

# Werden wir dümmer?

## Der Flynn-Effekt im Rückwärtsgang

M. Spitzer, Ulm

Wenn man Kinder oder erwachsene Menschen ganz viele, ganz unterschiedliche Aufgaben zum Lösen gibt, dann fällt auf, dass manche mehr und manche weniger Aufgaben lösen können, wobei es nahezu egal ist, um welche Aufgaben es geht. Schon vor mehr als hundert Jahren wurden hierzu standardisierte Verfahren zur Messung der ganz allgemeinen geistigen Leistungsfähigkeit entwickelt, verbreitet und verwendet, um beispielsweise geistig behinderte Kinder zu identifizieren, um sie in besonderen Bildungseinrichtungen besser fördern zu können (1, 2). Der Ausdruck „Intelligenz“ für diese *allgemeine Aufgabenlösefähigkeit* wurde schon am Ende des vorletzten Jahrhunderts vom französischen Psychologen Alfred Binet eingeführt (▶ Abb 1).

Kinder verschiedenen Alters musste vielfältige Aufgaben unterschiedlicher Schwierigkeit (▶ Abb. 2–5), die ganz unterschiedliche kognitive Funktionen erforderten (Sprachfähigkeit, Wortschatz, Merkfähigkeit, Vorstellungsvermögen, logisches Schließen etc.) lösen. Daraus bestimmte man dann, in welchem Alter Kinder im Mittel welche Aufgaben (Schwierigkeitsgrad) lösen können. Man experimentierte über Jahre mit den Aufgaben herum und testete dann 30 Aufgaben bei 50 Kindern

aus 5 Altersgruppen aus. Die einfachsten Aufgaben (mit den Augen dem Licht eines Streichholzes folgen, dem Versuchsleiter die Hand geben) konnten von allen Kindern ausgeführt werden; das Benennen von Gegenständen (▶ Abb. 2) oder Körperteilen, Nachsprechen kurzer Sätze oder Nachzeichnen einfacher Zeichnungen (▶ Abb. 3), Definieren einfacher Wörter („Haus“, „Gabel“) oder drei Zahlen in umgekehrter Reihenfolge nachsprechen waren etwas schwieriger. Das Bilden von Sätzen aus vorgegebenen Wörtern, das Nachzeichnen aus dem Kopf (▶ Abb. 3), das Angeben von Unterschieden zwischen bestimmten Sachen oder das Nachzeichnen von Mustern aus dem Kopf (▶ Abb 4) waren noch schwieriger. Nur ältere Kinder konnten komplexe Vorstellungsaufgaben lösen (▶ Abb. 5) und zu den schwierigsten Aufgaben gehörten „Mein Nachbar hat eigenartigen Besuch. Erst kam ein Arzt, dann ein Anwalt und dann ein Pfarrer. Was war geschehen?“ und das Nachsprechen von sieben Zahlen in umgekehrter Reihenfolge.



**Abb. 1** Alfred Binet (1857–1911) arbeitete von 1883 bis 1889 als Schüler von Jean-Martin Charcot in dessen neurologischem Labor an der Salpêtrière in Paris. Im Jahr 1889 gründete er das erste psychologische Forschungslabor Frankreichs und entwickelte zusammen mit seinem Schüler, dem Arzt Théodore Simon (1873–1961), die ersten brauchbaren Intelligenztests für Kinder.

1 Im Original: „Nous montrons cette gravure à l'enfant, et nous lui demandons de nous désigner successivement les objets suivants : la fenêtre, la maman, la grande soeur, la petite fille, le chat, le balai, le panier, le bouquet, le plumeau, le moulin à café. La question est ainsi posée : «Où est la fenêtre? Dis moi où est la fenêtre? Montre-moi la fenêtre? Mets ton doigt sur la fenêtre?»“ (Binet & Simon 1904, S. 204).

Nervenheilkunde 2018; 37: 617–625

### Korrespondenzadresse

Prof. Dr. Dr. Manfred Spitzer, Universitätsklinikum Ulm  
Klinik für Psychiatrie und Psychotherapie III  
Leimgrubenweg 12, 89075 Ulm

### Abb 2

Aufgabe zur Messung der Intelligenz (aus 1, S. 205), bei der auf einem Bild dargestellte Menschen und Objekte zu benennen sind: „Zeig mir das Fenster...“<sup>1</sup>

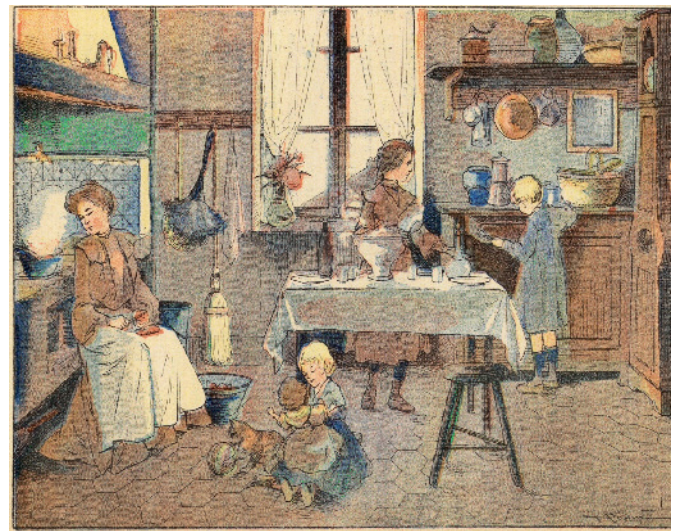


Fig. 1. — Gravure servant à rechercher si l'enfant, normal ou anormal, découvre et indique correctement les objets qu'on lui nomme.

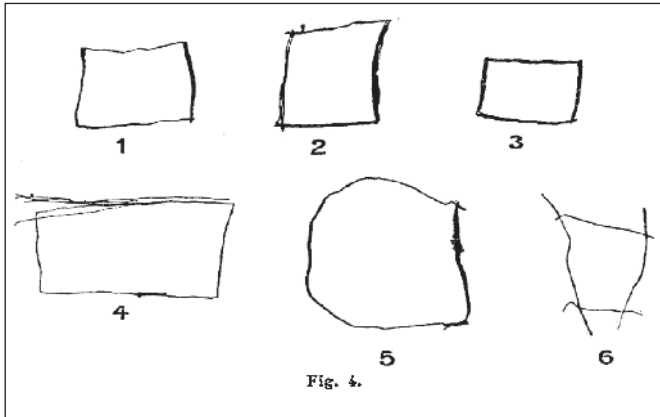


Abb 3

Aufgabe für ein 5-jähriges Kind (aus 2, S. 18), bei der ein vorgegebenes Rechteck im Format 3 x 4 cm nachzuzeichnen ist. Die oberen drei Lösungen halten die Autoren für akzeptabel, die unteren drei nicht.

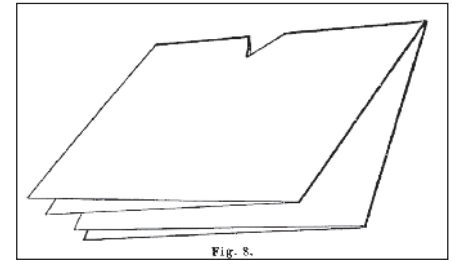


Abb. 5 Aufgabe für ein 13-jähriges Kind (aus 2, S. 55): Angegeben werden soll, wie ein zweimal gefaltetes Papier, in das mit der Schere, wie in der Abbildung ersichtlich, ein Dreieck herausgeschnitten wurde, nach dem Auffalten aussieht.

