

Wie gut schlagen wir uns als Internistinnen und Internisten?

# Systematische Denkfehler im klinischen Alltag

Levy Jäger<sup>a</sup>, Stefan Markun<sup>a</sup>

<sup>a</sup> Institut für Hausarztmedizin, Universitätsspital Zürich, Zürich

Dieser Artikel ist eine Zusammenfassung des gleichnamigen Workshops am SGAIM-Herbstkongress 2020. Der Artikel umfasst theoretische Grundlagen sowie die präsentierten Fallvignetten und deren Lösungen inklusive der Antworten von 48 Ärztinnen und Ärzten, die am Workshop teilgenommen haben. Für eine tiefe Auseinandersetzung und Selbsttestung empfehlen die Autoren, die Fragen und Vignetten (Link im Kasten auf dieser Seite) noch vor dem Weiterlesen dieses Artikels zu beantworten.

## Einleitung / Hintergrund

Irren ist menschlich. Dennoch gibt es zahlreiche Situationen, in denen Fehler schwerwiegende Konsequenzen nach sich ziehen können – besonders dann, wenn es um Patientinnen und Patienten geht.

Die Ursache von Fehlern im ärztlichen Alltag liegt meistens nicht etwa in fehlender Kompetenz oder unzureichender Intelligenz, sondern man findet oft eine einfache, einleuchtende und nicht ganz so entmutigende Erklärung. Sie stammt aus den Ergebnissen der Forschung in der kognitiven Psychologie der letzten Jahrzehnte und bildet die Anwendung eines Modells, das unsere Denkprozesse allgemein sehr gut zu beschreiben vermag. Gemäss dieser *dualen Prozess-theorie* lassen sich Entscheidungsprozesse als das Zusammenspiel von zwei kognitiven Systemen beschreiben [1]. Erste, schnelle Schlussfolgerungen werden vom intuitiven «System 1» gezogen. Dieses unterliegt nur bedingt unserer bewussten Kontrolle, erfordert aber geringe Anstrengung. Feinere Überlegungen gehen dagegen auf das eher langsame, analytische «System 2» zurück, das uns zwar mehr Mühen kostet, dafür aber bewusster zugänglich ist. System 1 ist dafür verantwortlich, dass wir beim ersten Patientenkontakt bereits gewisse Muster erkennen, die uns gleich zu einer Arbeitsdiagnose führen. Besonders in Notfallsituationen ist das sicher von grossem Nutzen. Diese schnellen Schlüsse können uns aber auch in die Irre führen. Hier kommt System 2 ins Spiel, das durch rationale Überlegungen die von System 1 getrof-

Die am Workshop präsentierten Fragen und Fallvignetten sind für die Leser/-innen unter diesem Link verfügbar: <https://de.surveymonkey.com/r/ihamz2020>.



fenen Entscheidungen modulieren und eventuell korrigieren kann. Dabei ist es keineswegs so, dass uns das System 1 nur auf falsche Fährten locken will. Ganz im Gegenteil: Wir haben es ihm zu verdanken, dass wir die meisten Situationen im Alltag ohne allzu grosse Anstrengung bewältigen können. Es stellt uns einfache Heuristiken zur Verfügung, die wir auf einen Grossteil unserer Entscheidungen erfolgreich anwenden können.

Idealerweise würden wir Entscheidungen stets über eine Kalibrierung der beiden Systeme treffen. Nicht immer reichen aber unsere Energiereserven aus, um das eher anstrengende System 2 herbeizuziehen. Der anstrengende Klinikalltag, Konflikte mit Mitarbeitenden oder zu wenig Schlaf können schon dazu ausreichen, unseren logischen Apparat ausser Gefecht zu setzen. Schlussfolgerungen unterliegen dann weniger einer Modulation durch System 2, und es besteht vermehrt die Gefahr von Fehlern. Es gilt also: Fehler passieren uns oft nicht etwa aus Inkompetenz, sondern vielmehr, weil die Energiereserven unseres Denkapparats beschränkt sind.

