung geführt haben.). Abduktive Schlüsse kommen aber nicht nur in der Medizin und bei Sherlock Holmes vor — im Grunde üben wir uns jedes Mal in Abduktion, wenn wir erklären oder verstehen wollen, wie eine Situation entstanden ist oder wie ein Problem gelöst werden kann. Das Smartphone funktioniert nicht? Abduktiv schlussfolgern wir, dass der Akku leer sein könnte. Die neue Mitarbeiterin kündigt in der Probezeit? Abduktiv schlussfolgern wir, dass sie ein besseres Jobangebot erhalten haben könnte. Das grosse Projekt im Unternehmen ist gescheitert? Abduktiv suchen wir nach Gründen, warum.

4.2 Logische Fehlschlüsse

Logik wird dann zu einem potenziellen Problem für Rationalität, wenn die Schlussfolgerungen, welche wir treffen, fehlerhaft sind. In solchen Fällen (um erneut die weiter oben eingesetzte Metapher zu bemühen) wollen wir mit dem Auto von A nach B fahren, aber etwas funktioniert mit unserem Auto nicht so richtig, sodass wir vom Weg abkommen, ohne es zu merken.

Es gibt zwei wesentliche Gruppen von Fehlschlüssen: *Formale* und *informelle* Fehlschlüsse. Formale Fehlschlüsse sind deduktive Argumente, deren Struktur rein formal falsch ist. Ein Beispiel: «*Menschen aus Mexiko trinken gerne Tequila. José trinkt gerne Tequila. José ist also Mexikaner.*» ist eine formal immer falsche Schlussfolgerung, weil sie deduktiv nicht abgeleitet werden kann⁸. Ein anderes Beispiel: «*Alle Katzen sind Säugetiere. Julia ist ein Säugetier. Julia ist also eine Katze*»⁹. Bei formalen Fehlschlüssen ist nicht die inhaltliche Wahrheit der Prämissen relevant, sondern nur deren formale, deduktive Struktur. Das Argument «*Alle Schweden sind Ausserirdische vom Planeten Venus. Erik ist ein Schwede. Erik ist als ein Ausserirdischer vom Planeten Venus.*» ist formal richtig, obwohl der Inhalt der Behauptung völliger Quatsch ist.

Formale Fehlschlüsse sind grundsätzlich ein Problem, aber wir können sie verhältnismässig klar identifizieren, weil es bei formalen Fehlschlüssen keinen Diskussions- und Interpretationsspielraum gibt; sie sind einfach und klar

⁸Wir können aus der Schlussfolgerung (José ist Mexikaner) nicht auf die Ursache schliessen. Eine formal korrekte deduktive Schlussfolgerung würde lauten: «*Menschen aus Mexiko trinken gerne Tequila. José ist aus Mexiko. José trinkt also gerne Tequila.*».

⁹Die Regel besagt, dass alle Katzen Säugetiere sind, aber nicht, dass alle Säugetiere Katzen sind.

falsch. Anders sieht es bei *informellen Fehlschlüssen* aus. Informelle Fehlschlüsse sind Fehlschlüsse, deren formale Struktur oft korrekt ist. Das, was bei ihnen nicht stimmt, ist stattdessen inhaltlicher Natur. Das macht informelle Fehlschlüsse zu einem schwierigen Phänomen, weil sie logisch korrekt wirken können und inhaltlich eine überzeugende, wenn auch trügerische Geschichte erzählen. Informelle Fehlschlüsse können sowohl bei deduktiven als auch bei induktiven Schlussfolgerungen auftreten. Einige der wichtigeren informellen Fehlschlüsse sind die folgenden:

- ***Ad hominem***. Der Ad hominem-Fehlschluss wird begangen, wenn als Argument gegen eine Idee, Sache, Gegenstand usf. nicht die Idee, Sache, oder Gegenstand an sich kritisiert wird, sondern die Person hinter der Idee, Sache, Gegenstand. Ein Beispiel: «A: *Klimawandel ist ein Problem für die Menschheit*. B: *Nein, das stimmt nicht. Du bist doch letzten Sommer mit dem Flugzeug in die USA geflogen, und dein Auto ist ein Diesel*.».
- ***Post hoc, ergo propter hoc***. Post hoc, ergo propter hoc ist lateinisch und bedeutet «Nach diesem, also wegen diesem»¹⁰. Das ist der Fehlschluss, zu glauben, dass eine bloße Korrelation auch Kausalität bedeutet. Ein Beispiel: «*Ich habe homöopathische Globuli geschluckt und dadurch wurde ich wieder gesund*.».
- ***Appeal to nature***. Weil etwas «natürlich» ist, glauben wir oft, dass es darum gut oder wahr oder richtig ist. Ein Beispiel: «*Ich impfe meine Kinder nicht gegen Masern. Masern sind etwas ganz Natürliches*.».
- ***Appeal to popularity***. Weil etwas beliebt ist, glauben wir oft, dass es darum gut oder wahr oder richtig ist. Ein Beispiel: «*Milliarden von Menschen glauben an Astrologie und Horoskope. Darum muss an Astrologie etwas dran sein*.».
- ***Appeal to tradition***. Weil etwas Tradition hat und schon lange vorhanden ist, glauben wir oft, dass es darum gut oder wahr oder richtig ist. Ein Beispiel: «*Unser Unternehmen setzt seit 30 Jahren den gleichen Prozess für die Rekrutierung und Bewerbungsprozesse ein. Darum muss der Prozess gut sein*.».
- ***Appeal to authority***. Weil eine (wie auch immer geartete) Autoritätsfigur etwas behauptet, glauben wir oft, dass das, was die Autoritätsfigur behauptet, darum gut oder wahr oder richtig ist. Ein Beispiel: «*Auf Instagram*

¹⁰Manchmal wird auch die Beschreibung «Cum hoc, ergo propter hoc» verwendet: Mit diesem, also wegen diesem.

hat die Influencerin XY, die 15 Millionen Follower hat, behauptet, Coca Cola sei besser als Pepsi. Ich trinke ab jetzt nur noch Coca Cola!«.

- **Zirkuläres Argument.** Wir verheddern uns oft in Argumenten, die in Tat und Wahrheit zirkulär und damit nichtssagend sind. Ein Beispiel: *«Die Bibel ist das Wort Gottes, denn in der Bibel steht, dass sie das Wort Gottes ist.»*
- **Red Herring.** In Argumenten begehen wir oft «Ausweichmanöver», die vom eigentlichen Inhalt, um den es geht, ablenken sollen¹¹. Ein Beispiel: *«A: Donald Trump handelt unmoralisch und gesetzeswidrig. B: Ach ja? Und was ist mir Hillary Clinton?».*
- **Argument der Mässigung.** Wir glauben oftmals, dass die wahre oder richtige oder gute Entscheidung in der Mitte zwischen zwei vermeintlichen Polen liegen muss. Ein Beispiel: *«A: Infektionskrankheiten sind schlecht! B: Nein, Infektionskrankheiten sind gut! C: Die Wahrheit liegt in der Mitte — manche Infektionskrankheiten sind gut und manche schlecht.»*
- **No True Scotsman.** Menschen neigen dazu, wie auch immer geartete Klassen oder Gruppen generalisiert, universal zu beschreiben. Wenn sie mit einem Beispiel, das ihre allgemeine Regel widerlegt, konfrontiert sind, wird dieses Beispiel gerne als Anomalie oder Ausreisser oder als kein echtes Mitglied dieser Klasse beschrieben. Ein Beispiel: *«A: Leute, die Videospiele spielen, lesen keine Bücher. B: Jörg spielt Videospiele und ist ein richtiger Bücherwurm. B: Nun gut, Jörg spielt ja auch keine normalen Videospiele.».*
- **Argument der Ignoranz.** Wir glauben bisweilen, dass etwas, was bisher nicht widerlegt worden ist, wahr sein muss; oder, dass etwas, was bisher nicht bestätigt wurde, falsch sein muss. Ein Beispiel: *«Bisher gibt es keinerlei Anzeichen für ausserirdisches Lebens. Darum wissen wir, dass es kein Leben jenseits der Erde gibt.».*
- **Kompositions-Fehlschluss.** Oft machen wir die zu schnelle Schlussfolgerung, dass etwas, was für einen Teil des Ganzen zutrifft, auch für das Ganze zutreffen muss. Ein Beispiel: *«Max und Julia gehen in dieselbe Klasse und haben sehr gute Noten. Das bedeutet, dass die ganze Schulklasse sehr gut ist.».*

¹¹Dieser Fehlschluss ist auch als «Whataboutism» bekannt, nach dem Ausdruck «What about...?» («Und was ist mit...?»), der bei dem Red Herring oft direkt oder indirekt verwendet wird. Der Begriff «Red Herring» stammt offenbar von der Vorstellung, Jagdhunde liessen sich durch den starken Geruch eines geräucherten Salzherings auf eine falsche Fährte locken.

- **Special Pleading.** Wenn es generelle, universale Eigenschaften einer Klasse oder Gruppe gibt, suchen wir oft nach Ausreden, warum diese Eigenschaften für einen Teil dieser Klasse oder Gruppe nicht zutreffen soll¹². Ein Beispiel: «A: Arzneimittel können in randomisierten Experimenten auf ihre Wirksamkeit getestet werden, auch die Homöopathie. B: Nein, die Homöopathie funktioniert ganz anders!».
- **Strohmann-Argument.** Bei Sachverhalten, bei denen starke Meinungen gegeben sind, werden die jeweiligen Argumente der Gegenseits oftmals stark überzeichnet und ins Absurde geführt. Damit wird nicht das eigentliche Argument, um welches es geht, kritisiert, sondern ein bildlicher Strohmann; eine verzerrte Karikatur des Argumentes. Ein Beispiel: «A: Wir sollten etwas gegen Klimawandel machen. B: Du findest also ernsthaft, wir dürfen keine Autos mehr fahren und müssen uns nur noch von Gras ernähren?».

Informelle Fehlschlüsse sind also nicht einfach formallogisch falsch, sondern inhaltlich fehlerhaft. Ein wichtiger Grund, warum wir ihnen auf dem Leim gehen, ist der Umstand, dass sie der Art und Weise, wie wir intuitiv denken, entspringen: Sie sind eng mit kognitiven Heuristiken verbunden [15]. Das Problem der Heuristiken ist im nachfolgenden Teil II beschrieben.

4.3 Denkfallen bei Abduktion

Abduktion ist womöglich die anspruchsvollste Form des logischen Schliessens. Gleichzeitig ist Abduktion auch jene Form des Schliessens, derer wir uns in der Regel am wenigsten bewusst sind, obwohl wir uns tagtäglich darin üben. Die Folge davon ist, dass wir bei abduktivem Denken Gefahr laufen, in unterschiedliche Denkfallen zu tappen und abduktive Schlüsse fehlerhaft zu machen.

Erstens ist Abduktion abhängig von unserem bestehenden Wissen über Regeln und Ursachen, die bei einer beobachteten Wirkung (dem Problem, um das es geht) in Frage kommen. Ein gutes Sinnbild hierfür ist Medizin: Der Grund, warum eine Ärztin nützlich abduktive Schlussfolgerungen für Symptome bei Patienten treffen kann, ist, dass sie über sehr viel Wissen verfügt, das sie rasch abrufen kann. Über je weniger Wissen wir im Kontext einer abduktiven Schlussfolgerung verfügen, desto unwahrscheinlicher ist es, dass wir einen guten abduktiven Schluss machen.

¹²Daher der Name Special Pleading: Wir behaupten, der eine Fall, um den es geht, sei etwas ganz Besonderes, und darum würde die Eigenschaft der Klassen dieses Falls für ihn nicht zutreffen.

Zweitens besteht bei abduktiven Schlussfolgerung die Gefahr, dass wir in ein deterministisches Denkmuster hineinrutschen. Abduktive Schlussfolgerungen können immer nur probabilistisch sein, also rationalerweise nur mit einer bestimmten Wahrscheinlichkeit zutreffen. In der Hitze des Gefechtes kann es aber gut sein, dass wir verzweifelt nach einer Erklärung für ein Problem suchen und glauben, mit Sicherheit die richtige abduktive Schlussfolgerung getroffen zu haben, sobald wir eine halbwegs plausible Erklärung zusammengeschustert haben.

Drittens besteht bei abduktiven Schlussfolgerungen die Gefahr, dass wir uns in eine spezifische abduktive Schlussfolgerung verbeissen und nicht gewillt sind, unsere Meinung zu revidieren. Wenn beispielsweise Anita einen Job nicht kriegt, ist es denkbar, dass sie als abduktive Erklärung zum Schluss kommt, dass das Unternehmen Frauen diskriminiert. Diese Erklärung ist möglicherweise korrekt, aber es könnte auch sein, dass eine andere Kandidatin oder Kandidat besser qualifiziert war. Das Festhalten an einer spezifischen abduktiven Schlussfolgerung zuungunsten der rationalen Revision der eigenen Meinung ist bedingt durch irrationale Denkmuster wie den Confirmation Bias und Anchoring; mehr zu diesen in Abschnitt 6 in Teil II.