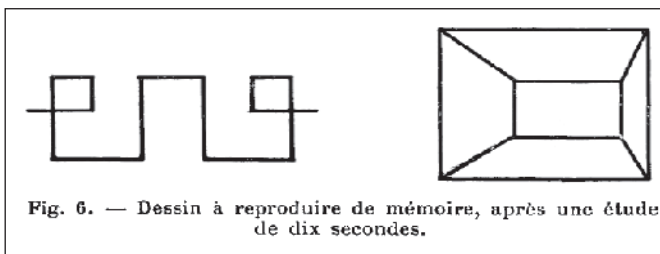


Abb. 4

Zwei Aufgaben zur Messung der Intelligenz (aus 1, S. 216), bei denen Muster nach 6 Sekunden Betrachtung auswendig nachgezeichnet werden sollen.

**Der Ausdruck „Intelligenz“ für allgemeine Aufgabenlösefähigkeit wurde am Ende des vorletzten Jahrhunderts von Alfred Binet eingeführt.**

Binet und Simon bestimmten mit diesen Aufgaben das Intelligenzalter von Kindern im Alter von 3 bis 13 Jahren. Mittels der gesamten Aufgabenbatterie (5 bis 7 Tests für jede Altersstufe) konnten sie nach einer solchen Testung Kinder mit höherer Aussicht auf Förderung der für sie passenden Beschulung zuteilen. Heute würde man sagen: Sie versuchten einen höheren Grad der Individualisierung des Unterrichts. Nicht mehr und nicht weniger.

Wenige Jahre später entwickelte der deutsche Psychologe William Stern<sup>2</sup> diese Überlegungen weiter und erfand dabei den IQ (40, S. 28; ► Abb. 6): Wenn ein Kind mit 6 Jahren die Aufgaben lösen kann, die Sechsjährige lösen können, so entspricht sein „Intelligenzalter“ auch seinem Lebensalter. Setzt man beide ins Verhältnis, ergibt sich der „Intelligenz-Quotient“ (der dann noch zur besseren Handhabung mit 100 multipliziert wird) wie folgt:

$$\begin{aligned} & (\text{Intelligenzalter/Lebensalter}) \times 100 \\ &= 6/6 \times 100 \\ &= 100. \end{aligned}$$

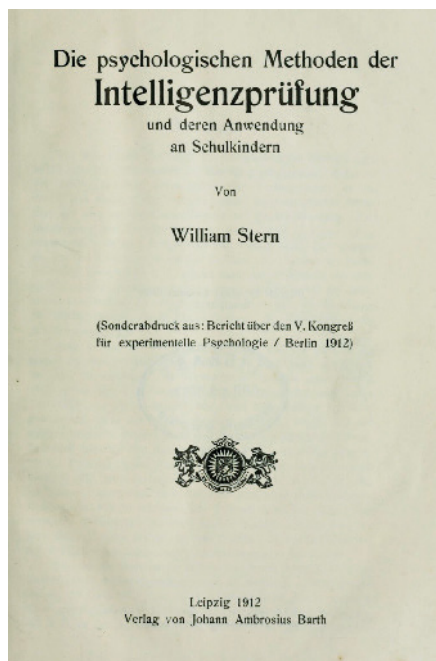
Kann ein fünfjähriges Kind diese Aufgaben (die sechsjährige Kinder lösen können) schon lösen, dann sagt man, sein Intelligenzalter ist 6 Jahre. (Heute würde man sagen: Intelligenzmäßig ist das Kind schon sechs Jahre alt.) Sein Intelligenzquotient (Intelligenzalter/Lebensalter) mal 100 beträgt damit  $(6/5 = 1,2) 120$ .

<sup>2</sup> William Stern (1871–1938) war einer der bedeutendsten deutschen Psychologen, Begründer der Differenziellen Psychologie, Mitbegründer der Universität Hamburg, der Deutschen Gesellschaft für Psychologie (DGPs) und der Zeitschrift für angewandte Psychologie. Stern erhielt im Jahr 1909 zusammen mit Sigmund Freud und Carl Gustav Jung die Ehrendoktorwürde der Clark University, obwohl er der Psychoanalyse zeitlebens kritisch gegenüber eingestellt war. Im Jahr 1905 verwendete er erstmals den Ausdruck „Deutungspfücher“ zur Bezeichnung von Psychologen, „die ihren Beruf dazu nutzen, ihre Privatmeinung und ihre persönlichen Einstellungen und Vorurteile als psychologische, wissenschaftliche Erkenntnis zu verkaufen“ (wie im Eintrag der Wikipedia unter „Wissenswertes“ vermerkt ist; der Autor konnte die Quelle nicht verifizieren und ist für entsprechende Hinweise dankbar).

Später verwendete man „Intelligenz“ auch zur Bezeichnung der gleichen allgemeinen geistigen Leistungsfähigkeit bei Erwachsenen. Allerdings klappt in dieser Gruppe der „Quotient“ (ein 50-Jähriger der die Aufgaben eines 60-Jährigen löst etc.) nicht mehr, weil die Gehirnentwicklung längst abgeschlossen ist. Stattdessen lässt man viele Erwachsene viele Aufgaben unterschiedlichen Schwierigkeitsgrades lösen, und findet, dass die meisten Leute eine mittlere Anzahl von Aufgaben lösen können. Einige Wenige können nur ganz wenige (leichte) Aufgaben lösen, die Mehrheit liegt in der Mitte und wieder nur ganz Wenige lösen fast alle Aufgaben, also auch die schwierigsten und schweren. Die Verteilung sieht ähnlich aus wie eine Glockenkurve, die man aus der Statistik kennt und *Gauss'sche Normalverteilung* heißt (► Abb. 7).

Wählt man nun die sie beschreibenden Parameter bei Erwachsenen so (Mittelwert: 100; Standardabweichung: 15) wie sie bei Kindern empirisch gefunden worden waren (Quotient aus Intelligenz- und Lebensalter mal 100), dann kann man bei Erwachsenen so etwas wie einen IQ bestimmen (obwohl das gar kein Quotient mehr ist, sondern „nur“ noch ein Punkt auf einer Kurve).