Wir werden anhand der am Workshop präsentierten Fragen und Fallvignetten verschiedene Denkfehler

vorstellen, die im klinischen Alltag häufig vorkommen. Dafür wollen wir zuerst zwei Grundkategorien definieren:

- Unter einem **Fehlschluss** (*fallacy*) versteht man jegliche Verwendung einer ungültigen Schlussfolgerung in einer Argumentationskette. Wir konzentrieren uns dabei nur auf Fehlschlüsse, die nicht direkt mit dem Inhalt der dabei verwendeten Argumente zu tun haben. Insbesondere wollen wir hier auf Fehlschlüsse eingehen, die mit Fehleinschätzungen und Fehlinterpretationen der Häufigkeiten bestimmter Ereignisse einhergehen.
- Von einer **kognitiven Verzerrung** (*bias*) spricht man, wenn ein systematischer Fehler in Denk- und Entscheidungsprozessen vorliegt, der von den Regeln rationalen Denkens abweicht.

Seit Einführung dieser Begriffe durch die bahnbrechenden Arbeiten der Psychologen Tversky und Kahneman konnten dafür in verschiedensten Bereichen zahlreiche Beispiele identifiziert werden [2]. Mittlerweile sind Denkfehler in der klinischen Medizin Gegenstand aktiver Forschung und werden immer wieder in umfassenden Übersichten präsentiert (z.B. [3]).

## Fragen und Vignetten

### 1. Aufwärmübungen (CRT)

Die Fragen 1–3 sind Elemente des sogenannten *Cognitive Reflection Test* (CRT) [4]. Der CRT testet die Neigung von Probanden, ihre schnelle, intuitive Antwort (System 1) bewusst zu reflektieren (System 2).

#### Frage 1.1

«Ein Schläger und ein Ball kosten zusammen 110 Schweizer Franken. Der Schläger kostet 100 Franken mehr als der Ball. Wie viele Franken kostet der Ball?»

Bei dieser Frage bildet System 1 rasch die Differenz zwischen 110 und 100 und drängt auf, dass der Ball 10 Franken kosten muss. System 2 rechnet nach und stellt fest: Wenn der Ball 10 Franken kostet, dann kostet der Schläger aber 90 Franken mehr als der Ball und nicht 100 Franken, wie in der Aufgabestellung definiert. Die korrekte Lösung, bei der sowohl die Summe von 110 als auch die Differenz von 100 eingehalten werden, ist: Ball = 5 Franken, Schläger = 105 Franken.

Die korrekte Antwort wurde von 48% der Teilnehmenden gegeben.

#### Frage 1.2

«Wenn 5 Maschinen 5 Minuten benötigen, um 5 Objekte herzustellen, wie viele Minuten brauchen 100 Maschinen, um 100 Objekte herzustellen?»

Bei dieser Frage erkennt System 1 die konstanten Proportionen zwischen der Anzahl Maschinen, der Anzahl Objekte und der Herstellungszeit (5,5,5) und schlägt eine analoge Lösung vor (100,100,100). System 2 erkennt jedoch, dass die Herstellungszeit nicht zunimmt, wenn die Anzahl Maschinen im gleichen Ausmass wie die zu produzierenden Objekte zunimmt. Die korrekte Lösung ist somit: 100 Maschinen brauchen für 100 Objekte ebenfalls 5 Minuten (100,100,5).

Die korrekte Antwort wurde von 73% der Teilnehmenden gegeben.

#### Frage 1.3

«Auf einem See befindet sich ein Geflecht aus Seerosen. Jeden Tag verdoppelt das Geflecht seine Grösse. Wenn es 48 Tage dauert, bis das Geflecht den gesamten See bedeckt hat, wie lange braucht es dann für die Hälfte des Sees?»