Teil II

Irrationalität

5 Was Menschen irrational macht

5.1 Die besondere Bedeutung kognitiver Biases

Eine Person ist dann irrational, wenn sie etwas ohne gute Gründe glaubt, oder, wenn sie eine Entscheidung trifft, mit der sie das Ziel, das sie eigentlich erreichen will, teilweise oder komplett verfehlt. Für irrationales Denken und Entscheiden kann es grundsätzlich viele Gründe geben. Vielleicht sind wir gerade betrunken oder sonstwie auf Drogen; vielleicht sind wir gerade emotional aufgewühlt; vielleicht haben wir die letzte Nacht kaum geschlafen und sind todmüde; vielleicht sind wir als Kind in ein religiöses oder sonstwie ideologisches Weltbild hineinsozialisiert worden, das wir nicht zu hinterfragen gelernt haben. Und so weiter, und so fort.

Diese und viele andere Extremsituation und Spezialfälle können zweifellos einen negativen Impact auf das Denken und Entscheiden in der Praxis haben. Doch sie sind nicht die wichtigste Ursache von Irrationalität. Zum einen sind derartige Extremsituationen und Spezialfälle nicht der Standardzustand, sondern eben Extremsituationen und Spezialfälle. Zum anderen können wir derartige Extremsituationen und Spezialfälle relativ gut als solche erkennen. Wenn jemand beispielsweise LSD konsumiert und berichtet, einen rosaroten Elefanten gesehen zu haben, können wir mit hoher Zuversicht einschätzen, dass die Person irrationalerweise glaubt, einen rosaroten Elefanten gesehen zu haben.

Es gibt aber auch eine Form der Irrationalität, die nicht die Ausnahme, sondern die Regel bei unserem Denken und Entscheiden ist. Eine Irrationalität, die zudem nicht einfach als solche zu erkennen ist: *Kognitive Biases*; subtile, aber systematische kognitive Verzerrungen in unserem Denken und Entscheiden.

5.2 Nützliche Heuristiken und gefährliche Biases

Kognitive Biases sind eine Folge unserer *kognitiven Heuristiken*. Heuristiken sind automatisierte Denkmuster, eine Art kognitive Daumenregeln, mit denen wir ohne grossen Aufwand relativ schnell Entscheidungen treffen können [16]. Die Natur kognitiver Heuristiken wird oft mit der Metapher des System 1- und System 2-Denkens beschrieben [17, 18]. In dieser Metapher wird das System 1 als eine Art Autopilot gedacht: Ein schnelles, automatisiertes Denken, das uns erlaubt, viele Entscheidungen sehr effizient zu treffen. Im Kontrast dazu erlaubt uns das System 2, bei Bedarf in einen langsamen, bewussten und überlegten Modus zu schalten, in welchem wir Entscheidungen bedachter und expliziter

treffen.

Kognitive Heuristiken sind an und für sich nicht schlimm oder unerwünscht. Im Gegenteil: Heuristiken erlauben es uns überhaupt erst, uns halbwegs erfolgreich durch den Alltag zu manövrieren. Hätten Heuristiken nämlich gar keine nützliche Funktion, gäbe es sie gar nicht. Evolutionsbiologisch gesehen dürften Heuristiken nämlich gewisse Überlebensvorteile mit sich gebracht haben, was dazu geführt hat, dass Heuristiken zu einem universalen biologischen Merkmal von uns Menschen wurden [19]. Der Überlebensvorteil von Heuristiken kann anhand eines vereinfachten Beispiels illustriert werden. Angenommen, ein früher Mensch in der Savanne Ostafrikas wäre immer perfekt rational gewesen, wenn er ein Rascheln im Gras hört. Er hätte jedes Mal versucht, zuerst eine Hypothese aufzustellen und dann diese wie ein Wissenschaftler empirisch getestet, um daraus abzuleiten, was er glauben soll. Das Problem mit einer solchen Strategie ist aber, dass im Zweifelsfall ein Löwe im Gras lauerte, der den rationalen frühen Menschen gefressen hat. Handkehrum wäre der irrationale frühe Mensch jedes Mal davongerannt, wenn es im Gras raschelt, obwohl meistens nur der Wind schuld war. Im Zweifelsfall also konnte die schnelle, automatisierte Entscheidung des irrationalen frühen Menschen sein Leben retten.

Nicht nur für die frühen Menschen in der Savanne waren Heuristiken nützlich. Auch im heutigen modernen Alltag helfen uns Heuristiken enorm, Entscheidungen zu treffen, die oftmals gut genug sind. Sie können in gewissen Situationen sogar zu besseren Ergebnissen als formal rationale Entscheidungsfindungsprozesse führen [20]¹³. Wir dürfen nicht den Fehler machen und Heuristiken dämonisieren — sie sind ein fester und wichtiger Bestandteil des menschlichen Denkapparates. Doch leider können Heuristiken in vielen Entscheidungssituationen auch einen systematisch verzerrenden Effekt haben. Die Folge davon ist, dass wir zwar immer noch schnell und spontan Entscheidungen treffen, die Entscheidungen aber irrational sind, weil wir die Ziele, welche wir verfolgen, damit nicht optimal erreichen oder komplett verfehlen.

Dann, wenn kognitive Heuristiken systematisch verzerrende Einflüsse haben, werden sie *kognitive Verzerrungen* oder, vom englischen Ausdruck für Verzerrung abgeleitet, *kognitive Biases* genannt. [21]. Im nachfolgenden Abschnitt 6 sind einige wichtige kognitive Biases beschrieben.

¹³Das mag überraschend klingen, ist aber auf den zweiten Blick offensichtlich. Angenommen, jemand wirft Hans auf der Wiese im Park einen Ball zu, und Hans möchte ihn fangen. Hans könnte im Prinzip hoch rational entscheiden, wie er den Ball fangen soll, indem er die Flugbahn berechnet. Das dürfte aber nicht funktionieren (ausser, Hans ist ein Physikgenie). Stattdessen verlässt sich Hans auf eine einfache, intuitive Heuristik: Er verfolgt den Ball mit seinem Blick und positioniert sich so, dass der Blickwinkel immer in etwa derselbe bleibt.

6 Einige wichtige kognitive Biases

Es gibt Dutzende kognitive Biases, die in der Kognitionspsychologie und Verhaltensökonomie erforscht wurden und werden. Sie alle in dem vorliegenden Dokument aufzulisten, würde den Rahmen des Zumutbaren sprengen¹⁴. Stattdessen sind einige der bedeutenderen und gut dokumentierten Biases in den nachfolgenden Unterabschnitten beschrieben. Die Biases sind dabei in unterschiedliche Gruppen eingeordnet. Diese Gruppierung dient der Veranschaulichung und dem Verständnis; in der Forschung zu kognitiven Biases gibt es (noch) keine fixen Gruppen oder Typen von Biases.

6.1 Wahrscheinlichkeit

Im Alltag müssen wir oft in Wahrscheinlichkeiten denken und entscheiden, aber intuitiv fällt uns das nicht immer leicht. Einige der wichtigen Biases, die mit Wahrscheinlichkeit zusammenhängen, sind die nachfolgenden:

- **Base Rate Neglect.** Wir ignorieren beim intuitiven Einschätzen der Wahrscheinlichkeit von etwas oft die sogenannte Base Rate, die Ausgangswahrscheinlichkeit dessen, worum es geht [23]. Ein klassisches Beispiel: Viele Menschen reisen nicht gerne mit dem Flugzeug, weil sie das Risiko eines Absturzes meiden wollen, und nehmen stattdessen lieber das Auto. Dabei ignorieren sie die viel höhere Ausgangswahrscheinlichkeit, im Auto einen schweren Unfall zu erleiden.
- **Conjunction Fallacy.** Wir schätzen intuitiv die Wahrscheinlichkeit, dass zwei oder noch mehr Bedingungen zutreffen, oft als höher ein als die Wahrscheinlichkeit, dass nur eine der Bedingungen zutrifft. Das ist mathematisch unmöglich, ist für uns aber die bessere Geschichte [24]. Ein Beispiel: Angenommen, Roger Federer spielt gegen einen Amateur ein Freundschaftsmatch. Viele Leute schätzen die Aussage «Roger Federer verliert einen Satz und gewinnt das Spiel» als wahrscheinlicher ein als die Aussage «Roger Federer verliert einen Satz».
- **Gambler's Fallacy.** Wir glauben manchmal, dass eine Pechsträhne bedeutet, dass wir bald Glück haben müssen, obwohl vergangene Misserfolge keinen Einfluss auf die Wahrscheinlichkeit des zukünftigen Erfolgs haben [25]. Ein Beispiel: Ein Unternehmen hat schon drei Personen für eine

¹⁴Eine empfehlenswerte, weitgehend aktuell gehaltene Übersicht über so gut wie alle Biases findet sich in der «List of cognitive biases» auf Wikipedia [22].

Management-Stelle während der Probezeit gekündigt. Die HR-Abteilung ist überzeugt, dass darum die vierte Person, die eingestellt wird, die richtige sein wird; die Pechsträhne muss ja irgendwann aufhören.

- **Hot Hand Fallacy.** Das Gegenteil der Gambler's Fallacy: Der Glaube, dass eine «Glückssträhne» bedeutet, dass ein zukünftiger Erfolg wahrscheinlicher wird [26]. Ein Beispiel: Ein Roulette-Spieler im Casino hat zwei Mal hintereinander auf die richtige Farbe gesetzt und ist überzeugt, dass er mit seinem glücklichen Händchen auch ein drittes Mal gewinnt.
- **Representativeness.** Wir schätzen die Wahrscheinlichkeit von Dingen als höher ein, wenn die Art und Weise, wie das Ding aussieht, für uns repräsentativer für das Ding an sich ist [27]. Ein Beispiel: Angenommen, wir werfen eine normale, faire Münze 10 Mal. Viele Menschen glauben, dass die Sequenz KKKKKKKKKK unwahrscheinlicher ist als die Sequenz KZZKKZKZKK, weil die zweite Sequenz «normaler» aussieht (Die Wahrscheinlichkeit ist in beiden Fällen identisch.).
- **Black Swan Events.** Menschen sind nicht gut darin, mit sehr kleinen Wahrscheinlichkeiten umzugehen. Wenn ein sogenannter Black Swan Event eintritt — ein sehr unwahrscheinliches und sehr folgenreiches Ereignis —, haben wir Mühe damit, das als Zufall anzusehen [28]. Ein Beispiel: Die Terroranschläge vom 11. September 2001 in den USA waren extrem unwahrscheinliche Ereignisse. Viele Menschen glauben nicht, dass das «zufällig» möglich war; dahinter muss eine Verschwörung der US-Regierung stecken.

6.2 Informationsverarbeitung

Informationsverarbeitung ist ein sehr breiter Begriff, und als Prozess ist Informationsverarbeitung ein wesentlicher Bestandteil dessen, was unsere Denkleistung ausmacht. Es gibt eine Reihe von Biases, die diesen Prozess stören oder negativ beeinflussen können. Einige wichtige davon:

- **Anchoring.** Die erste Information, die wir erhalten, funktioniert als eine Art Anker: Die weiteren Informationen zum Sachverhalt, um den es geht, vergleichen wir mit dieser ersten Information, die besonders tief sitzt [29]. Ein klassisches Beispiel hierfür sind Lohnverhandlungen. Die erste Zahl, die genannt wird, ist der Anker, über den verhandelt wird.

- **Framing.** Die Art und Weise, wie Information präsentiert wird, hat einen Einfluss darauf, wie sie wahrgenommen und verarbeitet wird [30]. Zum Beispiel wirkt der Satz «Du hast 100 Franken und es bleiben dir 90.» weniger schlimm als «Du hast 100 Franken und du verlierst 10.».
- **Priming.** Ein vorangehender Reiz beeinflusst, wie wir darauffolgende Reize wahrnehmen [31]. Wenn eine Person zum Beispiel die Wörter «Blatt; Wiese; Natur; Kleeblatt» hört, wird sie das unvollständige Farbe «gr**» eher als «grün» denn als «grau» vervollständigen.
- **Recency.** Die Information, die am wenigsten weit in der Vergangenheit liegt, beeinflusst uns oft stärker als andere Informationen [32]. Zum Beispiel erinnern wir uns bei einer Präsentation tendenziell besser an den Schluss als auf die Inhalte in der Mitte der Präsentation.
- **Availability.** Wenn wir über Themen, Konzepte oder Entscheidungen nachdenken, orientieren wir uns an starken Beispielen und Bildern in unserem Kopf [33]. Zum Beispiel kaufen viele Menschen Lotto-Tickets, weil die starken Bilder des Lottogewinns uns zum Irrglauben verführen, unsere Chancen stünden ganz gut.