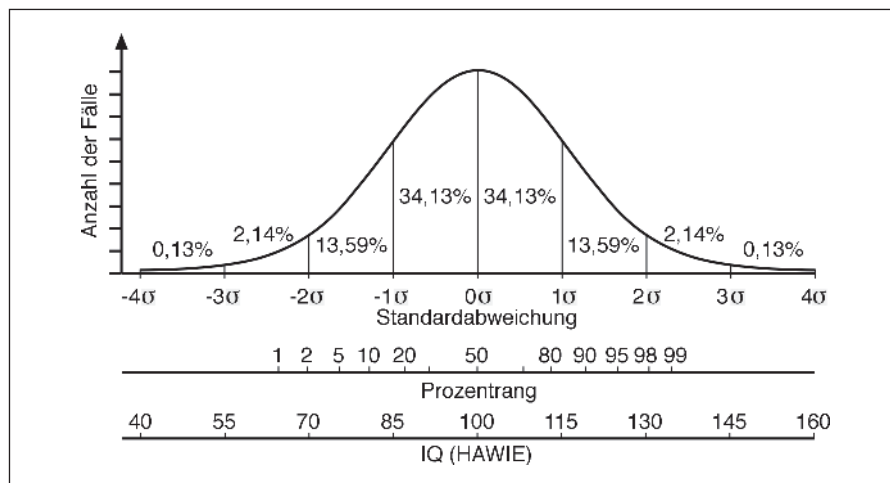
Bald wurde erkannt, wie praktisch das sein kann, und so fand der IQ in den USA beispielsweise bald Verwendung bei der Zulassung zum College. Wenig später wurde bzw. wird der IQ noch immer weltweit beim Militär bzw. in Personalabteilungen



**Abb. 6** „Geburt“ des Intelligenzquotienten in einem unscheinbaren Sonderdruck aus einem Kongressband. Diese gut 100 Seiten lange Schrift wurde von Stern ausdrücklich angefertigt, weil dies – wie man im Vorwort liest – „das jetzt ständig wachsende Interesse des In- und Auslandes [...] und zugleich die außerordentlich zersplitterte, z. T. schwer zugängliche Literatur eine erstmalige Gesamtdarstellung“ erfordere (40, S. III).

vieler Firmen zur Identifikation von Menschen für Führungsaufgaben verwendet. Diese Aspekte der Anwendung von Intelligenztests in Schulen und anderswo beschrieb schon vor fast hundert Jahren der Schulrat (in den USA klingt der Name des Jobs erheblich besser: „School District Superintendent“) Samuel Brooks aus Winchester im US-Bundesstaat New Hampshire. Sein Text beginnt jedoch mit einer gewissen Einschränkung: „In den Händen von praktischen Menschen stellen Intelligenztests praktische Werkzeuge dar, um praktische Zwecke zu erreichen“ (5, S. 217). Weniger praktisch veranlagte Menschen sollten sich also fern von diesem Werkzeug halten, was aus meiner Sicht für die meisten Werkzeuge gilt.

Vor allem die Anwendung von Intelligenztests beim Militär führte zu unglaublich großen Datenmengen. In den USA wurde beispielsweise im Jahr 1917 vor deren Eintritt in den Ersten Weltkrieg der IQ von 1 726 966 Soldaten verwendet, um



**Abb. 7** Trägt man die Intelligenz nach rechts und die Anzahl der Menschen mit entsprechender Intelligenz nach oben auf, ergibt sich die bekannte Glockenkurve, deren Eigenschaften der Mathematiker Carl-Friedrich Gauss erstmalig beschrieben hat, und die daher auch *Gauss-Kurve* heißt. Messgrößen, die eine solche Verteilung aufweisen, bezeichnet man auch als normalverteilt oder Gauss-verteilt. Weil das bessere oder schlechtere Lösen von ganz unterschiedlichen Aufgaben durch ganz unterschiedliche Menschen einigermassen<sup>3</sup> normalverteilt ist, und weil es zu den Eigenschaften der Glockenkurve gehört, dass im Bereich von einer Standardabweichung um den Mittelwert herum 68,27% aller Werte liegen, haben 68,27% der Menschen (gut zwei Drittel) einen IQ zwischen 85 und 115.

Führungskräfte auszuwählen. Der gerade erwähnte School District Superintendent kommentierte: „When we entered the war against Germany our military authorities were faced with the problem of selecting and training thousands of new officers to lead the millions of raw recruits furnished by the draft. Time was at a premium. The psychologists offered their assistance, and after a period of trial it was decided to permit them to select the new officer material by giving intelligence tests to the more promising of the drafted and enlisted men“ (5, S. 221).

Halten wir fest: Wenn man die Intelligenz von Erwachsenen so definiert, dass ihr Mittelwert bei 100 liegt, dann kann die Menschheit als Ganze eigentlich nicht intelligenter werden. Schon in der ersten Hälfte des letzten Jahrhunderts fiel jedoch auf, dass die Leute immer besser bei IQ-Tests abschnitten. Dies veranlasste die Testtheoretiker zunächst dazu, dass Intelligenztests neu normiert wurden, um den Mittelwert von 100 wieder herzustellen. Das Problem mit solchen Re-Normierungen ist, dass dann alle wieder „dümmer“ eingestuft werden, was zu nachweislichen Verwerfungen bei der Diagnose „geistige Behinderung“ und bei der Einstellung von

„Grenzbegabten“ beim Militär geführt hat (22).

Bereits im Jahr 1938 hatte A. Merrill die Daten von 905 Personen, bei denen im Jahr 1916 der Stanford-Binet IQ-Test durchgeführt worden war (mit damaliger Normierung), mit den Daten von 2904 neu durchgeführten IQ-Tests (gleicher Test mit neuer Normierung) verglichen und auf die Probleme durch die neue Normierung hingewiesen: Sie machte die Schwachen schwächer (30).

**Bei Erwachsenen wird Intelligenz nicht als „Quotient“ berechnet, der Name Intelligenzquotient wurde jedoch beibehalten.**

Zehn Jahre später zeigte eine Untersuchung von R. Tuddenham an 768 US-Soldaten des Zweiten Weltkriegs, dass deren IQ-Werte bei Zugrundelegung der Nor-

<sup>3</sup> Bei genauerer Betrachtung zeigt sich, dass die empirisch bestimmten Werte nicht ganz der Gauss-Verteilung entsprechen: Es gibt empirisch mehr Minderbegabte als der (theoretischen) Gauss-Verteilung entsprechen. Man berücksichtigt dies bei der „Normierung“ der Intelligenzwerte, also bei der Umrechnung der Rohwerte (erreichten Punktzahlen in den Untertests) in den IQ.

men des Ersten Weltkriegs signifikant höher waren. Er interpretierte seine Daten wie folgt: „Der Schreiber neigt dazu, sie als Ausdruck der Überlegenheit der Testergebnisse der untersuchten Population gegenüber der vorhergehenden Generation zu interpretieren, und dass ein großer Teil dieser Verbesserung die Konsequenz einer besseren (Aus-)Bildung von mehr Menschen ist (45, S. 54).

In der Folge betrachtete man immer öfter große Datensätze von Intelligenzmessungen (z. B. vom Militär) mit verschiedenen Kohorten über die Zeit hinweg. Eine der neuesten und größten Metaanalysen hierzu (271 unabhängige Datensätze, fast 4 Millionen Testergebnisse aus 31 Ländern) zeigte eine deutliche Zunahme des IQ zwischen den Jahren 1909 und 2013 (32). Wie aus ►Tabelle 1 ersichtlich, betrug das Ausmaß dieser Zunahme in den meisten Ländern zwischen knapp einem und zwei bis drei IQ-Punkten pro Jahrzehnt (mit einem „Ausreißer“ nach unten von 1,2 IQ-Punkten Abnahme pro Jahrzehnt in Brasilien). Die Aussagekraft der Daten ist aufgrund der erheblich unterschiedlichen Datenmengen (N) sehr unterschiedlich: Die Daten von über einer Million Schweden bzw. über einer halben Million Dänen muss man ernster nehmen als die von 89 Südkoreanern oder 124 Schweizern!