System 1 reagiert auf das Wort «verdoppelt» sowie die Aufgabe einer Rückrechnung mit einer Halbierung und schlägt  $48/2=24$  als Lösung vor. System 2 stellt fest, dass die Verdoppelung täglich ist und das halbe Ausmass somit korrekterweise am Tag  $48-1=47$  vorgelegen haben muss.

Die korrekte Antwort wurde von 69% der Teilnehmenden gegeben.

### 2. Selbsteinschätzung und Gegenüberstellung zum CRT

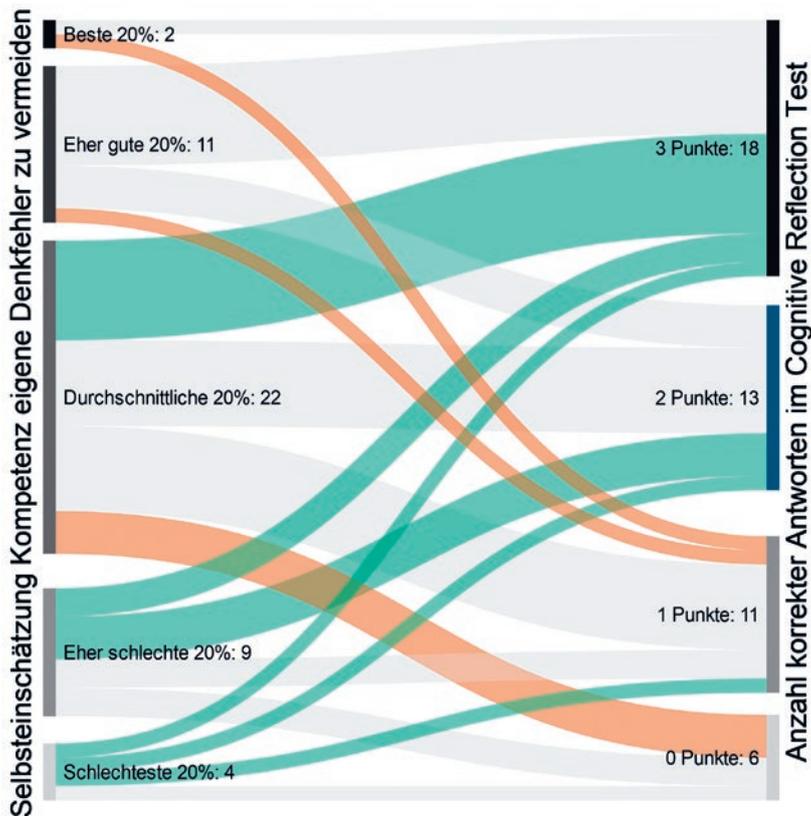
«Wie gut sind Sie im Vermeiden von eigenen Denkfehlern (Selbsteinschätzung im Vergleich mit «durchschnittlichen» Workshopeteilnehmenden).» Zur Auswahl standen fünf Quintile aus dem Bereich «zu den besten 20%» bis «zu den schlechtesten 20%».

Unter der Annahme einer perfekten Selbsteinschätzung im Vergleich mit anderen müssten sich die Teilnehmenden in gleich grosse Gruppen einschätzen. Tatsächlich schätzten sich 46% als «zu den durchschnittlichen 20%» ein. Die Selbsteinschätzungen einzelner Teilnehmenden wurden nachträglich von uns mit der Anzahl korrekter Antworten (Punkte) im CRT gegenübergestellt (Abb. 1). Hier zeigte sich bezüglich des Abschneidens im CRT, dass sich, 31% der Teilnehmenden unterschätzt und 10% überschätzt haben.