6.3 Ursache und Wirkung

Kausale Zusammenhänge, also Kausalketten von Ursache und Wirkung, zu erkennen, ist ganz allgemein sehr schwierig, sodass uns dabei immer wieder Fehler passieren. Einige der Biases in diesem Bereich sind die nachfolgenden:

- **Survivorship Bias.** Der Survivorship Bias (Überlebensfehler) ist unsere Tendenz, bei Informationen oder Daten den Selektionsprozess, der zu diesen Informationen oder Daten geführt hat, zu ignorieren. Das bedeutet konkreter, dass wir uns oft für Personen oder Organisationen oder sonstige Dinge interessieren, die einen Filter überstanden haben (die also «überlebt» haben) und dabei vergessen, all die Personen oder Organisationen oder sonstigen Dinge, die den Filter nicht überstanden haben, auch mitzuberücksichtigen [34]. Ein typisches Beispiel für den Survivorship Bias sind «Best Practice»-Studien und -Anleitungen, aus denen man angeblich lernen kann, wie Unternehmen erfolgreich werden. Das Problem dabei ist, dass solche Ratgeber meistens nur erfolgreiche Unternehmen anschauen und darum übersehen, dass die Eigenschaften, die angeblich den Erfolg

ausmachen, oft genauso bei den gescheiterten Unternehmen vorhanden waren.

- ***Illusorische Korrelation.*** Wir alle wissen, dass Korrelation nicht Kausalität bedeuten muss. Das Problem beginnt aber noch einen Schritt vorher: Wir bilden uns oft Korrelationen ein, wo rein statistisch keine vorhanden sind [35]. Zum Beispiel meinen viele Leute, dass schlechtes Wetter bei ihnen Schmerzen verursacht. Sie denken dann an die wenigen Fälle, in denen das Wetter schlecht war und sie z.B. Knieschmerzen hatten, aber sie ignorieren all die Schlechtwettertage, an denen sie beschwerdefrei waren (Weil sich die Tage, an denen das Knie weh tut, emotional stärker in unser Gedächtnis einbrennen als «normale» Tage.).
- ***Barnum Effect.*** Wir haben die Tendenz, sehr vage, allgemeine Persönlichkeitsbeschreibungen als für uns persönlich sehr zutreffend und genau wahrzunehmen. Das klassische Beispiel ist Astrologie, wo sehr vage und sich widersprechende Allgemeinplätze als sehr genaue Beschreibungen der eigenen Persönlichkeit empfunden werden [36].
- ***Hindsight Bias.*** Im Nachhinein sind wir immer schlauer. Der Hindsight Bias (Rückschaufehler) ist unsere Tendenz, die Vorhersehbarkeit von Ereignissen im Nachhinein zu überschätzen [37]. Zum Beispiel denken sich heute viele Leute, dass sie damals, als die Kryptowährung Bitcoin praktisch wertlos war, hätten erkennen müssen, dass das eine einmalige Investment-Chance ist.

6.4 Gruppendynamiken

Viele Entscheidungen, die wir tagein, tagaus treffen müssen, haben auf die eine oder andere Art mit anderen Menschen und mit Gruppen zu tun. Dabei können wir in unterschiedliche Denkfallen tappen. Einige wichtige sind die nachfolgenden:

- ***Groupthink.*** Tendenziell glauben wir, dass mehr Menschen gemeinsam eine bessere Entscheidung treffen als eine Person alleine (Vier-Augen-Prinzip, etc.). Es kommt aber auch vor, dass Gruppen einer kollektiven Täuschung erliegen, getrieben von Optimismus, Harmoniebedürftigkeit und Konformismus [38]. Ein klassisches Beispiel für Groupthink ist die Invasion in der kubanischen Schweinebucht 1961, bei der die damalige US Regierung

von John F. Kennedy für eigentlich offensichtliche Schwachpunkte am Invasionsplan blind war.

- **Bandwagon Effect.** Wir alle lassen uns davon beeindrucken, was andere machen. Wenn wir beobachten, dass viele Menschen eine bestimmte Präferenz haben, tendieren wir, uns dieser Mehrheitsmeinung anzuschliessen [39]. Ein Beispiel sind Online-Rezensionen: Je mehr und je bessere Bewertungen ein Produkt hat, desto eher wird es gekauft.
- **Bystander Effect.** Wenn eine Person Hilfe braucht, helfen wir am ehesten dann, wenn keine anderen Personen anwesend sind. Wenn viele Leute anwesend sind, fühlt sich niemand direkt verpflichtet, zu helfen [40].
- **Spotlight Effect.** Wir alle haben instinktiv Angst, irgendwie negativ in der Öffentlichkeit aufzufallen. Darum überschätzen wir systematisch, wie viel Beachtung uns Menschen im öffentlichen Raum tatsächlich schenken [41]. Das klassische Beispiel ist das unangenehme Gefühl, dass die lachende Person im Restaurant sich über mich lustig macht, obwohl die Person in Tat und Wahrheit über etwas ganz anderes lacht.

6.5 Dinge, Menschen und Organisationen einschätzen

Einschätzungen über Menschen oder Organisationen oder sonstige Dinge zu treffen, ist alles andere als einfach. Wir nutzen darum intuitiv Heuristiken, die leider nicht immer zielführend sind. Einige der wichtigeren Biases in diesem Kontext sind die folgenden:

- **Stereotyping.** Wenn wir eine Person, die wir nicht kennen, einschätzen müssen, haben wir es mit grosser Ungewissheit, also mit unbekannten Faktoren zu tun. Um diese Ungewissheit zu reduzieren, nutzen wir intuitiv Stereotypisierung: Wir versuchen, aus den (vermeintlichen) Eigenschaften einer Gruppe, zu der die Person gehört, Eigenschaften der Person selber abzuleiten. Das Problem dabei ist, dass die Eigenschaften der Gruppe oftmals nicht evidenzbasiert, sondern reines irrationales Bauchgefühl sind [42]. Ein Beispiel: In Lohnverhandlungen wird aggressives Verhalten von Männern oft als Selbstsicherheit und Stärke interpretiert, bei Frauen hingegen als Zickigkeit und Überheblichkeit.
- **Halo Effect.** Wenn eine Person oder Organisation eine positive Eigenschaft hat, glauben wir intuitiv, dass auch andere Eigenschaften der Person oder

Organisation positiv sein müssen [43]. Ein klassisches Beispiel ist körperliche Schönheit: Bei gleicher Qualifikation haben schönere Menschen einen höheren Lohn und steigen schneller zu höheren Posten auf.

- **Intergroup Bias.** Das menschliche Denken ist fundamental ein Gruppendenken. Früher war jeder Mensch Mitglied in einem Stamm, der sich von den anderen Stämmen abgrenzen musste — der eigene Stamm war gut, die anderen waren böse. Das Überbleibsel davon ist unsere Tendenz, nach wie vor stark in Gruppen zu denken und unsere Innengruppe gegen die Aussengruppe abgrenzen zu wollen [44]. Darum sehen wir Mitglieder der Aussengruppe (z.B. einem anderen Team als dem eigenen im Unternehmen) als moralisch schlechter, als inkompetenter und ihre Motive und Ziele als verwerflich.
- **Attributionsfehler.** Wir neigen dazu, den Einfluss äusserer Faktoren auf das Verhalten bei anderen Menschen systematisch zu unterschätzen, während wir es bei unserem eigenen Verhalten überschätzen [45]. Wenn ich zum Beispiel einen Apfel stehle, tue ich das, weil ich in einer Krise stecke und mir Essen nicht leisten kann. Wenn jemand anderes aber einen Apfel stiehlt, so meine verzerrte Interpretation, tut die Person das, weil sie kriminell veranlagt ist.

6.6 Nutzen und Risiken

In der Theorie können wir jede Entscheidung rational treffen, indem wir den erwarteten Nutzen mit den erwarteten Risiken abwägen. In der Realität funktioniert dieses Kalkül aber nicht, wie die nachfolgenden Biases aufzeigen:

- **Omission Bias.** Menschen neigen dazu, Entscheidungen zu vermeiden, die zwar einen klaren positiven Effekt hätten, aber auch das Risiko bergen, einen kleineren negativen Effekt zu haben. Das ist der sogenannte Omission Bias [46], der Unterlassungsfehler. Ein Beispiel für den Unterlassungsfehler ist Ablehnung von Impfungen: Impfungen können Nebenwirkungen haben, aber in der Regel sind die Nebenwirkungen deutlich kleiner als die Risiken einer Infektion ohne Impfschutz. Der Omission Bias führt dazu, dass viele Menschen lieber passiv die grösseren Risiken einer Infektion ohne Impfschutz eingehen als aktiv die viel kleineren Risiken einer Impfung.

- **Verlustaversion.** Im Prinzip müssten wir kühle Optimiererinnen und Optimierer sein, die Dinge kaufen und dann wieder veräussern, wenn es sich lohnt. In der Realität aber steigt der subjektive Nutzen von etwas, wenn wir es besitzen — wenn wir etwas haben, geben wir es nicht gerne wieder weg [47]. Zum Beispiel zeigt sich, dass Menschen es als deutlich schlimmer empfinden, eine bestimmte Summe Geld zu verlieren, als sie es positiv finden, dieselbe Summe Geld zu gewinnen.
- **Hyperbolic Discounting.** Wir haben generell lieber den Spatz in der Hand als die Taube auf dem Dach: Der Nutzen einer Sache im Hier und Jetzt ist für uns meistens grösser als der erwartete Nutzen in der Zukunft, weil die Zukunft ungewiss ist. Bei der Berechnung des zukünftigen Nutzens sind wir aber nicht kohärent [48]. Zum Beispiel würden viele Leute lieber heute sofort 10 Franken kriegen als 11 Franken in einem Monat. Wenn man dieselben Leute aber fragt, ob sie lieber 10 Franken in einem Jahr oder 11 Franken in einem Jahr und einem Monat kriegen wollen, entscheiden sich viel mehr Menschen für die 11 Franken.
- **IKEA Effect.** Dadurch, dass wir etwas selber (mit-)erschaffen, steigt unsere Einschätzung des Wertes der Sache, die dabei herauskommt [49]. Ein Beispiel dafür sind IKEA-Möbel: Möbel wie jene von IKEA, die wir selber zusammenschrauben, schätzen wir das Möbel stärker, als wenn es ohne unser Zutun fertiggestellt würde.
- **Affektheuristik.** Wenn wir die Risiken von etwas einschätzen, spielt unser Affekt gegenüber der einzuschätzenden Sache eine grosse Rolle [50]. Wer zum Beispiel beim Thema Atomenergie vor allem einen negativen Affekt hat, schätzt die Risiken der Atomenergie deutlich höher ein als jemand, der Atomenergie gegenüber grundsätzlich positiv eingestellt ist.
- **Risikokompensation.** Risiken managen wir, damit die Eintrittswahrscheinlichkeit eines unerwünschten Ereignisses minimiert werden kann. Manchmal aber bewirkt Risikomanagement, dass wir uns zu sicher fühlen. In der Folge gehen wir viel mehr Risiken ein, als wir es ohne das Risikomanagement getan hätten, was die Eintretenswahrscheinlichkeit des unerwünschten Ereignisses letztlich wieder hebt. Ein Beispiel für diese Risikokompensation sind Velohelme: Menschen fahren bisweilen riskanter Fahrrad, wenn sie einen Helm haben, weil sie sich sicherer fühlen — mit der Folge, dass der Nutzen der Helme vom zusätzlichen risikofreudigen Verhalten überschattet wird [51].

6.7 Bestehende Überzeugungen

Etwas vom Schwierigsten, was es gibt, ist, die eigene Meinung zu revidieren. Es verwundert darum nicht, dass es zahlreiche Biases gibt, die uns dazu verführen, weiterhin das zu glauben, was wir glauben wollen. Einige dieser Biases sind die folgenden:

- **Status Quo Bias.** Der Mensch schätzt Bequemlichkeit über alles. So auch bei Entscheidungen: Am liebsten lassen wir alles beim Alten und ändern nichts. Nicht zuletzt, weil wir den Status Quo kennen und Veränderung immer ein gewisses Mass an Risiko enthält [52]. Ein typisches Beispiel für diesen Status Quo Bias sind Organspenden. In Ländern, in denen man aktiv einwilligen muss, um Organspenderin oder -spender zu werden, sind die Spenderaten viel tiefer als in Ländern mit dem umgekehrten System (Man ist automatisch Organspenderin oder -spender und muss aktiv widersprechen, wenn man nicht spenden will.).
- **Confirmation Bias.** Das, was wir bereits glauben, wollen wir auch weiterhin glauben — das ist, in einem Satz, der berühmte Confirmation Bias (Bestätigungsfehler) [53]. Der Confirmation Bias tritt auf unzählige Arten im Alltag auf. Ein Beispiel für den Confirmation Bias ist, wenn in einem Mitarbeitergespräch unbewusst nur negative Punkte zu einem Mitarbeiter gesucht werden, weil das Management vom Bauchgefühl her glaubt, dass der betroffene Mitarbeitende keinen guten Job machen kann.
- **Overconfidence.** Overconfidence (Selbstüberschätzung) ist unsere Tendenz, die eigene Meinung und die eigenen Entscheidungen zu überschätzen. Es gibt drei Varianten von Overconfidence: Overestimation (Überschätzung der eigenen Leistung), Overprecision (zu hohe Sicherheit, die Wahrheit zu kennen) und Overplacement (relative Selbstüberschätzung im Vergleich mit anderen Menschen oder Organisationen) [54]. Ein typisches Beispiel für Selbstüberschätzung als Overplacement ist die Frage an Autofahrerinnen und Autofahrer, wie sie ihre Fahrkenntnisse einschätzen: Viel mehr Menschen glauben, sie seien überdurchschnittlich gute Fahrerinnen und Fahrer, als mathematisch möglich ist.
- **Sunk Cost Fallacy.** Wir alle kennen das Gefühl, ein Projekt, das schon lange dauert, noch durchziehen zu wollen, weil wir schon so viel Ressourcen eingesteckt haben. Dieses Gefühl ist die sogenannte Sunk Cost Fallacy, der Trugschluss der versunkenen Kosten [55]. Wir glauben, dass wir noch

mehr investieren müssen, obwohl es besser wäre, das Projekt zu beenden und die weiteren Kosten zu vermeiden. Ein Beispiel für die Sunk Cost Fallacy sind glücklose Ehen, die nur aufrechterhalten werden, weil das Ehepaar schon so viele Jahre zusammen ist.