Das Phänomen des über die Jahrzehnte steigenden IQ wurde von Herrnstein und Murray in ihrem 1994 erschienenen, viel-diskutierten Buch über Intelligenz als „Flynn-Effekt“ bezeichnet (S. 307), da es vom neuseeländischen Wissenschaftler James Flynn als erstem ausführlich beschrieben worden war – zunächst für die USA (Flynn 1984) und drei Jahre später für weitere 14 Länder einschließlich Deutschland (15).<sup>4</sup>

4 Dazu Herr Flynn (16, S. 851) selbst: „Ich habe selbst niemals einen IQ-Test durchgeführt“. In der Tat war er Jurist von Beruf. Aus seiner Sicht hätte dem Briten Richard Lynn die Ehre der Namensgebung dieses Effekts zu Teil werden sollen, da dieser bereits 1982 im Fachblatt *Nature* eine Zunahme des IQ für Japan gezeigt hatte. Dies erörternd räumt Robert Williams (47) in seiner Übersicht zum Flynn-Effekt eingangs ein, dass dieser eigentlich Lynn-Flynn-Effekt heißen müsste oder wenigstens „FLynn“ Effekt (mit einem großen zweiten „L“ für „Lynn“).

Flynn diskutiert diesen Effekt sehr vorsichtig und unterscheidet drei Möglichkeiten im Hinblick auf seine Bedeutung. „Nicht real“: Es könnte sich um ein statistisches Artefakt handeln, d. h. der IQ und die damit gemessene Intelligenz der Menschen steigt nicht wirklich – beispielsweise

weil über die Zeit hinweg andere Personengruppen gemessen wurden. So könnten durch die zunehmende Anwendung von IQ-Tests über die Zeit hinweg Menschen mit geistiger Behinderung in zunehmendem Ausmaß in spezielle Einrichtungen verbracht und damit „ausortiert“ worden

**Tab. 1** Zunahme des IQ pro Jahrzehnt (Flynn-Effekt) nach Ländern geordnet (nach Daten der Metaanalyse von Pietschnig und Voracek (33) aus Tabelle S2 im Online Supplement).

Land	Studien	N (Personen)	Zeitraum	IQ-Zunahme Punkte/Jahrzehnt
Argentinien	2	2 270	1964–1998	7,2
Australien	8	41 432	1936–1995	4,2
Belgien	2	102 400	1958–1967	8,1
Brasilien	13	1 906	1930–2007	-1,2
Bulgarien	1	1 657	1941–1973	2,1
China	2	3 231	1984–2011	2,2
Dänemark	1	549 148	1959–2004	1,7
Deutschland	13	504 284	1956–2008	6,0
Dominikanische Republik	1	725	1947–1983	4,9
Estland	1	1 812	1935–2006	1,7
Finnland	1	75 144	1988–2009	2,0
Frankreich	6	146 377	1938–1993	4,4
Großbritannien	24	164 937	1932–2008	1,1
Irland	2	117	1942–1972	2,2
Israel	3	492 000	1971–1984	5,8
Japan	3	4 466	1951–1975	10,2
Kanada	2	9 583	1946–1976	4,4
Kenia	1	655	1984–1998	17,9
Neuseeland	3	320 42	1936–1978	2,4
Niederlande	9	133 460	1952–2005	3,6
Norwegen	1	210 000	1954–2002	2,3
Österreich	4	10 602	1962–2000	2,4
Saudi-Arabien	4	4 628	1977–2013	3,5
Schweden	5	1 084 903	1912–1937 1961–2004	2,5
Schweiz	1	124	1954–1981	8,0
Spanien	9	43 239	1963–2000	4,6
Südkorea	1	89	1986–1999	5,6
Südafrika	4	3 239	1963–1987	1,6
Sudan	2	4 292	1964–2007	2,8
Türkei	1	476	1977–2010	1,6
USA	136	54,243	1909–2006	3,4

sein, sodass immer weniger von ihnen in Schulen oder beim Militär getestet wurden.

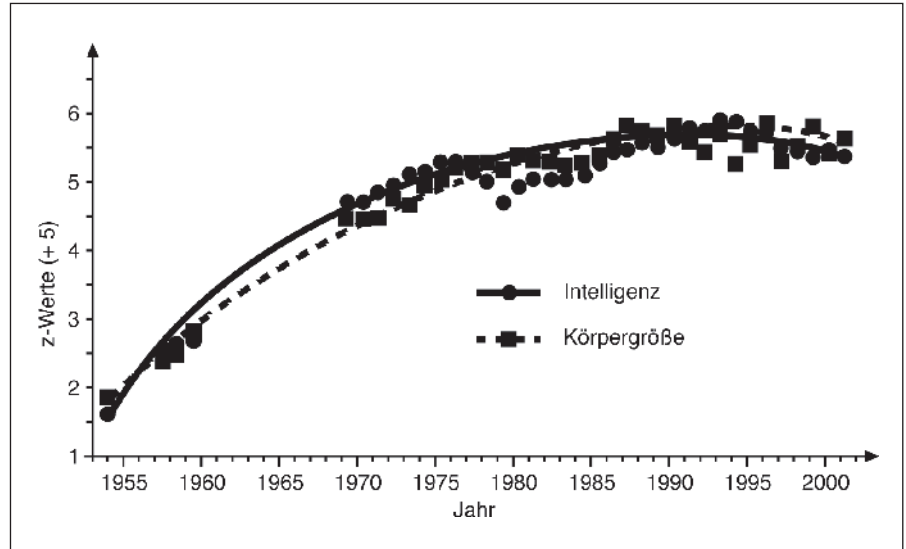
„Semi-real“: Es könnte sich um ein Artefakt der Messung handeln. Die Menschen werden immer besser in der Bearbeitung von Tests oder (wie später vorgeschlagen wurde) im Erraten der richtigen Antworten („improved guessing“). In diesem Fall würde nur der gemessene IQ steigen, die Intelligenz jedoch „in Wahrheit“ nicht.

„Real“: In diesem Fall hätte man ein Problem. Es folgte nämlich beispielsweise, dass ein Mensch, der vor 100 Jahren einem IQ von 100 hatte (also „durchschnittlich intelligent“ war), heute nur noch einen IQ von 70 haben und damit als am Rande der Behinderung eingestuft werden würde. Umgekehrt hätte ein Mensch, der heute einen IQ von 100 hat, vor hundert Jahren einen IQ von 130 gehabt und hätte damit als hochbegabt gegolten. Kann das wirklich sein? Herr Flynn selbst jedenfalls hielt dies eher für unwahrscheinlich.

### Das Phänomen des über die Jahrzehnte steigenden IQ wurde 1994 als „Flynn-Effekt“ bezeichnet.

Trotz dieser recht skeptischen Interpretation seines Namensgebers wurde die Steigerung des IQ von vielen Wissenschaftlern ernst genommen. Selbst wenn es so ist, dass ein Teil des Effekts darauf zurückgeht, dass man heute nicht mehr alle Menschen eines Jahrgangs testet und dass Wiederholungsmessungen des IQ gezeigt haben, dass man den Test teilweise „lernen“ kann, reichen diese Erklärungen für das ganze Ausmaß des Effekts nicht aus. Wenn aber die Intelligenz der Menschen wirklich zunimmt, stellt sich die Frage, woran das liegt.