### 3. Klinische Vignetten

#### Frage 3.1

«In einer grossen urbanen Praxis existiert seit zehn Jahren ein COPD-Screeningprogramm für >60 Jahre alte Raucher oder Ex-Raucher mit >40 Pack-Years. Es ist aus langjähriger Erfahrung bekannt, dass bei den gescreenten Patientinnen und Patienten



**Abbildung 1:** Sankey-Diagramm der Antworten auf die Frage nach Selbsteinschätzung der Kompetenz eigene Denkfehler zu vermeiden (Proportionen repräsentiert durch die Höhe der linken vertikalen Balken), sowie Gegenüberstellung mit der Anzahl korrekter Antworten im *Cognitive Reflection Test* (Proportionen repräsentiert durch die Höhe der rechten vertikalen Balken). Die Breite der Verbindungen zwischen linken und rechten Balken ist proportional zur Anzahl Teilnehmenden mit entsprechenden Antwortkombinationen (rote Verbindungen stehen für Selbstüberschätzung, grüne Verbindungen für Selbstunterschätzung).

in 50% aller Fälle eine COPD nachweisbar ist. Beim Screening der letzten vier Fälle wurde viermal in Folge keine COPD festgestellt. Wie hoch schätzen Sie die Wahrscheinlichkeit für eine COPD beim nächsten gescreenten Fall?»

System 1 vereinfacht die Situation und geht intuitiv von einer überschaubaren Gesamtmenge an Patienten aus, aus welcher Stichproben gezogen wird. In einer kleinen Gesamtmenge wäre der Pool an COPD-Patienten tatsächlich erschöpfbar und es liesse sich aus vorhergehenden Diagnosen auf zukünftige Diagnosewahrscheinlichkeiten schliessen. In obigen Fall ist man verleitet anzunehmen, dass beim nächsten Patienten eine COPD weniger wahrscheinlich ist als 50%. In der klinischen Realität ist der Pool an Patienten mit einer bestimmten Diagnose allerdings sehr gross und aufgrund von wenigen Beobachtungen oder einer kurzfristigen Häufung lassen sich keine sicheren Prognosen für zukünftige Patienten machen. Im obigen Fall ist die Wahrscheinlichkeit aufgrund der langjährigen

Erfahrung somit weiterhin 50%. Der Fehlschluss, der hier gemacht wird, heisst «*Gambler's Fallacy*» und ist ähnlich zum Roulettespieler, der nach mehrmaligem Auftauchen einer Farbe auf die Gegenfarbe setzt, um vermeintlich seine Gewinnchance zu steigern.

Am Workshop haben 85% der Teilnehmenden diese Vignette korrekt beantwortet.

### Frage 3.2

«Eine 72-jährige Frau wird von ihrer Tochter in eine Praxis gebracht, weil sie heute Morgen kurzzeitig nicht mehr wusste, welches Jahr wir haben. Die Tochter hatte unter der Hypothese einer Dehydratation ihrer Mutter viel zu trinken gegeben, worauf sie wieder zeitlich orientiert war. Allerdings fiel eine Urininkontinenz auf. Die MPA nimmt sogleich einen Urinstreifentest vor. Dieser zeigt eine Leukozyturie und ist positiv auf Nitrit. Körpertemperatur, Blutdruck, Puls der Patientin sind normwertig. Die Patientin ist aktuell zeitlich, örtlich, zur Person und Situation orientiert. Die Nierenlager zeigen keine Klopfdolenz. Die Patientin hat einen HWI. Nach zusätzlichem Ausschluss einer relevanten Inflammation durch normwertiges CRP und unauffälliges Blutbild ist eine ambulante Behandlung mit Antibiotika mit Verlaufskontrolle in ein bis drei Tagen ausreichend. Stimmen Sie diesem Vorgehen zu?»