- ***Moralische Lizenzierung.*** Den Nutzen einer Entscheidung wägen wir oft nicht für sich alleine ab, sondern vergleichen, wie gut unsere früheren Entscheidungen waren. Wenn wir glauben, dass eine oder mehrere unserer früheren Entscheidungen gut waren, sehen wir das als Lizenz, jetzt eine schlechte Entscheidung treffen zu dürfen [56]. Ein typisches Beispiel hierfür ist die Extraportion Schokolade oder das zweite und dritte Bier, die wir uns gönnen, weil wir vorher eine Stunde lang Sport getrieben haben.
- ***Planungsfehlschluss.*** Wenn wir in die Zukunft planen, zum Beispiel für ein Projekt, haben wir die Tendenz, viel zu optimistisch zu sein und die tatsächlichen zeitlichen Anforderungen systematisch und bisweilen massiv zu unterschätzen [57]. Ein berühmtes Beispiel für diesen Planungsfehlschluss ist der Flughafen Berlin Brandenburg. Der neue Berliner Flughafen hätte 2011 eröffnet werden sollen. Aktuell wird die Eröffnung für 2021 geplant; 10 Jahre später als ursprünglich geplant.

Teil III

Rationalität stärken

7 Drei Ansätze für bessere Entscheidungen

Es gibt, wie in Teil II ausgeführt ist, zahlreiche Biases, die uns im Sinne von Denkfallen das Denken und Entscheiden schwer machen. Diesen Denkfallen sind wir aber nicht hilflos ausgeliefert: Der Umstand, dass wir etwas über die Natur dieser Denkfallen wissen, erlaubt uns, etwas gegen sie zu unternehmen. Das Ziel ist dabei nicht, uns selber oder unser Team oder unsere Organisation perfekt rational zu machen — perfekte Rationalität ist kein realistisch erreichbares Ziel. Es ist aber durchaus möglich, Rationalität bei einzelnen wichtigen Entscheidungen und in ausgewählten Kontexten zu steigern, um die Gefahr irrationaler Denkmuster und irrationaler Entscheidungen zu reduzieren. Zu diesem Zweck gibt es unterschiedliche Ansätze oder Methoden. Drei der bedeutendsten von ihnen sind *Nudging*, *Debiasing* und *Red Teaming*.

7.1 Nudging

Nudging kann am besten als eine Art kognitives Aikido verstanden werden. Bei der Kampfkunst Aikido steht nicht im Vordergrund, den Gegner aktiv anzugreifen. Das Ziel ist stattdessen, den Schwung und die Energie des Gegners gezielt gegen ihn auszunutzen. Bei Nudging geht es entsprechend nicht darum, irgendetwas an der Irrationalität von Menschen zu ändern. Stattdessen wird ihre Irrationalität gezielt ausgenutzt, um sie zu jenem Verhalten zu bewegen, das erwünscht und idealerweise auch für sie selber am besten ist. Das bedeutet konkret, dass ein Entscheidungskontext an die menschliche Irrationalität angepasst wird, sodass die Entscheidung ausfällt, wie es gewünscht ist. Nudging wird bisweilen auch als «Choice Architecture», als Entscheidungs-Architektur, beschrieben [58], weil bei Nudging nicht am Denkapparat der Betroffenen geschraubt wird, sondern an der Entscheidungssituation, in der sie sich befinden. Ein klassisches Beispiel für Nudging ist die Anordnung von Menu-Optionen in Schulkantinen: Wenn gesündere Nahrungsmittel besser sichtbar und einfacher zu erreichen sind als ungesunde, werden sie auch viel häufiger ausgewählt.

Ein Kernprinzip von Nudging ist, dass es wirklich bei Nudging, also bei sanftem Stupsen, bleibt: Mit Nudging werden weder Optionen gestrichen, noch wird irgendeine Form von Zwang eingeführt. Mit Nudging werden also alle Freiheiten in der konkreten Entscheidungssituation beibehalten [58]. Niemand wird zu etwas gedrängt, denn im Prinzip kann jede und jeder so entscheiden, wie sie oder er es vorher getan hätte.

Der vielleicht wesentliche Grund, warum Nudging ein attraktiver Ansatz

ist, ist die relative Einfachheit von Nudging-Massnahmen, mit denen Menschen positiv beeinflusst werden können. Im Vereinigten Königreich beispielsweise konnte mit einem einfachen Nudge das Verhalten bei der Steuererklärung verändert werden: Ein einfacher schriftlicher Hinweis, dass die Mehrheit der Leute ihre Steuern rechtzeitig bezahlen, hat stark beeinflusst, ob die Steuern rechtzeitig und vollständig bezahlt wurden [59]. Eine fast trivial einfache Intervention, die gesamthaft einen starken Effekt hatte.

Nudging hat aber auch einige Nachteile. Der wichtigste Nachteil ist der Umstand, dass Nudging Menschen nicht rationaler macht — Nudging nutzt lediglich die Irrationalität von Menschen aus, um sie zu einem gewünschten Verhalten zu bewegen. Nudging ist entsprechend auch nicht nachhaltig, denn ein Nudge wirkt genau dort, wo er eingesetzt wird, aber das Verhalten der Betroffenen wird nicht jenseits dieses Entscheidungskontextes verändert. Spezifische Nudges sind zudem auch sehr kontextgebunden. Ein Nudge, der für eine bestimmte Entscheidungssituation zugeschnitten wurde (z.B. das Zahlen von Steuern), kann nicht ohne Weiteres für eine andere Entscheidungssituation übernommen werden.

7.2 Debiasing

Debiasing ist ein Ansatz, der das macht, was der Name andeutet: Das Ziel ist, aktiv zu «entzerren», also die Wurzel von Irrationalität, unsere kognitiven Biases, aktiv zu tilgen. Debiasing ist also gewissermassen das Gegenteil von Nudging, denn mittels Debiasing wird Irrationalität nicht passiv ausgenutzt, sondern aktiv bekämpft, indem die individuelle Anfälligkeit für kognitive Biases aktiv reduziert wird. Während bei Nudging also der Entscheidungskontext verändert wird, werden bei Debiasing das Denken und die Entscheidungsfindung verändert.

Der grösste Vorteil von Debiasing ist dessen Nachhaltigkeit. Wenn es gelingt, die Art und Weise, wie Menschen (in bestimmten Situationen) denken, zu verändern, dann ist der Impact potenziell viel langlebiger als jener beim Nudging. Debiasing stimuliert nämlich die sogenannte *Metakognition*, das Denken über das Denken. Derartige Prozesse sind grundsätzlich sehr gut auf andere Denk- und Entscheidungssituationen übertragbar und haben damit also eine gewisse Universalität.

Debiasing hat aber auch Nachteile. Menschen aktiv zu verhelfen, weniger anfällig für kognitive Verzerrungen zu sein, ist aufwendig und nicht immer unmittelbar praxisbezogen. Auch kann der Aufwand von Debiasing schnell relativ

hoch werden, weil wirksame Debiasing-Massnahmen Zeit und viel Eigenarbeit erfordern. Weil die Thematik für die meisten Menschen neu und abstrakt ist, kann Debiasing auch innere Widerstände auslösen.

7.3 Red Teaming

Die Terroranschläge vom 11. September 2001 in den USA und die darauffolgende Invasion des Irak 2003 waren kolossale Fehlentscheide. Vor den Terroranschlägen und vor der Irak-Invasion waren alle nötigen Fakten vorhanden, um richtig zu entscheiden — die Terroranschläge hätten verhindert werden können, und die analytischen Fehler bei der Einschätzung der angeblichen irakischen Massenvernichtungswaffen und der Verbindung des Iraks zu islamistischem Terrorismus waren vermeidbar. Ein wesentlicher Faktor, warum diese zwei Mega-Fehler begangen wurden, sind kognitive Biases und Irrationalität. Bei den Terroranschlägen vom 11. September spielten Faktoren wie Selbstüberschätzung bei der Einschätzung der Risikolage und mangelnde Kooperation unter den unterschiedlichen Nachrichtendiensten aufgrund des Intergroup-Bias eine Rolle [60, 61]. Bei der Irak-Invasion hatten Faktoren wie ein starker Groupthink bei den Entscheidungsträgern in Kombination mit einem Confirmation Bias einen starken verzerrenden Effekt [62, 63, 64].

Dieses historische doppelte Versagen hat das US-Militär sowie verschiedene US-Nachrichtendienste dazu bewegt, die Fallstricke der menschlichen Irrationalität ernst(er) zu nehmen. Zu diesem Zweck haben sie den Ansatz des *Red Teaming* wiederbelebt [65]. Red Teaming bedeutet, Entscheidungen, Prozesse und Strategien systematisch zu hinterfragen, um blinde Flecken aufzudecken und Entscheidungen dadurch zu verbessern. Ein gutes Stichwort, um den Ansatz des Red Teaming zusammenzufassen, ist *Advocatus Diaboli*: Das Ziel von Red Teaming ist immer, den Status Quo der etablierten Mehrheitsmeinung zu durchbrechen und in den bestehenden Überzeugungen Schwachstellen zu suchen. Eines der ersten Länder, das Red Teaming fest verankert hat, ist Israel. Nachdem Israel 1973 vom Yom Kippur-Krieg überrascht wurde, obwohl alle Anzeichen auf einen Angriff deuteten, entschied der wichtigste israelische militärische Nachrichtendienst, das Risiko ähnlicher Fehlentscheide in Zukunft mittels fest institutionalisiertem Red Teaming zu reduzieren.

Der grösste Vorteil von Red Teaming ist der Umstand, dass Red Teaming fundamental praxisorientiert ist. Während Debiasing darauf abzielt, das Denken von Menschen zu verändern (was bisweilen ein wenig theoretisch daher kommen kann), ist Red Teaming sehr praktisch auf ein konkretes Problem fo-

kussiert, weil das Ziel einer jeden Red Teaming-Massnahme immer ist, eine konkrete Entscheidung, eine konkrete Strategie, einen konkreten Prozess usw. kritisch unter die Lupe zu nehmen. Es geht bei Red Teaming also nicht primär darum, über Biases zu reden und diese zu identifizieren; Biases werden im Red Teaming indirekt getilgt, indem praktische Sachverhalte hinterfragt werden. Als positive Nebenwirkung regt Red Teaming indirekt aber durchaus auch Metakognition an. Dadurch, dass Entscheidungsträgerinnen und Entscheidungsträger im Rahmen unterschiedlicher Red Teaming-Methoden mit ihren blinden Flecken konfrontiert werden, ist ein Denken über das eigene Denken oft eine fast zwangsläufige Konsequenz.

Ein wesentlicher Nachteil von, oder zumindest eine Herausforderung bei Red Teaming ist eine Art Huhn-Ei-Problem: Damit Red Teaming erfolgreich in einer Organisation eingesetzt werden kann, müssen die relevanten Entscheidungsträgerinnen und Entscheidungsträger den Wert von Red Teaming erkennen. Damit sie aber erkennen, warum Red Teaming nützlich ist, müssen sie bereits einen gewissen Grad an metakognitiver Reflektiertheit haben, der nicht *a priori* vorausgesetzt werden kann — diese Form der Reflektiertheit ist im Grunde die *Folge* von Red Teaming, welche sich als ein Kulturwandel in der Organisation manifestiert. Eine weitere Herausforderung in diesem Kontext ist, dass Red Teaming in vielen Organisationen z.B. aufgrund starrer Hierarchien oder sonstiger struktureller oder kultureller Faktoren eine Art Alibi-Übung sein kann. Wenn man Red Teaming einsetzt, aber alle Betroffenen unausgesprochen wissen, dass der Chef sich eh nicht hinterfragen lässt, bringt die ganze Übung nicht viel.