Genetische Ursachen wurden diskutiert, scheiden jedoch aus, weil sie die Geschwindigkeit des Effekts nicht erklären können. So bleiben Umwelteinflüsse, und in der Tat hat man gute Gründe für ihre Annahme. In ►Abbildung 8 ist beispielsweise die Entwicklung der Körpergröße und des gemessenen IQ in Norwegen von 1954 bis 2002 (entspricht den Geburtskohorten von etwa 1935 bis 1984) dargestellt. Eine der Ursachen der Entwicklung beider gemessener Größen besteht wahrscheinlich in einer besseren Ernährung und Gesundheitsvor-



**Abb. 8** Entwicklung von Körpergröße und Intelligenz am Beispiel von Norwegen in den Jahren 1954 bis 2002 (nach 42, S. 357). Die Ursachen hierfür reichen von einer besseren Ernährung (erklärt beide Kurven) bis zu einer besseren Ausbildung (erklärt die Kurve der Intelligenz). Es muss sich jedoch in jedem Fall um eine Änderung von Umweltfaktoren handeln.

sorge großer Teile der Bevölkerung über den Beobachtungszeitraum hinweg. In Schweden lagen die Dinge allerdings anders, denn dort wurde kein Zusammenhang zwischen der Entwicklung von Körpergröße und IQ über die Zeit hinweg beobachtet (38).

Auch der sozioökonomische Status der Menschen in Norwegen hat sich über diesen Zeitraum verbessert und dürfte daher eine Rolle spielen. Für diese Annahmen spricht auch die Tatsache, dass beispielsweise in Japan ebenfalls starke Zunahmen von IQ und Körpergröße gemessen wurden. Der wirtschaftliche Aufschwung des Landes nach dem Zweiten Weltkrieg und eine damit einhergehende verbesserte Ernährung haben hieran sicherlich einen nicht geringen Anteil.

Neben verbesserten ökonomischen Rahmenbedingungen des Lebens spielen auch Verbesserungen im Bildungsbereich bei der Zunahme des gemessenen IQ eine Rolle. Eine ebenfalls in Norwegen durchgeführte quasi-experimentelle Untersuchung zu den Auswirkungen einer in den 1960er-Jahren implementierten Schulreform im Sinne einer Erhöhung der allgemeinen Schulpflicht um zwei Jahre (von 7 auf 9 Jahre) zeigte eine deutliche Steigerung des IQ durch die Reform (4). Gemessen wurde jeweils der IQ mit 19 Jahren (als Teil der

Untersuchung junger Männer durch das Militär). Die Reform betraf Schüler im Alter von 14 bis 16 Jahren, die nun *alle* (im Gegensatz zur Zeit vor der Reform) zur Schule gehen mussten. Die Reform wurde je nach Gemeinde nicht im gleichen Jahr, sondern über einen Zeitraum von mehreren Jahren implementiert. Dies erschwerte einerseits die Auswertung (weil für etwa 20% der Schüler der genaue Zeitpunkt der Reform nicht mehr feststellbar war), erhöhte jedoch die Robustheit der Daten (weil die Verteilung der Reform über mehrere Jahre die Auswirkungen mancher Idiosynkrasien einzelner Jahre verringerte). Der Effekt der Reform auf den IQ wurde mit 0,6 Punkten ermittelt. Weil der Flynn-Effekt in der entsprechenden Kohorte mit 1,6 IQ-Punkten gemessen worden war, folgern die Autoren, dass mehr als ein Drittel des Flynn-Effekts im betreffenden Zeitraum der 1960er-Jahre auf das Konto der Schulreform geht.

Eine kürzlich publizierte Metaanalyse (37) untersuchte Studien zu drei unterschiedlichen quasi-experimentellen Ansätzen (einer davon der gerade beschriebene) an insgesamt über 600 000 Schülern zu den Auswirkungen von Schule auf den IQ. Sie ergab, dass ein zusätzliches Schuljahr den IQ um 1 bis 5 IQ-Punkte erhöhen kann. Dass dies überhaupt möglich sein könne,

war noch in den 1980er-Jahren des letzten Jahrhunderts heftig umstritten: Der IQ war damals (nicht zuletzt aus ideologischen Gründen) für im Wesentlichen genetisch bedingt und damit durch Bildungsmaßnahmen kaum beeinflussbar betrachtet worden. Die Zusammenhänge zwischen IQ und Bildung erklärte man ausschließlich durch einen Selektionseffekt (wer schlauer ist, geht länger zur Schule!). Dass dieser Selektionseffekt zwar besteht, aber nicht ausschließlich den Zusammenhang zwischen IQ und Bildung erklärt, ist mittlerweile durch eine größere Zahl von Studien sehr gut belegt. Mit den Worten der Autoren: „Education appears to be the most consistent, robust, and durable method yet to be identified for raising intelligence“ (37, S. 1) (► Tab. 1).

**„Es wird Zeit, dass wir Bildung als die beständigste, stärkste und nachhaltigste Methode zur Förderung der Intelligenz anerkennen.“**

Entsprechend beurteilt auch Flynn die Steigerung des IQ mittlerweile weniger als Artefakt, sondern als durchaus plausible, *reale* Änderung geistiger Leistungsfähigkeit (17). „[...] mehr und bessere Beschulung, kognitiv stärker herausfordernde Arbeitsplätze und eine bessere Gesundheit [...] verursachen große Steigerungen des IQ über mehrere Generationen hinweg“.

Halten wir fest:

- Intelligenz wird seit über hundert Jahren gemessen und bezeichnet die Fähigkeit von Menschen, Aufgaben bzw. Probleme zu lösen.
- Sie zeigt sich schon im Kindesalter, ist in ihrer Entwicklung während Kindheit und Jugend (Lebensphase, in der sich das Gehirn entwickelt) jedoch auch abhängig von der Qualität und Quantität von Bildungsprozessen.
- Fast solange, wie Intelligenz in Form des IQ gemessen wird, zeigt sich fast überall ein Anstieg des IQ über die Jahrzehnte hinweg.

Dies war der Stand der Dinge bis etwa zur Jahrtausendwende. Seitdem erschien eine Reihe von Arbeiten, in denen über eine in jüngster Zeit stattfindende Abnahme des IQ über die Jahre hinweg berichtet wird.

Zu den ersten<sup>5</sup> diesbezüglichen Daten gehörten die in ► Abbildung 8 bereits dargestellten, an norwegischen Rekruten gewonnenen IQ-Werte (42). Auch Daten aus Dänemark der beiden Psychologen Thomas Teasdale und David Owen, die in der Arbeit *Der Flynn-Effekt im Rückwärtsgang* im Jahr 2005 publiziert wurden, beschreiben den gleichen Sachverhalt. In einer Stellungnahme wird Teasdale mit den Worten zitiert (8): „Mit Beginn der neunziger Jahre hörte die Steigerung der IQ-Werte auf. Seit 1999 beobachten wir einen Rückgang.“

Seit etwa Mitte der 1990er-Jahre erwiesen sich die Werte, zunächst in den genannten skandinavischen Ländern und später in weiteren Ländern Europas als rückläufig; die Studien zeigten weiterhin, dass der Trend anhält (► Tab. 2).