Diese Vignette enthält Fallstricke auf zwei Stufen. Auf der ersten Stufe präsentiert sich die Patientin schon bei der MPA mit einem suggestiven Muster. System 1 hat bereits verfügbare Befunde, die für einen Harnwegsinfekt sprechen. System 1 versucht dann noch weitere vorhandene Befunde zu generieren, welche die Hypothese stützen (positiver Urinstreifentest), ohne jedoch weitere Befunde zu erheben, welche die Hypothese infrage stellen würden (z.B. Hypercalcämie-Testung). Diese selektive Berücksichtigung von bestätigender Information auf Kosten wiederlegender Information wird *Confirmation Bias* genannt. Auf der zweiten Stufe folgt die ärztliche Präsentation mit einer vorgespurten Diagnose. Die weiteren Abklärungen dienen nur noch der Bestimmung des Schweregrades der vorgespurten Diagnose, alternative Diagnosen werden weiterhin nicht in Betracht gezogen. Dieses Einrasten in eine bestimmte Spur aufgrund einer vorhergehenden (Fehl-)Information wird als *Anchoring* bezeichnet. Am Workshop waren 42% der Teilnehmenden mit dem Vorgehen in dieser Vignette einverstanden (Antibiotika unter der Hypothese eines HWI ohne weitere Ursachensuche bei Verwirrung). Die Konstruktion dieser Vignette erlaubt allerdings keine sichere Unterscheidung zwischen einer unbewussten Unterlassung (Fehlschluss) und einer bewussten Unterlassung (z.B. wegen zu geringer Vortestwahrscheinlichkeit) von weiteren Abklärungen.

### Frage 3.3

«Eine 25-jährige Patientin hat sich bei Ihnen zur Sprechstunde einschreiben lassen wegen fluktuierender Müdigkeit seit mehreren Jahren. Sie ist engagierte Tierschutzaktivistin, Veganerin

und macht viel Sport. Was ist bei dieser Patientin wahrscheinlicher: 'Sie hat eine Anämie' oder 'Sie hat eine Anämie mit Eisenmangel'?)»

Diese Vignette enthält Risikofaktoren für einen Eisenmangel. System 1 beschwört umgehend dieses Stereotyp und vernachlässigt, dass Eisenmangelanämien nur eine Subgruppe aller möglichen Anämien sind. Die Gesamtheit der nicht näher bezeichneten Anämien ist somit wahrscheinlicher als die Subgruppe der Eisenmangelanämien. Diese fälschliche Erwartung eines spezifischen Stereotypes anstelle eines weniger spezifischen Falles wird als *Conjunction Fallacy* bezeichnet.

Am Workshop haben 38% der Teilnehmenden diese Vignette korrekt beantwortet.

### Frage 3.4

«Bei der Patientin aus Frage 3.3 zeigen hinreichende Abklärungen weder eine Anämie (Hb 13,5 g/dl), noch einen Eisenmangel (Ferritin 64 ng/ml), noch sonstige objektivierbare Müdigkeitsursachen. Die Patientin wurde bereits vor einem Jahr mit ähnlichen Untersuchungsergebnissen abgeklärt, insbesondere fand sich ein Ferritin von 70 ng/ml und es wurden keine spezifischen Massnahmen eingeleitet. Da die Müdigkeit aber besonders stark war, suchte die Patientin auf eigene Initiative einen Eisenspezialisten auf. Dieser verabreichte ihr eine Eiseninfusion, worauf die Müdigkeit sich deutlich verbesserte. Die Patientin meint, dies spreche eindeutig für eine Wirkung der Eiseninfusion gegen die Müdigkeit. Stimmen Sie dieser Ansicht zu?»

Aus der unkontrollierten Selbstbeobachtung der Patientin lässt sich kein sicherer ursächlicher Zusammenhang herleiten. Es kommen andere Ursachen für die Verbesserung der Symptome infrage, wie zum Beispiel der natürliche Verlauf oder ein Placeboeffekt. Die Annahme eines ursächlichen Zusammenhangs aufgrund einer zeitlichen Abfolge wird als *Post hoc ergo propter hoc*-Fehlschluss bezeichnet.

Am Workshop haben 71% der Teilnehmenden der Ansicht der Patientin nicht zugestimmt.