7.4 Nudging, Debiasing und Red Teaming im Vergleich

In Tabelle 2 sind die Ansätze des Nudging, Debiasing und Red Teaming vergleichend zusammengefasst.

Es gibt nicht den einen einzigen richtigen Ansatz. Nudging, Debiasing und Red Teaming eignen sich für unterschiedliche Problemstellungen und Kontexte. *Nudging* eignet sich als Ansatz, wenn das Ziel ist, relativ viele Menschen in einer ganz konkreten Entscheidungssituation positiv zu beeinflussen, und zwar ohne Anspruch auf metakognitive Veränderung in irgendeiner Art. *Debiasing* eignet sich als Ansatz, wenn das Ziel ist, positive metakognitive Veränderungen bei einem kleineren Zielpublikum zu bewirken, mit der Folge, dass bestimmte Entscheidungen oder Prozesse nachhaltig verbessert werden. *Red Teaming* eignet sich als Ansatz, wenn das Ziel ist, ganz konkrete und bedeutende Entschei-

Tabelle 2: *Tabellarischer Vergleich von Nudging, Debiasing und Red Teaming.*

	Nudging	Debiasing	Red Teaming
Optimierungsziel	Verhalten	Denken	Denken und Verhalten
Angriffsfläche	Verhalten	Denken	Denken und Verhalten
Publikumsgrösse	hoch	tief bis mittel	tief bis mittel
Metakognition	keine	mittel bis hoch	mittel bis hoch
Praxisbezug	hoch	tief bis mittel	hoch
Nachhaltigkeit	tief	mittel bis hoch	mittel bis hoch
Übertragbarkeit	tief	hoch	mittel bis hoch
Effektivität	tief bis mittel	mittel	mittel bis hoch
Aufwand	tief bis mittel	mittel bis hoch	tief bis mittel

dungen, Prozesse, Strategien und dergleichen kritisch zu durchleuchten und zu challengen, mit der Option, sie komplett zu verwerfen.

In den nachfolgenden Abschnitten sind einige konkrete Techniken bzw. Methoden für Nudging, Debiasing und Red Teaming beschrieben.

8 Generelle Nudging-Techniken

Ein Nudge, der in einer spezifischen Situation eingesetzt wird, kann nicht ohne Weiteres auf andere Situationen übertragen werden. Es gibt aber allgemeine Nudging-Techniken oder -Prinzipien, die für unterschiedliche konkrete Einsatzbereiche und Kontexte angepasst werden können. Einige dieser Nudging-Techniken sind nachfolgend in fünf Kategorien zusammengefasst: *Exposure*, *Choice Architecture* (im engeren Sinn), *Signalling*, *Information Design* und *Motivation Hacking*.

Einige Exposure-Nudges sind in Tabelle 3 zusammengefasst. Das Grundprinzip von Exposure-Nudges ist, Dinge sichtbar und auffällig zu machen.

Tabelle 3: *Exposure-Nudges.*

Nudge	Beschreibung	Beispiel
Mere Exposure	Je öfter wir mit etwas in Kontakt kommen, desto besser finden wir es.	Werbung: Wir sehen wiederholt ein Produkt und wollen es kaufen, weil wir es kennen.
Priming	Ein vorangehender Reiz kann beeinflussen, wie wir nachfolgende Reize verarbeiten.	Wir riechen im Supermarkt frisch gebackenes Brot und entscheiden spontan, Brot zu kaufen.
Picture Superiority	Bilder bleiben stärker im Gedächtnis als Wörter.	Kunden erinnern sich eher an eine Firma, die ein bildhaftes Logo hat.
Konditionierung	Konditioniertes Verhalten gezielt ausnutzen.	Wenn ein Ampelsystem-Label bei Nahrung rot anzeigt, schreckt das ab.

Einige Choice Architecture-Nudges sind in Tabelle 4 zusammengefasst¹⁵. Das Grundprinzip von Choice Architecture-Nudges ist, ganz konkrete Entscheidungsoptionen so zu gestalten, dass die Entscheidung eher so ausfällt, wie es gewünscht ist.

Einige Signalling-Nudges sind in Tabelle 5 zusammengefasst. Das Grundprinzip der Signalling-Nudges ist, gewisse Informationen und Aufforderungen implizit angedeutet werden.

Einige Information Design-Nudges sind in Tabelle 6 zusammengefasst. Das Grundprinzip der Information Design-Nudges ist, dass Information so präsentiert wird, dass sie einen gewünschten Effekt erzielt.

Einige Motivation Hacking-Nudges sind in Tabelle 7 zusammengefasst. Das Grundprinzip der Motivation Hacking-Nudges ist, die intrinsische Einstellung

¹⁵Der Begriff Choice Architecture ist hier in einem engeren Sinn verwendet: Es geht um Nudges, bei denen der Umgang mit unterschiedlichen Optionen im Vordergrund steht. Weiter oben im Unterabschnitt 7.1 wird Choice Architecture als allgemeine Beschreibung für Nudging verwendet.

Tabelle 4: Choice Architecture-Nudges.

Nudge	Beschreibung	Beispiel
Sticky Defaults	Standardoptionen werden wegen des Status Quo-Bias oft beibehalten.	Organspende: Bei Opt-Out spenden viel mehr Leute Organe als bei Opt-In.
Anordnungseffekte	Die Anordnung von Optionen beeinflusst unsere Wahrnehmung und Präferenzen.	Wenn die erste Option auf einer Speisekarte relativ gesehen die teuerste ist, wirken alle anderen günstig.
Vereinfachung	Zu viele Optionen frustrieren und machen unglücklich. Optionen bündeln und vereinfachen macht glücklich	Ein Webshop zeigt zunächst eine Auswahl an Suchergebnissen; auf Wunsch werden weitere Ergebnisse eingeblendet.

oder Motivation gezielt zu beeinflussen.

Tabelle 5: Signalling-Nudges.

Nudge	Beschreibung	Beispiel
Bandwagon Effect	Wir richten uns danach, was andere machen.	Wir buchen online eher ein Hotel, wenn es gute Nutzer-Bewertungen hat.
Halo Effect	Eine positive Eigenschaft sehen wir als Anzeichen für weitere positive Eigenschaften.	Attraktive Mitarbeitende werden bei objektiv gleicher Kompetenz als kompetenter in ihrem Job wahrgenommen als weniger attraktive.
Authority Bias	Wir richten uns nach der Meinung von Experten und Autoritätsfiguren.	Wenn in Zahnpasta-Werbung Schauspieler als Zahnärzte verkleidet sind, glauben wir, die Zahnpasta sei gut.
Scarcity Effect	Was selten ist, ist begehrt, weil es gefragt scheint.	Wenn «nur noch wenige» Stücke eines Produktes übrig sind, greifen wir eher zu.
Surveillance Bias	Wenn wir uns beobachtet fühlen, verhalten wir uns regelkonform und prosozial.	Wenn auf Mülleimern Augen aufgemalt sind, gibt es weniger Littering.
Soziale Norm	Wir orientieren uns an sozialen Normen.	«9 von 10 Leuten zahlen ihre Steuern fristgerecht.» bewegt uns dazu, Steuern pünktlich zu bezahlen.

Tabelle 6: Information Design-Nudges.

Nudge	Beschreibung	Beispiel
Framing	Wie Information präsentiert wird, beeinflusst die Wahrnehmung und Verarbeitung der Information.	«Du hast 100 Franken und verlierst 20» wird subjektiv als schlimmer empfunden als «Du hast 100 Franken und es bleiben dir 80».
Decoy Effect	Asymmetrische Optionen können Präferenzen unverhältnismässig stark beeinflussen.	Wenn Klein 5 Franken, Mittel 8 Franken und Gross 10 Franken kostet, tendieren wir zu Gross (weil es «sich lohnt»).
Anchoring	Die erste Information bleibt besonders stark hängen.	Sonderangebote: Wenn der Originalpreis von 100 Franken nun auf 70 gesenkt wurde, wirkt 70 im Vergleich zum Anker günstig.

Tabelle 7: Motivation Hacking-Nudges.

Nudge	Beschreibung	Beispiel
Affekt-Heuristik	Unser Affekt beeinflusst unsere Präferenzen und unsere Risikowahrnehmung.	Werbung vermittelt uns den Nutzen von Zigaretten mit schönen Bildern. Darum unterschätzen wir die Risiken.
Portionierung	Mehrere kleine Ziele motivieren mehr als ein grosses Ziel.	3 x 5 Liegestützen fallen uns leichter als 1 x 15 Liegestützen, weil wir mehrere kleine Teilziele erreichen.
Countdowns	Das Frustempfinden beim Warten wird durch Countdowns abgebaut.	Countdowns bei Bus- und Tramstationen haben einen besänftigenden und beruhigenden Effekt.
Gamification	Spielerische Anreize bewegen unaufdringlich zu bestimmten Handlungen.	Die Fliege im Pissoir: Männer zielen beim Pinkeln auf die Fliege (und verursachen dadurch weniger Verunreinigung).
Curiosity Gap	Genug Info zum neugierig machen; zu wenig Info, um schon befriedigt zu sein.	Clickbait: «Die 10 besten Katzenbilder aller Zeiten. Nummer 5 ist schockierend!»

9 Debiasing-Techniken

Bei Debiasing gibt es zwei grundsätzliche Stossrichtungen [66]. Zum einen gibt es Debiasing-Massnahmen, welche direkt und primär auf die Ebene der Metakognition abzielen. Diese Massnahmen haben also zum Ziel, Denken ganz direkt positiv zu beeinflussen. Diese Familie der Debiasing-Massnahmen wird bisweilen als «metacognitive forcing» beschrieben, weil ihr Ziel ist, in einer Entscheidungssituation das Denken über das Denken anzuregen [67]. Zum anderen gibt es Debiasing-Massnahmen, welche eher auf die Prozessebene abzielen und Metakognition zwar auch fördern, aber dafür prozess- und kontextorientierten Werkzeugen arbeiten.

Einige Debiasing-Techniken, welche direkt Metakognition ansprechen, sind die nachfolgenden:

- **Lernen.** Es klingt trivial, aber Lernen ist gewissermassen eine zentrale Debiasing-Massnahme, um das Denken zu verändern. Je mehr und besser eine Person über Irrationalität und kognitive Biases Bescheid weiss, desto eher ist sie bereit, über das eigene Denken zu denken und eigene Denkmuster und Entscheidungen zu hinterfragen.
- **Workshops und Anleitungen.** Workshops und Anleitungen stellen eine Form des zielgerichteten, angeleiteten Lernens dar. Der Vorteil solcher Formate ist, dass damit beispielsweise eine Gruppe an Mitarbeitenden in einer Organisation gemeinsam erreicht werden kann. Diese Form der Debiasing-Massnahme kann einen unmittelbaren sowie einen nachhaltigen Effekt haben [68]. Der Nachteil ist aber, wie bei sonstigen Workshops und Anleitungen, der Aufwand für die Vorbereitung, Durchführung und Nachbereitung, sowie auch die nicht immer gegebene intrinsische Motivation oder Willigkeit der Teilnehmenden, sich wirklich offen mit der Thematik auseinanderzusetzen.
- **Explizite Begründungen.** In Anlehnung an die Diskussion zur Definition von Rationalität in Teil I ist eine Debiasing-Massnahme, bei Einschätzungen und Entscheidungen aktiv nach Gründen zu fragen. Diese Massnahme kann sogar als eine Art Mikro-Intervention einen gewissen positiven Effekt zeitigen, indem wir uns selber die Frage stellen: «Warum?» [69]. Die Aufforderung, eine Begründung zu formulieren, garantiert nicht, dass die Begründung qualitativ hochstehend wird, aber diese Massnahme erhöht die Wahrscheinlichkeit, dass wir uns Gedanken machen.