Wie kommt es zum negativen Flynn-Effekt? Die wohl am häufigsten hierfür diskutierte Ursache ist ein genetischer Effekt: Schon lange ist bekannt, dass der Bildungsstand eines Menschen negativ mit der Anzahl der Kinder korreliert, was insbesondere bei Frauen gilt, jedoch in schwächerem Ausmaß auch bei Männern gefunden wurde (6, 7, 36). Wenn nun Menschen umso weniger Kinder bekommen, je gebildeter sie sind, sollte der IQ in hochentwickelten Ländern langfristig fallen, und genau darauf wurde der negative Flynn-Effekt zurückgeführt (10, 28). Er wurde bislang nur in hochentwickelten Ländern gefunden, wohingegen der (positive) Flynn-Effekt in den noch weniger entwickelten Ländern nach wie vor zu finden ist. Nach einer Befragung von 75 Experten im Bereich „Flynn-Effekt“ und „negativer Flynn-Effekt“ aus dem Jahr 2017 wurde dieser genetische Mechanismus mit am häufigsten als Ursache genannt (35).

Aus dem gleichen Jahr stammt eine große isländische Studie, die diese Überlegung zunächst zu bestätigen scheint. Island ist bekanntermaßen das Eldorado der Humangenetik. Dort ist man dabei die gesamte Bevölkerung – etwa 317 000 Menschen – genetisch zu untersuchen, um diese Daten

dann mit biologischen, psychologischen und sozialen Variablen (z. B. zur Bildung oder Gesundheit) in Verbindung zu bringen. Eine im Fachblatt *Nature* publizierte genomweite Assoziationsstudie (GWAS) konnte 74 signifikante Genloci identifizieren, die mit der Anzahl der Jahre, die ein Mensch in Bildungseinrichtungen verbrachte, in Zusammenhang stand (31). Zwar ist die Bildung eines Menschen (gemessen als Anzahl der Jahre in Ausbildung) nur zu etwa 20% genetisch beeinflusst, wohingegen 80% der Varianz auf das Konto von sozialen oder anderen Umweltfaktoren gehen, aber man kann diese Zahl relativ leicht und objektiv bestimmen. Aus diesen Genloci ließ sich ein Score berechnen, der mit einer erhöhten Bildung einhergeht. Im Jahr 2017 wurde nun dieser Score bei 129 808 Personen, die zwischen 1910 und 1990 geboren worden waren, bestimmt und seine Auswirkung auf die Reproduktion („reproduction history“) von 109 120 Isländern aus dieser Personen-Gruppe ermittelt (23). Dabei fand man tatsächlich eine Assoziation dieses Scores mit einer verminderten Anzahl von Kindern und mit einem höheren Lebensalter bei der Geburt des ersten Kindes – mit einer Irrtumswahrscheinlichkeit von  $p < 10^{-100}$  (!). Der Effekt war bei Frauen größer als bei Männern (► Tab. 3).

Hätte man nicht einfach auch mit einem Heer von Psychologen bei hunderttausend Isländern den IQ messen und die Zahl der Kinder sowie das Alter von Mutter und Vater bei deren Geburt erfragen können? Weil der polygenetische Score nur 3,7% der Varianz der Anzahl der Bildungsjahre erklärt und diese ja auch nur zu 50% (Größenordnung!) mit der Intelligenz korreliert, müsste man mit dieser ziemlich einfachen Methodik sogar wesentlich bessere Ergebnisse erwarten!<sup>6</sup> Aber: Weil man aber

<sup>5</sup> Im Nachhinein findet man schon frühere Erwähnungen des Phänomens (z. B. 13), denen jedoch vergleichsweise wenig Beachtung geschenkt wurde.

<sup>6</sup> „When applied to 46,079 Icelanders with educational attainment data POLY<sub>EDU</sub> was found to explain 3.74% of the trait variance ( $P < 10^{-300}$ ). By contrast, the strongest single variant only explains 0.10% of the variance, indicating that educational attainment is a complex trait influenced by many variants in the genome and highlighting the increased power of using the polygenic score for our analyses,“ schreiben die Autoren im Originaltext (23; s. E727).

**Tab. 2** Der Flynn-Effekt im Rückwärtsgang: Fallende Messwerte des IQ, absolut und in IQ-Punkten, nach Ländern geordnet (*n.a.*: nicht anwendbar, da sehr viele Einzelsamples zusammengefasst wurden; *BPP*: Børge Priens Prøve, ein Paper&Pencil IQ-Test für Gruppen bis 30; *Raven SPM*: Raven Standard Progressive Matrices; *Peruskoe*: bedeutet aus dem Finnischen ins Englische

übersetzt „basic test“, ein in Finnland entwickelter Intelligenztest; *WAIS III & IV*: Wechsler adult intelligence scale, Versionen 3 und 4; *GATB*: general aptitude test battery; *General Ability*: der Test kombinierte mehrere Verfahren; *3DC*: three dimensional cubes test, ein Test zum räumlichen Denkvermögen)

Autor, Jahr	Land	Altersgruppe (Jahre)	Testverfahren	N	Zeitraum	IQ-Abfall Punkte/Jahrzehnt
Teasdale, Owen 2005, 2008	Dänemark	18–19	BPP	ca. 25 000/Jahr	1998–2004	2,7
KoPrgeaar 2013	Estland	18–19	Raven SPM	552; 411; 304	2001; 2005; 2012	8,4
Dutton, Lynn 2013	Finnland	18–19	Peruskoe	ca. 25 000/Jahr	1997–2009	2,0
Dutton, Lynn 2015	Frankreich	Erwachsene	WAIS III & IV	2 x 79	1999; 2008/9	3,8
Shayer, Ginsburg 2007	Großbritannien	11–12	Piaget (Volumen, Gewicht)	10 023 (über 5 Kohorten verteilt)	1975; 2000; 2001; 2002; 2003	4,3
Shayer, Ginsburg 2009	Großbritannien	13–14	Piaget (Pendel und Gleichgewicht)	793	1976; 2006; 2007	2,5
Woodley, Meisenberg 2013	Niederlande	Erwachsene	GATB	n.a.	1975/2005	1,35
Sundet et al. 2004	Norwegen	18–19	General Ability	ca. 25 000/Jahr	1996–2002	3,8
Bratsberg, Rogeberg, 2018	Norwegen	18–19	General Ability	ca. 23 000 bis 32 000 /Jahr	1975–1991	3,4
Pietschnig, Gittler 2015	Österreich	n.a.	3DC	13 172	1977–2014	4,8

**Tab. 3** Einfluss des mit längerer Ausbildung assoziierten polygenetischen Scores auf die Anzahl der Kinder und das Alter bei der ersten Geburt sowie das Durchschnittsalter bei allen Geburten bei Frauen und Männern (aus 23, Table 1).