## Konklusion

Die Beobachtungen am Workshop zeigten uns: Die Teilnehmenden schienen geistig fit genug, um ihr System 2 einzusetzen. Schlüsse auf allgemeine kognitive Reserven von Internistinnen und Internisten lassen sich jedoch daraus natürlich nicht ziehen: Zum Beispiel kommt ein *Selection Bias* infrage, wobei besonders interessierte und vorinformierte Ärztinnen und Ärzte für unseren Workshop angemeldet gewesen wären. Interessanterweise haben deutlich mehr Teilnehmende sich im Vergleich mit anderen im Raum unterschätzt als überschätzt. Die Vignette, bei der die Teilnehmenden am schlechtesten abschnitten, war

jene mit der *Conjunction Fallacy*. Dies mag ein Hinweis darauf sein, dass die Teilnehmenden anfällig sein könnten, sich von Stereotypen in die Irre führen zu lassen, idem vorschnell eine ungerechtfertigt genaue Diagnose unterstellt wird. Am besten abgeschnitten haben die Ärztinnen und Ärzte beim *Post hoc ergo propter hoc*-Fehlschluss. Dies spricht dafür, dass die meisten Teilnehmenden korrekterweise bei der Beurteilung von Wirksamkeitsfragen auf kontrollierte Beobachtungen setzen.

Schliesslich bleibt die Frage: Können wir nun etwas gegen solche Denkfehler tun? Oder müssen wir uns damit abfinden, dass gewisse Fehler für uns unvermeidbar bleiben? Die Lage ist glücklicherweise nicht hoffnungslos und es sind verschiedene Techniken beschrieben, allgemein unter dem Begriff *Debiasing* zusammengefasst, die uns bei der Vermeidung von Verzerrungen und Fehlschlüssen helfen sollen. Zwar hält sich die Evidenz zu ihrer Wirksamkeit in Grenzen, jedoch lohnt sich eine Auseinandersetzung mit ihnen durchaus [5]. Als effektiv für den klinischen Alltag hat sich die Technik des *Cognitive forcing* (deutsch etwa «kognitives Erzwingen») erwiesen [3]. Dabei geht es darum, bewusst den eigenen Denkprozess zu bremsen und über ihn nachzudenken, also Metakognition zu betreiben. Beispielsweise kann es bereits hilfreich sein, vor einer Diagnosestellung inne zu halten und sich zu vergewissern, dass man alle infrage kommenden alternativen Differenzialdiagnosen des Leitsymptomes und speziell die gefährlichen darunter ausreichend gewürdigt und aktiv entkräftet hat.

Diese Ansätze beruhen auf der Hoffnung, dass bereits die Kenntnis und das Bewusstsein über solche Fehler dazu verhelfen können, weitere zu vermeiden. Durch wiederholtes Beobachten von Fehlschlüssen und Verzerrungen bei sich selbst wie bei anderen im (nicht nur) klinischen Alltag erreicht man sicher eine erhöhte Sensibilität für das Thema. Und wer weiss, vielleicht führen bereits die Teilnahme an unserem Workshop oder das Lesen dieses Artikels dazu, den einen oder anderen Fehlschluss zu vermeiden.

### Literatur

- 1 Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ Theory Pract.* 2009;14(Suppl 1):27–35.
- 2 Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. 1974;185(4157):1124–31.
- 3 O'Sullivan ED, Schofield SJ. Cognitive bias in clinical medicine. *J R Coll Physicians Edinb.* 2018;48(3):225–32.
- 4 Frederick S. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives.* 2005;19(4):25–42.
- 5 Ludolph R, Schulz PJ. Debiasing Health-Related Judgments and Decision Making: A Systematic Review. *Med Decis Making.* 2018;38(1):3–13.

Korrespondenz:  
Dr. med. Stefan Markun  
Institut für Hausarzt-  
medizin  
Universitätsspital Zürich  
Pestalozzistr. 24  
CH-8091 Zürich  
stefan.markun[at]usz.ch