- **Überzeugung quantifizieren.** Eine weitere erkenntnistheoretisch inspirierte Debiasing-Massnahme ist, die Stärke einer Überzeugung explizit als Wahrscheinlichkeit von 0 bis 1 (oder 0 bis 100) zu quantifizieren. Eine Überzeugung auf diese Art zu quantifizieren leitet automatisch einen metakognitiven Prozess ein, weil das Quantifizieren von Überzeugungen nicht etwas ist, was wir spontan aus dem Bauch heraus machen können — um auf eine Zahl zu kommen, müssen wir aktiv nachdenken. Quantifizierte Überzeugungen sind zudem auch eine belastbare Grundlage, um in Teams Entscheidungen zu diskutieren und aktiv zu treffen.
- **Consider the Opposite.** Eine wirksame Massnahme, um bei automatisierten Entscheidungen die metakognitive Komponente zu aktivieren, ist die Consider the Opposite- oder Consider an Alternative-Strategie [70] (auf Deutsch «erwäge das Gegenteil» und «erwäge Alternativen»). Consider the Opposite und Consider the Alternative sind nützlich, weil sie uns dazu bewegen, die kognitiven Scheuklappen abzulegen und aktiv andere Erklärungen oder Hypothesen oder Szenarien in Betracht zu ziehen. Ein Beispiel für die praktische Umsetzung dieser Massnahme ist der Prozess der Differenzialdiagnose in der Medizin, bei der aktiv Unterschiedliche abduktive Erklärungen (Diagnosen) für die beobachteten Symptome in Betracht gezogen werden.
- **Principle of Charity.** Eine häufige Folge irrationaler Biases ist, dass Menschen das Verhalten anderer Menschen negativer interpretieren, als es angebracht ist. Zum Beispiel verfallen die Konfliktparteien in Konfliktsituationen schnell in einen Confirmation Bias, bei dem alles, was die Gegenseite macht, negativ interpretiert wird. Ein Weg, um solche Dynamiken zu verhindern, ist das Principle of Charity (Prinzip des Wohlwollens) [71]. Das Principle of Charity besagt, dass wir Äusserungen und Handlungen des Gegenüber maximal wohlwollend interpretieren sollen, weil wir nicht ausschliessen können, dass der Grund für unsere negative Einschätzung ein «Übersetzungsfehler» ist (zum Beispiel in Form eines kognitiven Bias unsererseits). Mit dem Principle of Charity kann innerhalb von Sekunden ein fast totaler Perspektivenwechsel stattfinden.
- **Debunking.** Unsere Denkfehler machen uns anfällig für unterschiedliche Formen der Falschinformation: Gerüchte, Verschwörungstheorien, Fake News, gezielt gestreute Desinformation, und so fort. Um gegen Falschinformation korrigierend vorzugehen, werden bisweilen Debunking-Interventionen

eingesetzt, um die Falschinformation zu entkräften. Klassisches Debunking bedeutet, korrekte Informationen als Korrektiv gegenüber der Falschinformation aufzuzeigen. Der positive Impact klassischen Debunkings, das gegen spezifische Falschinformation vorgeht, ist aber limitiert [72]. Eine wirksamere Methode von Debunking ist, nicht bloss korrigierende Informationen, sondern die Logik- und Denkfehler bei Falschinformationen aufzuzeigen [73, 74]. Der zusätzliche Vorteil dieser Debunking-Methode ist, dass sie universalen Charakter hat und die allgemeine Anfälligkeit für Falschinformation senkt.

- **Prebunking.** Prebunking ist Debunking, das *vorbeugend* eingesetzt wird. Prebunking-Interventionen sind somit Debiasing-Massnahmen gegen Falschinformation, die zum Einsatz kommen, bevor bestimmte Falschinformationen allzu beliebt werden, z.B. auf Social Media viral gehen und sich in den Köpfen festsetzen können. Die Logik von Prebunking ist damit jene einer kognitiven Impfung: Indem Menschen vorbeugend gegen Falschinformation gestärkt werden, sinkt ihre Anfälligkeit für Falschinformation, wenn sie ihr in der Praxis begegnen. In zahlreichen Studien hat sich Prebunking als hochwirksame Debiasing-Intervention gezeigt; besonders, wenn sie für universale Logik- und Denkfehler sensibilisieren [75, 76].

Einige Debiasing-Techniken, die mehr auf die Prozess- oder Kontextebene des Entscheidens abzielen, sind die folgenden:

- **Checklisten.** Checklisten sind eine enorm wirksame Massnahme, um in hektischen und vielschichten Prozessen Fehler zu vermeiden [77]. Checklisten zwingen uns, über Dinge nachzudenken, über die wir nicht aktiv nachdenken, oder bei denen wir aufgrund von Faktoren wie Overconfidence oder Confirmation Bias davon ausgehen, dass alles ist, wie wir es uns wünschen. Checklisten haben in vielen Domänen die Fehlerraten drastisch gesenkt, etwa in der Luftfahrt (Piloten gehen vor jedem Flug eine Checkliste durch) oder in der Medizin.
- **Entschleunigung.** Kognitive Verzerrungen schlagen dann besonders oft zu, wenn wir schnell und automatisiert denken. Wenn Entscheidungsprozesse aktiv verlangsamt werden, sinkt entsprechend die Wahrscheinlichkeit, irrational zu entscheiden. Entschleunigung ist eine fast triviale Massnahme, die aber bemerkenswerten Impact zeitigen kann. Ein Bereich, wo Entschleunigungs-Massnahmen erfolgreich eingesetzt werden, ist Medizin [78].

- **Anonymisierung.** Anonymisierung kann helfen, den Einfluss subjektiver, irrationaler Bewertungen zum Beispiel des Aussehens, des Geschlechts oder der Herkunft auszuschliessen. Ein klassisches Beispiel für erfolgreiche Anonymisierung sind Orchester. Wenn Orchester anonyme Bewerbungsverfahren einführen, bei denen die Künstlerinnen und Künstler «versteckt» vorspielen (also ohne, dass ihr Geschlecht erkennbar ist), werden im Schnitt fünf- bis sechs Mal mehr Frauen als bei einem offenen Vorspielen eingestellt [79].
- **Commitment Devices.** Wenn wir heute aktiv und bewusst unsere zukünftigen Handlungsoptionen einschränken, um irrationale Entscheidungen zu vermeiden, setzen wir ein sogenanntes Commitment Device, also einen «Verpflichtungs-Apparat», ein [80]. Ein Commitment Device beruht auf der rationalen Einsicht, dass gewisse zukünftige Handlungsoptionen dazu führen könnten, dass wir schlecht entscheiden. Wenn wir uns heute rationalerweise entscheiden, diese zukünftigen Optionen zu eliminieren, reduzieren wir das Risiko. Ein triviales Beispiel: Wenn ich weiss, dass ich nachts immer Süssigkeiten esse, obwohl ich weiss, dass das nicht gut für mich ist, dann ist ein wirksames Commitment Device, zuhause gar keine Süssigkeiten zu haben.

10 Red Teaming-Techniken

Red Teaming ist nicht einfach eine Ansammlung von Techniken, sondern Primär eine Einstellung: Red Teaming bedeutet, dass jemand in einer Organisation jene Fragen stellt, welche sonst niemand stellt, um jene Probleme aufzudecken, welche lieber ignoriert werden [81]. Das bedeutet, dass der wohl schwierigste und bedeutendste Teil von Red Teaming darin besteht, das Prinzip des Red Team überhaupt in einer Organisation zum Einsatz kommen zu lassen. Um das zu bewerkstelligen, ist in der Regel viel Überzeugungsarbeit notwendig.

Für die konkrete Red Teaming-Arbeit gibt es zahlreiche Techniken, die je nach Problemstellung und Kontext eingesetzt werden können [82, 83, 84]. Einige nützliche Red Teaming-Techniken sind nachfolgend aufgelistet:

- **Reversal Test.** Der Reversal Test ist eine philosophische Technik, um aufzudecken, ob Widerstand gegen eine bestimmte Veränderung durch den Status Quo Bias bedingt ist [85]. Wenn sich Entscheidungsträgerinnen und Entscheidungsträger gegen eine bestimmte Veränderung sträuben, wird als Gedankenexperiment eine Umkehrung der Veränderung vorgeschlagen: Wie wäre es, wenn wir genau das Gegenteil dessen, was vorgeschlagen ist, machen? Falls es auch Widerstand gegen die umgekehrte Veränderung gibt, kann das darauf hindeuten, dass ein Status Quo Bias vorliegt. Dass der Status Quo nämlich im ganzen Spektrum der möglichen Veränderungen die beste Option ist, ist statistisch gesehen eher unwahrscheinlich. Wenn die Geschäftsleitung zum Beispiel die Marketing-Abteilung nicht um 5 Personen vergrössern will, aber auch der Meinung ist, dass eine Verkleinerung um 5 Personen falsch wäre, muss begründet werden, warum der Status Quo die beste Option darstellt.
- **Premortem.** Nach Abschluss (oder dem vorzeitigen Scheitern) von Projekten wird bisweilen eine «Postmortem»-Analyse durchgeführt, um festzuhalten, was warum wie schief ging. Die Premortem-Technik ist ein Postmortem, das sozusagen auf den Kopf gesetzt wird: Ein Premortem wird *vor* Projektbeginn eingesetzt, um zu erarbeiten, was im Projekt alles schief gehen könnte [86]. Im Rahmen einer Premortem-Analyse werden zunächst Erfahrungswerte aus früheren, ähnlichen Projekten bezüglich der Frage, was warum schief ging, zusammengetragen. Danach werden in einer kreativen Ideenphase alle möglichen Gründe zusammengetragen, warum welche Dinge beim bevorstehenden Projekt schief gehen könnten. Abschliessend werden die Prioritäten bei den potenziellen Stolpersteinen bestimmt,

sowie Massnahmen definiert, um diese Stolpersteine zu eliminieren.

- ***Fail on Purpose.*** Bei der Technik Fail on Purpose geht es darum, in einer Trockenübung zu überlegen, wie ein Projekt oder eine Organisation o.ä. so schnell wie möglich zum totalen Scheitern gebracht werden kann. Fail on Purpose aktiviert schnell kreative Kräfte, und die Technik zeigt auf, welche der Dinge, die zum Scheitern führen würden, schon heute im Alltag des Projektes oder der Organisation vorkommen (ggf. in abgeschwächter Form). Fail on Purpose zeigt somit indirekt Schwachstellen und blinde Flecken auf.
- ***Black Swan Analysis.*** Black Swan Events sind Ereignisse, die sehr unwahrscheinlich sind, aber einen grossen positiven oder negativen Impact haben. Die kreative Suche nach Black Swan-Ereignissen öffnet den Horizont von Teams und Organisationen massiv und erlaubt, über Risiken und Chancen nachzudenken, die sonst gar nicht auf dem Radar sind.
- ***Key Assumptions Check.*** Jeder Strategie, jedem Projekt und letztlich jeder Entscheidung liegen einige wesentliche Annahmen oder Überzeugungen zugrunde. Diese bleiben aber oftmals unausgesprochen oder unhinterfragt. In einem Key Assumptions Check werden die Kernannahmen einer bestimmten Entscheidungssituation explizit gemacht, um sie in einem zweiten Schritt kritisch zu hinterfragen. Das offenbart so gut wie immer Lücken und Schachstellen.
- ***Epistemic Shift.*** In Abschnitt 9 wurde das Quantifizieren von Überzeugungen in Form von Wahrscheinlichkeiten als einfache, aber wirksame Debiasing-Technik beschrieben. Das Quantifizieren von Überzeugungen kann auch als Red Teaming-Technik eingesetzt werden. Einerseits kann diese Quantifizierung zum Beispiel im oben beschriebenen Key Assumptions Check eingesetzt werden, um die zentralen Überzeugungen so präzise wie möglich zu beschreiben. Die Quantifizierung von Überzeugungen kann aber auch über die Zeit angewendet werden, um aufzuzeigen, wie sich Überzeugungen wandeln. Die einfachste Variante dieser epistemischen Verschiebung ist eine Vorher-Nachher-Anordnung. Zum Beispiel können HR-Verantwortliche ihre Einschätzung über einen Bewerber zunächst vor dem Bewerbungsgespräch quantifizieren, sowie dann erneut im Anschluss an das Bewerbungsgespräch. Das offenbart, ob sich etwas verändert hat, und es erlaubt eine kritische Diskussion, ob die Veränderung gut begründet ist.