	Frauen			Männer		
	n	Effekt	p	n	Effekt	p
Anzahl der Kinder	58 560	-0,084	1,0 x 10 <sup>-43</sup>	50 560	-0,054	2,2 x 10 <sup>-15</sup>
Alter bei Geburt des Ersten Kindes	55 208	0,59	5,3 x 10 <sup>-155</sup>	45 669	0,44	6,2 x 10 <sup>-57</sup>
Mittleres Alter bei Geburt der Kinder	55 208	0,46	1,0 x 10 <sup>-117</sup>	45 669	0,37	6,5 x 10 <sup>-50</sup>

die Genetik sowieso gemacht hatte, alle Personen in entsprechenden Registern bestens erfasst sind und die Daten digital vorliegen, war es so herum ganz offensichtlich mit weniger Aufwand verbunden. – Könnte man meinen! Bei einem Teil der Studienteilnehmer – die zwischen 1910 und 1975 geborenen Frauen (n = 25 794) und Männer (n = 19 903) – hat man jedoch auch den Phänotyp (die Anzahl der Ausbildungsjahre) direkt ermittelt (in Island muss es für alles Register geben!) und konnte so nicht nur die Auswirkungen der für Bildung relevanten Gene (Genotyp) sondern auch der Bildungsjahre selbst (Phänotyp) mit der Reproduktion in Verbindung bringen. Hierbei zeigte sich dass die Effekte des Phä-

notyps tatsächlich größer sind, jedoch – bei gemeinsamer statistischer Betrachtung – zum großen Teil unabhängig voneinander: Die Auswirkung des genetischen Score für die Bildungsjahre auf die Anzahl der Kinder reduziert sich bei Frauen von -0,097 auf -0,053. Selbst wenn man nur Frauen mit 10 Jahren Bildung (n = 11 055) untersucht, ist der Einfluss des polygenetischen Scores auf die Anzahl der Kinder noch -0,079. „Dieses Ergebnis zeigt, dass der polygenetische Score eine direkte Auswirkung auf die Reproduktion hat, der unabhängig vom tatsächlich erreichten Bildungsgrad ist“ (23, S. e728).

Dies zeigt sich in besonderer Weise bei Männern, deren Bildungsjahre positiv –

wenn auch nur sehr gering (Effekt: 0,0063, p = 0,07) – mit ihrer Kinderzahl korrelierten. Bei gleichzeitiger statistischer Betrachtung des polygenetischen Scores für mehr Bildung und der Bildungsjahre (wenn man also die Auswirkungen von Genotyp und Phänotyp aufeinander mitberücksichtigt), verdoppelt sich der Effekt der Bildungsjahre knapp auf 0,011 und wird mit p=2,5 x 10<sup>-7</sup> signifikant.

Das in unserem Zusammenhang wichtigste Ergebnis der Studie enttäuscht trotz aberwitzig winziger Irrtumswahrscheinlichkeiten: Bei 1 577 Isländern wurde auch der IQ bestimmt. Weitere Berechnungen ergaben, dass die Auswirkung des polygenetischen Scores auf die Kinderzahl über

die Beobachtungszeit (1910 bis 1990) abgenommen hat und eine Extrapolation der Daten auf den IQ ergab eine Abnahme von 0,3 IQ-Punkten pro Jahrzehnt. „Wenn dieser Trend über Jahrhunderte anhielte, wäre dies ein bedeutsames Ergebnis“, schreiben die Autoren.<sup>7</sup> Den negativen Flynn-Effekt erklären können diese Daten jedoch definitiv nicht, dafür ist der in ihnen nachzuweisende Effekt etwa eine Größenordnung zu klein.

Wenn genetische Faktoren für den Rückgang des IQ nicht verantwortlich sein können, bleiben nur Umweltfaktoren zu Erklärung. Zu diesem Ergebnis kommen auch zwei in ►Tabelle 2 aufgeführten Studien, die daher abschließend erwähnt seien: Shayer und Ginsberg verwendeten eine Test, der auf Piagets-Konzepte der Entwicklung vom konkreten zum abstrakten Denken zurückgeht und dies mit den physikalischen Begriffen wie Volumenkonstanz, spezifischem Gewicht in Verbindung bringt. Die Fähigkeiten von 11- bis 12-jährigen Kindern in diesem Test haben seit den 1970er-Jahren deutlich abgenommen (►Abb. 9).

### Die Qualität der Schulen und der Medienkonsum sind wahrscheinlichste Ursachen der Abnahme des IQ seit Beginn der Jahrtausendwende in mehr als einem Dutzend Ländern.

Shayer und Ginsberg (51, S. 37) kommentieren: „On the reasons for the decline reported in this article one can only speculate. [...] Passive exposure to many hours of television a week has increased since the

7 „However, the possibility that such a phenomenon could be temporary or transitional was also raised. Indeed, there might be a cyclical element to this phenomenon, because it is only reasonable to assume that alleles associated with greater educational attainment must have been under positive selection at some time during the evolutionary history of Homo sapiens. The main message here is that the human race is genetically far from being stagnant with respect to one of its most important traits. It is remarkable to report changes in POLY<sub>EDU</sub> that are measurable across the several decades covered by this study. In evolutionary time, this is a blink of an eye. However, if this trend persists over many centuries, the impact could be profound“, kommentieren die Autoren (23, S. 730) am Ende ihrer Arbeit.

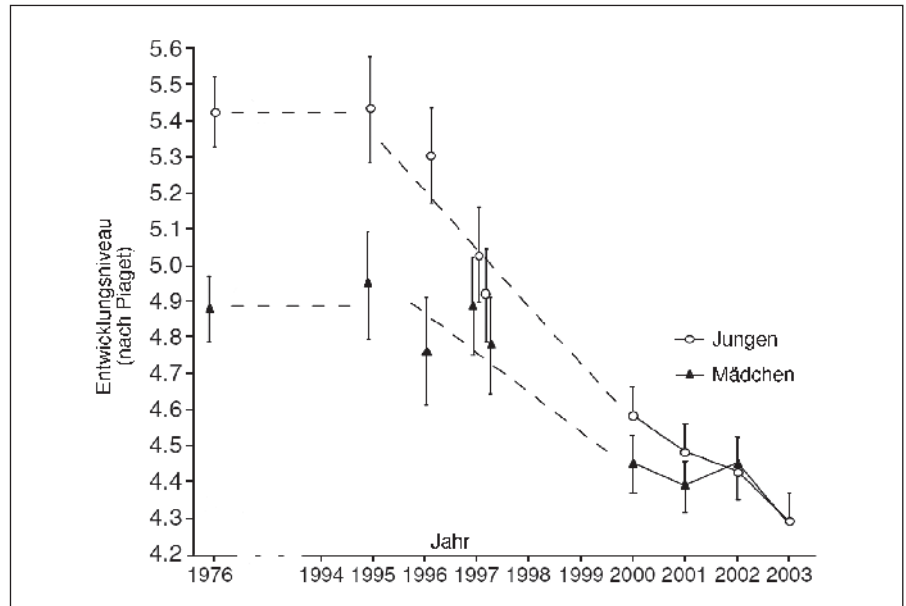


Abb. 9 Ergebnisse des immer gleichen Tests zur Entwicklung des Denkens von konkret nach abstrakt. Der schon lange bekannte und für diesen Test beschriebene leichte Vorteil der Jungen gegenüber den Mädchen verschwand mit der deutlichen Abnahme der Testleistungen bei beiden Geschlechtern gleich mit.