- ***Analysis of Competing Hypotheses.*** Wenn wir Erklärungen für Probleme suchen oder Prognosen für zukünftige Entwicklungen machen, stellen wir Hypothesen auf. Die Gefahr dabei ist, dass wir uns zu schnell in eine Hypothese hineinkaprizieren und andere mögliche Hypothesen ignorieren. Die Analyse konkurrierender Hypothesen ist eine Technik, bei der zunächst explizit alle vorhandenen und ggf. weitere unausgesprochene mögliche Hypothesen definiert werden, um in einem zweiten Schritt die Plausibilität aller Hypothesen zu prüfen (Wofür gibt es Evidenz? Welche Hypothesen lassen sich widerlegen?).
- ***Advocatus Diaboli.*** *Advocatus Diaboli* ist das Grundprinzip von Red Teaming, aber auch eine konkrete Technik. *Advocatus Diaboli* bedeutet, dass die vorherrschende Konsens-Meinung systematisch hinterfragt wird, indem ein Argumentarium für eine alternative Meinung oder Erklärung oder Sichtweise erstellt wird. Das Ziel von *Advocatus Diaboli* ist also, eine bestehende Überzeugung zu widerlegen und eine alternative Position als plausibler und wahrscheinlicher darzustellen. Diese Technik kann erstaunlich effektiv sein, weil Konsens-Meinungen in Organisationen oftmals als gegeben angesehen werden und nie kritisch hinterfragt werden.
- ***A-Team und B-Team.*** Die Technik des A-Team und B-Team bedeutet, dass mehrere Teams geschaffen werden, die aktiv unterschiedliche Ansätze verfolgen, um beispielsweise ein Problem zu erklären oder zu lösen. Der Vorteil dieser Technik ist, dass nicht eine dominante Konsens-Sichtweise erschaffen wird, die im Nachhinein hinterfragt wird. Stattdessen gibt es von Anfang an eine Art Wettbewerb der Sichtweisen, was es in der analytischen Auswertungsphase einfacher macht, die unterschiedlichen Ergebnisse miteinander zu vergleichen, weil Faktoren wie der Status Quo Bias oder der Confirmation Bias ausgeschaltet sind. Die A-Team und B-Team-Technik ist verhältnismässig aufwendig und eignet sich für grössere Projekte und bedeutende Problemstellungen.
- ***What If? Analysis.*** Die Was wenn?-Analyse ist eine Kreativtechnik für die Strategieplanung. Sie ist verwandt mit Fail on Purpose und der Black Swan-Analyse. Bei der Was wenn?-Analyse werden zunächst Ereignisse definiert, welche als unwahrscheinlich oder unerwünscht erachtet werden. In einem Gedankenexperiment ist dann die Ausgangslage, dass die Ereignisse eingetreten sind. Der analytische Teil dieser Technik besteht dann darin, über den Modus des «Reverse Engineering» nachzuzeichnen,

wie diese unerwarteten Ereignisse zustande hätten kommen können. Obwohl diese Technik sehr spekulativ ist, erlaubt sie einen schnellen Perspektivenwechsel, der Dinge aufzeigt, welche bisher nicht auf dem Radar waren.

- ***Kill your Organization.*** Bei der Kill your Organization-Technik werden in einem kreativen Workshop-Format ein oder mehrere fiktive Konkurrenten erschaffen. Diese Konkurrenten sind so viel besser, dass sie die eigene Organisation platt machen. Der Nutzen dieser Technik: Die Faktoren, welche in diesem Gedankenexperiment die fiktiven Konkurrenten überlegen machen, sind reale Schwächen und blinde Flecken der eigenen Organisation. Damit diese Technik erfolgreich ist, müssen z.B. die Moderierenden des Workshops dafür sorgen, dass die Teilnehmenden die Scheuklappen des operativen Alltags erfolgreich ablegen und Faktoren und Szenarien in Betracht ziehen, welche sie nicht unmittelbar auf dem Radar haben.
- ***Pro und Kontra.*** Eine Pro- und Kontra-Liste ist eine der einfachsten Red Teaming-Techniken. Eine Pro- und Kontra-Liste bedeutet, in einer Entscheidungssituation alle Argumente für und alle Argumente wider eine bestimmte Entscheidung zu verschriftlichen. Das erlaubt es, die Argumente dafür und dagegen gegeneinander abzuwägen, und zwar sowohl quantitativ als auch qualitativ.
- ***Erwartungswert.*** Der Erwartungswert ist eine Kennzahl, die sich aus dem erwarteten Nutzen einer Entscheidung multipliziert mit der Wahrscheinlichkeit, dass der Nutzen eintritt, ergibt. Die Berechnung von Erwartungswerten eignet sich sehr gut in riskanten Entscheidungssituationen mit unterschiedlichen Alternativen. Zudem zwingt die Quantifizierung des erwarteten Nutzens sowie der Eintretenswahrscheinlichkeit dazu, für diese Werte gute Gründe zu haben. Dadurch kann beispielsweise ein Team feststellen, ob überhaupt alle Teammitglieder einen ähnlichen Nutzen in den unterschiedlichen Entscheidungsmöglichkeiten sehen.
- ***50k-Experiment.*** Das 50k-Experiment ist eine einfache Technik, mit der Probleme und Problem-Priorisierung aufgezeigt werden. Teammitglieder oder Entscheidungsträgerinnen und Entscheidungsträger (abhängig vom konkreten Kontext) erhalten fiktive 50'000 Franken. Sie können komplett frei bestimmen, wofür sie dieses Geld in ihrer Organisation einsetzen. Im Anschluss an den kreativen Teil werden die Ergebnisse gesammelt

und verglichen. Das zeigt auf, ob Prioritäten grundsätzlich ähnlich gesetzt werden, oder, ob komplett unterschiedliche Problemwahrnehmungen bestehen.

- **Alternative Futures Analysis.** Die Ungewissheit der Zukunft ist die zentrale Herausforderung jeder strategischen Planung. Die Alternative Futures-Technik hilft in solchen Planungsprozessen, indem gezielt unterschiedliche Zukunftsszenarien erarbeitet werden. Die Ausgangslage für die Alternative Futures-Analyse ist dabei aber nicht, wie z.B. bei der Was wenn?-Analyse, das bereits vordefinierte zukünftige Endergebnis. Stattdessen werden die wichtigsten Faktoren der gegenwärtigen Situation, welche einen Einfluss auf zukünftige Ergebnisse und Ereignisse haben können, identifiziert. Anhand der identifizierten Faktoren werden dann Schritt für Schritt unterschiedliche Szenarien gebaut, je nach Ergebnis oder Ausprägung der Faktoren (Z.B. kommen Faktoren A und B in Szenario 1 vor; in Szenario 2 nur Faktor A; in Szenario 3 nur Faktor B; usw.). Wenn eine erschöpfende Anzahl an Zukunftsszenarien erarbeitet wurde, erhalten die Szenarien abschliessend einen Wahrscheinlichkeitswert, mit dem bestimmt wird, wie die Zukunft eher oder eher nicht aussehen wird. Diese Technik ist verhältnismässig zeitaufwendig, ist aber eine sehr effektive Art, um strategisch wichtige Faktoren und Ungewissheiten zu identifizieren.
- **Base Rate Building.** Für die Entscheidungsfindung ist es sehr wichtig, Base Rates, also Ausgangswahrscheinlichkeiten, zu haben. Im operativen Alltag kommen Base Rates aber praktisch nie zum Einsatz. Base Rate Building ist eine Technik, die z.B. in Kombination mit dem Key Assumptions Check eingesetzt werden kann. Das Ziel der Technik ist, alle relevanten Grundwahrscheinlichkeiten zu identifizieren und zu quantifizieren. Ein triviales Beispiel: Wenn eine Verkaufsabteilung überlegt, wie sie ihre Verkaufsstrategie optimieren soll, ist es wichtig, zu wissen, was die Erfolgsquoten je nach Kundensegment in der eigenen Organisation sowie bei Konkurrenzorganisationen sind.
- **SWOT-Aanalyse.** Die SWOT-Analyse ist ein bekanntes Werkzeug, dass auch im Rahmen von Red Teaming eingesetzt werden kann. SWOT steht für Strengths, Weaknesses, Opportunities, Threats. Strengths und Weaknesses sind dabei die Stärken und die Schwächen der eigenen Organisation oder des Projektes oder Produktes (je nach Einsatzgebiet). Opportunities und Threats sind Chancen und Risiken der Aussenwelt, die wir selber

nicht beeinflussen können, die aber einen Einfluss auf unser Vorhaben haben. Die SWOT-Analyse ist ein einfaches analytisches Werkzeug, um den Tunnelblick der eigenen Stärken (und damit den Optimism Bias und Overconfidence) abzulegen.

- **Stopping Rules.** In vielen Bereichen des beruflichen (und privaten) Lebens führt die Sunk Cost Fallacy dazu, dass Projekte oder Zustände oder sonstige Dinge immer weiter laufen, obwohl es sinnvoller wäre, sie abzubrechen. Stopping Rules ist eine Technik, mit der vor Beginn beispielsweise eines Projektes explizit bestimmt wird, wann das Projekt gestoppt wird. Die Anhalte-Regeln sollen dabei möglichst quantifizierbare Kriterien berücksichtigen.
- **Bias Hunting.** Bias Hunting ist eine spielerische Red Teaming-Technik, mit der in einem bestimmten Entscheidungskontext aktiv nach kognitiven Biases «gejagt» wird. Das kann beispielsweise in Form eines Workshops stattfinden, bei dem mehrere Gruppen eine Liste oder ein Kartenset ein Biases durchgehen und für jeden Bias einschätzen, ob und wie oft und mit welchen Auswirkungen er in dem Entscheidungskontext, welcher Gegenstand des Workshops ist, vorkommt. In der Synthese werden die Ergebnisse der unterschiedlichen Gruppen zusammengeführt und verglichen.
- **First Principles.** Wenn es darum geht, Probleme zu lösen, Sachverhalte zu verstehen, oder Prognosen und Schätzungen zu machen, nutzen die meisten Menschen und Organisationen Analogien. Wenn es zum Beispiel um die Frage geht, was für technologische Risiken es in Zukunft geben wird, wird in der Regel einfach eine Analogie zu bestehenden Risiken gemacht — was es heute gibt, wird in die Zukunft hinein extrapoliert. Analogien sind aber sehr limitiert und limitierend, denn dadurch entsteht weder wahre Innovation noch können Probleme damit aus neuen, bisher unbekannten Blickwinkeln betrachtet werden. First Principles (erste Prinzipien) als Gegenteil zu Analogien sind fundamentale, elementare Wahrheiten (oder Überzeugungen oder Axiome), die einer Fragestellung zugrundeliegen. Indem First Principles als essenzielle Bausteine einer Fragestellung identifiziert werden, können unterschiedliche und komplett neuartige Lösungsstrategien, Erklärungen und Vorgehensweisen erarbeitet werden, wie sie mit dem blossen Denken in Analogien unmöglich wären.

11 Zusammenfassung

Rationalität ist die Fähigkeit, Ziele sinnvoll zu verfolgen. Diese Fähigkeit setzt sich aus der epistemischen (etwas nur mit guten Gründen glauben) und der instrumentellen (nutzenmaximierend entscheiden) Rationalität zusammen. Was in der Theorie einfach klingt, ist in der Praxis eine grosse Herausforderung: Menschen sind systematisch irrational. Ein wesentlicher Grund dafür sind kognitive Biases (kognitive Verzerrungen): Automatisierte Denkmuster, mit denen wir uns selber täuschen. Das Problematische an kognitiven Biases ist, dass Menschen dann, wenn sie von ihnen betroffen sind, nicht merken, dass dem so ist.

Im Kampf gegen Irrationalität gibt es drei wichtige Ansätze: *Nudging*, *Debiasing* und *Red Teaming*. *Nudging* bedeutet, Irrationalität auszunutzen, indem der Entscheidungskontext so verändert wird, dass Menschen sich eher so verhalten, wie es gewünscht ist. Nudging macht Menschen nicht rationaler, denn Irrationalität wird durch Nudging lediglich ausgenutzt, nicht ausgemerzt. *Debiasing* ist ein Ansatz, der Irrationalität aktiv reduziert, indem die Anfälligkeit für kognitive Biases gesenkt wird. Debiasing funktioniert entweder über Massnahmen, welche Metakognition (das Denken über das Denken) aktiv stimulieren, oder über Massnahmen, welche stärker die Prozessebene des Entscheidens beeinflussen und Metakognition indirekt anregen. *Red Teaming* ist ein Ansatz, der Irrationalität ebenfalls aktiv reduziert, aber im Fokus steht dabei nicht das individuelle Denken, sondern vielmehr konkrete praktische Entscheidungen, Prozesse, Strategien und dergleichen. Das Kernprinzip von Red Teaming ist *Advocatus Diaboli*: Systematisch Fragen stellen und Positionen einnehmen, die nicht gestellt und eingenommen werden, um gewohnte Denk- und Entscheidungsmuster zu durchbrechen.

Alle drei Optimierungsansätze haben Vor- und Nachteile. Die Ansätze verfolgen unterschiedliche Ziele und eignen sich für unterschiedliche Problemstellungen und Kontexte. Zu entscheiden, welcher Ansatz und welche Techniken in der Praxis angebracht sind, ist die verantwortungsvolle, aber auch kreative und vielfältige Kernaufgabe von Rationalitäts-Praktikerinnen und -Praktikern.

Literatur

- [1] David C. Logan. Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. *Journal of Experimental Botany*, 60(3):712–714, March 2009.
- [2] Ray Pawson, Geoff Wong, and Lesley Owen. Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy. *American Journal of Evaluation*, 32(4):518–546, December 2011.
- [3] Keith E. Stanovich. On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In *The Oxford handbook of thinking and reasoning*, Oxford library of psychology, pages 433–455. Oxford University Press, New York, NY, US, 2012.
- [4] David Z. Hambrick and Alexander P. Burgoyne. The Difference Between Rationality and Intelligence. *The New York Times*, September 2016.
- [5] Galen V. Bodenhausen. Stereotypes as Judgmental Heuristics: Evidence of Circadian Variations in Discrimination. *Psychological Science*, 1(5):319–322, September 1990.
- [6] Matthias Steup. Epistemology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition, 2018.
- [7] Duncan Pritchard. Epistemic Luck. *Journal of Philosophical Research*, 29:191–220, July 2004.
- [8] Niko Kolodny and John Brunero. Instrumental Rationality. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition, 2018.
- [9] Matthew Kopec. A pluralistic account of epistemic rationality. *Synthese*, 195(8):3571–3596, August 2018.
- [10] William Talbott. Bayesian Epistemology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.
- [11] Richard Foley. The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly*, 29(2):111–124, 1992.

- [12] Itzhak Gilboa, Andrew W. Postlewaite, and David Schmeidler. Probability and Uncertainty in Economic Modeling. *Journal of Economic Perspectives*, 22(3):173–188, September 2008.
- [13] Keiran Sharpe. On Risk and Uncertainty, and Objective versus Subjective Probability. *Economic Record*, 94(S1):49–72, 2018.
- [14] Cassiano Terra Rodrigues. The Method of Scientific Discovery in Peirce’s Philosophy: Deduction, Induction, and Abduction. *Logica Universalis*, 5(1):127–164, April 2011.
- [15] Douglas Walton. Why Fallacies Appear to Be Better Arguments than They Are. SSRN Scholarly Paper ID 1759289, Social Science Research Network, Rochester, NY, 2010.
- [16] Gerd Gigerenzer and Wolfgang Gaissmaier. Heuristic Decision Making. *Annual Review of Psychology*, 62(1):451–482, 2011.
- [17] Keith Frankish. Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass*, 5(10):914–926, October 2010.
- [18] Daniel Kahneman. Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, 93(5):1449–1475, 2003.
- [19] Martie G. Haselton and Daniel Nettle. The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. *Personality and Social Psychology Review*, 10(1):47–66, February 2006.
- [20] Gerd Gigerenzer. *Gut feelings: the intelligence of the unconscious*. Viking, New York, 2007.
- [21] Amos Tversky and Daniel Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, September 1974.
- [22] List of cognitive biases, October 2019. Page Version ID: 921897950.
- [23] Maya Bar-Hillel. The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3):211–233, May 1980.
- [24] Amos Tversky and Daniel Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315, 1983.

- [25] William M. Lepley. "The Maturity of the Chances": A Gambler's Fallacy. *The Journal of Psychology*, 56(1):69–72, July 1963.
- [26] James Sundali and Rachel Croson. Biases in Casino Betting: The Hot Hand and the Gambler's Fallacy. *ResearchGate*, 1(July):1–12, February 2006.
- [27] Amos Tversky and Daniel Kahneman. Judgments of and by Representativeness. Technical Report TR-3, STANFORD UNIV CA DEPT OF PSYCHOLOGY, May 1981.
- [28] Marko Kovic and Tobias Fuchslin. Probability and conspiratorial thinking. *Applied Cognitive Psychology*, 32(3):390–400, May 2018.
- [29] Adrian Furnham and Hua Chu Boo. A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1):35–42, February 2011.
- [30] Benedetto De Martino, Dharshan Kumaran, Ben Seymour, and Raymond J. Dolan. Frames, Biases, and Rational Decision-Making in the Human Brain. *Science*, 313(5787):684–687, August 2006.
- [31] Y Kuzyakov, J. K Friedel, and K Stahr. Review of mechanisms and quantification of priming effects. *Soil Biology and Biochemistry*, 32(11):1485–1498, October 2000.
- [32] Bennet B. Murdock Jr. The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5):482–488, 1962.
- [33] Amos Tversky and Daniel Kahneman. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2):207–232, September 1973.
- [34] Stephen J. Brown, William Goetzmann, Roger G. Ibbotson, and Stephen A. Ross. Survivorship Bias in Performance Studies. *Review of Financial Studies*, 5(4):553–580, October 1992.
- [35] Loren J. Chapman. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 6(1):151–155, February 1967.
- [36] Catherine S. Fichten and Betty Sunerton. Popular Horoscopes and the "Barnum Effect". *The Journal of Psychology*, 114(1):123–134, May 1983.
- [37] Neal J. Roese and Kathleen D. Vohs. Hindsight Bias. *Perspectives on Psychological Science*, 7(5):411–426, September 2012.

- [38] Roland Bénabou. Groupthink: Collective Delusions in Organizations and Markets. *The Review of Economic Studies*, 80(2):429–462, April 2013.
- [39] Rüdiger Schmitt-Beck. Bandwagon Effect. In *The International Encyclopedia of Political Communication*, pages 1–5. American Cancer Society, 2015.
- [40] Peter Fischer, Joachim I. Krueger, Tobias Greitemeyer, Claudia Vogrincic, Andreas Kastenmüller, Dieter Frey, Moritz Heene, Magdalena Wicher, and Martina Kainbacher. The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, 137(4):517–537, 2011.
- [41] Thomas Gilovich, Victoria Husted Medvec, and Kenneth Savitsky. The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one’s own actions and appearance. *Journal of Personality and Social Psychology*, 78(2):211–222, 2000.
- [42] Mahzarin R. Banaji and Curtis D. Hardin. Automatic Stereotyping. *Psychological Science*, 7(3):136–141, May 1996.
- [43] E.L. Thorndike. A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1):25–29, 1920.
- [44] Miles Hewstone, Mark Rubin, and Hazel Willis. Intergroup Bias. *Annual Review of Psychology*, 53(1):575–604, 2002.
- [45] Daniel T. Gilbert and Patrick S. Malone. The correspondence bias. *Psychological Bulletin*, 117(1):21–38, 1995.
- [46] Siu Kit Yeung, Tijen Yay, and Gilad Feldman. Action and Inaction in Moral Judgments and Decisions: Meta-Analysis of Omission Bias Omission-Commission Asymmetries. *Personality and Social Psychology Bulletin*, 48(10):1499–1515, October 2022.
- [47] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1):193–206, February 1991.
- [48] David Laibson. Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, 112(2):443–478, May 1997.
- [49] Michael I. Norton, Daniel Mochon, and Dan Ariely. The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, 22(3):453–460, July 2012.

- [50] Paul Slovic, Melissa L. Finucane, Ellen Peters, and Donald G. MacGregor. The affect heuristic. *European Journal of Operational Research*, 177(3):1333–1352, March 2007.
- [51] Ross Owen Phillips, Aslak Fyhri, and Fridulv Sagberg. Risk Compensation and Bicycle Helmets. *Risk Analysis*, 31(8):1187–1195, 2011.
- [52] William Samuelson and Richard Zeckhauser. Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1):7–59, March 1988.
- [53] Raymond S Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175, 1998. bibtex: nickerson1998confirmation.
- [54] Don A. Moore and Paul J. Healy. The trouble with overconfidence. *Psychological Review*, 115(2):502–517, 2008.
- [55] Hal R Arkes and Catherine Blumer. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35(1):124–140, February 1985.
- [56] Anna C. Merritt, Daniel A. Effron, and Benoît Monin. Moral Self-Licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass*, 4(5):344–357, 2010.
- [57] Roger Buehler, Dale Griffin, and Michael Ross. Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions. In Thomas Gilovich, Dale Griffin, and Daniel Kahneman, editors, *Heuristics and Biases*, pages 250–270. Cambridge University Press, Cambridge, 2002.
- [58] Richard H Thaler, Cass R S Unstein, and John P Balz. Choice Architecture, 2012.
- [59] Michael Hallsworth, John A. List, Robert D. Metcalfe, and Ivo Vlaev. The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31, April 2017.
- [60] Charles F. Parker and Eric K. Stern. Blindsided? September 11 and the Origins of Strategic Surprise. *Political Psychology*, 23(3):601–630, September 2002.

- [61] Roderick M. Kramer. A Failure to Communicate: 9/11 and the Tragedy of the Informational Commons. *International Public Management Journal*, 8(3):397–416, December 2005.
- [62] Mark Phythian. The Perfect Intelligence Failure? U.S. Pre-War Intelligence on Iraqi Weapons of Mass Destruction. *Politics & Policy*, 34(2):400–424, 2006.
- [63] Dr David Patrick Houghton. Invading and Occupying Iraq: Some Insights From Political Psychology. *Peace and Conflict: Journal of Peace Psychology*, 14(2):169–192, April 2008.
- [64] Dina Badie. Groupthink, Iraq, and the War on Terror: Explaining US Policy Shift toward Iraq: Groupthink, Iraq, and the War on Terror. *Foreign Policy Analysis*, 6(4):277–296, September 2010.
- [65] Micah Zenko. *Red team: how to succeed by thinking like the enemy*. Basic Books, a member of the Perseus Books Group, New York, 2015.
- [66] Gideon Keren. Cognitive Aids and Debiasing Methods: Can Cognitive Pills Cure Cognitive Ills? In Jean-Marc Fabre and Michel Gonzalez Jean-Paul Caverni, editor, *Advances in Psychology*, volume 68, pages 523–552. North-Holland, 1990.
- [67] Pat Croskerry. Cognitive forcing strategies in clinical decisionmaking. *Annals of Emergency Medicine*, 41(1):110–120, January 2003.
- [68] Susan T. Fiske, Carey K. Morewedge, Haewon Yoon, Irene Scopelliti, Carl W. Symborski, James H. Korris, and Karim S. Kassam. Debiasing Decisions: Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioral and Brain Sciences*, 2(1):129–140, October 2015.
- [69] Marko Kovic. Debiasing in a Minute or Less, or Too Good to Be True? The Effect of Micro-Interventions on Decision-Making Quality. *Psych*, 1(1):220–239, June 2019.
- [70] Edward R. Hirt and Keith D. Markman. Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, 69(6):1069–1086, 1995.
- [71] Willard Van Orman Quine. *Word and Object*. MIT Press, 1964.

- [72] Man-pui Sally Chan, Christopher R. Jones, Kathleen Hall Jamieson, and Dolores Albarracín. Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. *Psychological Science*, 28(11):1531–1546, November 2017. Publisher: SAGE Publications Inc.
- [73] Philipp Schmid and Cornelia Betsch. Effective strategies for rebutting science denialism in public discussions. *Nature Human Behaviour*, 3(9):931–939, September 2019. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 9 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Decision making;Education;Human behaviour Subject_term_id: decision-making;education;human-behaviour.
- [74] Stephan Lewandowsky, John Cook, Ullrich Ecker, Dolores Albarracin, Michelle Amazeen, Panayiota Kendou, Doug Lombardi, Eryn J. Newman, Gordon Pennycook, Ethan Porter, David G. Rand, David N. Rapp, Jason Reifler, Jon Roozenbeek, Philipp Schmid, Colleen M. Seifert, Gale M. Sinatra, Briony Swire-Thompson, Sander van der Linden, Emily K. Vraga, Thomas J. Wood, and Maria S. Zaragoza. The Debunking Handbook 2020. Technical report, Center for Climate Change Communication, George Mason University, Fairfax, Virginia, 2020.
- [75] Sander van der Linden, Jon Roozenbeek, Rakoem Maertens, Melisa Basol, Ondřej Kácha, Steve Rathje, and Cecilie Steenbuch Traberg. How Can Psychological Science Help Counter the Spread of Fake News? *The Spanish Journal of Psychology*, 24, 2021. Publisher: Cambridge University Press.
- [76] Stephan Lewandowsky and Sander van der Linden. Countering Misinformation and Fake News Through Inoculation and Prebunking. *European Review of Social Psychology*, 0(0):1–38, February 2021. Publisher: Routledge _eprint: <https://doi.org/10.1080/10463283.2021.1876983>.
- [77] Brigitte M. Hales and Peter J. Pronovost. The checklist—a tool for error management and performance improvement. *Journal of Critical Care*, 21(3):231–235, September 2006.
- [78] Carol-anne Moulton, Glenn Regehr, Lorelei Lingard, Catherine Merritt, and Helen MacRae. Slowing Down to Stay Out of Trouble in the Operating Room: Remaining Attentive in Automaticity. *Academic Medicine*, 85(10):1571, October 2010.

- [79] Christy Krawietz. Overture to Equality: Preventing Subconscious Sex and Gender Biases from Influencing Hiring Decisions, An. *Seattle University Law Review*, 39:1051, 2015.
- [80] Gharad Bryan, Dean Karlan, and Scott Nelson. Commitment Devices. *Annual Review of Economics*, 2(1):671–698, 2010.
- [81] Susan Craig. Reflections from a Red Team Leader. *Military Review*, 87(2):57, March 2007.
- [82] Bryce G. Hoffman. *Red teaming: transform your business by thinking like the enemy*. Piatkus, London, 2017. OCLC: ocn953597921.
- [83] Monique Kardos and Patricia Dexter. A Simple Handbook for Non-Traditional Red Teaming. 2017.
- [84] Red Teaming Guide. Technical report, Development, Concepts and Doctrine Centre, Shrivenham, 2013.
- [85] Nick Bostrom and Toby Ord. The Reversal Test: Eliminating Status Quo Bias in Applied Ethics. *Ethics*, 116(4):656–679, July 2006.
- [86] Gary Klein. Performing a project premortem. *Harvard Business Review*, 85(9):18–19, 2007.