1960s [...]. Computer games may have usurped what might have been, for boys, many hours playing outside with friends with things, tools and mechanisms of various kinds rather than virtual reality.“

Bernt Bratsberg und Ole Rogeberg vom Ragnar Frisch Center for Economic Research in Oslo untersuchten nochmals die vom Militär durchgeführten IQ-Tests von 736808 19-jährigen Landsmännern der Geburtskohorten der Jahre 1962–1991 (3). Zusammen mit Daten aus den Einwohnerregistern zu den Familien wurden zudem familiäre Beziehungen in Erfahrung gebracht, also wer in der Familie der Erstgeborene, der Zweitgeborene etc. ist. Sie konnten mit ihren Daten klar zeigen, dass der IQ bis zur Geburtskohorte des Jahres 1975 zu- und danach wieder abgenommen hat (►Abb. 10).

Auch Bratsberg und Rogeberg (3, S. 4–5) schließen einen genetischen Zusammenhang weitgehend aus und bemerken abschließend: „[...] our results remain consistent with a number of proposed hypotheses of IQ decline: changes in educational exposure or quality, changing media exposure, worsening nutrition or health, and social spillovers from increased immigration“. Die Auswirkungen von schlechter Ernäh-

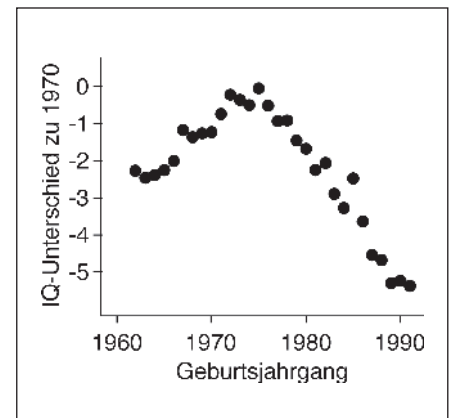


Abb. 10 Der Flynn-Effekt in Norwegen, bis 1975 vorwärts und danach rückwärts. Aufgetragen sind die Unterschiede zum im Jahr 1975 gemessenen IQ-Mittelwert aus den Jahren 1962 bis 1991 (nach 3, Abb 2c; um familiäre Effekte statistisch korrigierter Gesamteffekt).

runng kann man (vielleicht) in den entwickelten Ländern als eher vernachlässigbar einstufen und bei der Immigration spielen sehr viele unterschiedliche Effekte eine Rolle. Es bleiben somit die Qualität der Schulen und der Medienkonsum als mögliche Ursachen der gefundenen Abnahme des IQ seit Beginn der Jahrtausendwende.



Vielleicht ist es kein Zufall, dass Estland zum einen das digitalisierteste Land der Welt geworden ist und zum anderen den größten Abfall der Intelligenz zu verzeichnen hat (►Tab. 2).

Nach der neuesten und bislang umfangreichsten Metaanalyse zum negative Flynn-Effekt anhand von 40 Datensätzen aus 13 Ländern (gesamt N: 302 234 Teilnehmer) mit 66 Variablen beträgt diese Abnahme im Mittel 1,5 IQ-Punkte pro Jahrzehnt (50). Aus meiner Sicht sollten wir uns darüber Gedanken machen.

## Literatur

- Binet A, Simon T. Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. In: L'année psychologique 1904; 11: 191–244.
- Binet A, Simon T. Le développement de l'intelligence chez les enfants. In: L'année psychologique 1907; 14: 1–94.
- Bratsberg B, Rogeberg O. Flynn effect and its reversal are both environmentally caused. PNAS 2018; www.pnas.org/cgi/doi/10.1073/pnas.1718793115.
- Brinch CN, Gallowaya TA. Schooling in adolescence raises IQ scores. PNAS 2012; 109: 425–430.
- Brooks SS. Some uses for intelligence tests. Journal of Educational Research 1922; 5: 217–238.
- Caldwell JC. Mass education as a determinant of the timing of fertility decline. Popul Dev Rev 1980; 6: 225–255.
- Castro Martín T. Women's education and fertility: Results from 26 demographic and health surveys. Stud Fam Plann 1995; 26: 187–202.
- Donner S. Forscher schlagen Alarm: in den Industrieländern ist der IQ auf Talfahrt. Bild der Wissenschaft, 16.5.2005.
- Dutton E et al. A Flynn Effect in Khartoum, the Sudanese capital, 2004–2016. Intelligence 2018; 68: 82–86.
- Dutton E, Bakhiet SFA, Ziada KE, Essa YAS, Blahmar TAM. A negative Flynn effect in Khartoum, the Sudanese capital. Intelligence 2017; 63: 51–55.
- Dutton E, Lynn R. A negative Flynn in Finland, 1997–2009. Intelligence 2013; 41: 817–820.
- Dutton E, van der Linden D, Lynn R. The negative Flynn Effect: A systematic literature review. Intelligence 2016; 59: 163–169.
- Emanuelsson I, Reuterberg SE, Svensson A. Changing differences in intelligence? Comparisons between groups of 13-year-olds tested from 1960 to 1990. Scandinavian Journal of Educational Research 1993; 37: 259–276.
- Flynn JR. The Mean IQ of Americans: Massive Gains 1932 to 1978. Psychological Bulletin 1984; 95: 29–51.
- Flynn JR. Massive IQ Gains in 14 Nations: What IQ Tests Really Measure. Psychological Bulletin 1987; 101: 171–191.
- Flynn JR. The “Flynn Effect” and Flynn's paradox. Intelligence 2013; 41: 851–857.
- Flynn JR, Shayer M. IQ decline and Piaget: Does the rot start at the top? Intelligence 2018; 66: 112–121.
- Gray JR, Chabris CF, Braver TS. Neural mechanisms of general fluid intelligence. Nat Neurosci 2012; 6: 316–322.
- Hearne LJ, Mattingley JB, Cocchi L. Functional brain networks related to individual differences in human intelligence at rest. Sci Rep 2016; 6: 32328 doi: 10.1038/srep32328.
- Herrnstein RJ, Murray C. The bell curve: Intelligence and class structure in American life. New York: Free Press 1994.
- Jiménez Fanny. Warum der IQ der Menschen steigt und sinkt. Die Welt 2015.
- Kanaya T, Scullin MH, Ceci S J. The Flynn Effect and U.S. Policies. The Impact of Rising IQ Scores on American Society Via Mental Retardation Diagnoses. American Psychologist 2003; 58: 778–790.
- Kong A et al. Selection against variants in the genome associated with educational attainment. PNAS 2017; 114: E727–E732.
- Koßgesaar M. Flynn's Effekt Esinemine Eesti Abiturientide Seas Raveni Testi Pohjal [Presence of Flynn Effect Among Estonian School-Leavers in Raven Matrices. English abstract]. (Seminaritöö) Tartu Ulikool. Tartu 2013 ([http://dspace.ut.ee/bitstream/handle/10062/30644/korge-saar\\_merle.pdf?sequence=1&isAllowed=y](http://dspace.ut.ee/bitstream/handle/10062/30644/korge-saar_merle.pdf?sequence=1&isAllowed=y); abgerufen am 8.7.2018).
- Liu J, Lynn R. An Increase of Intelligence in China 1986–2012. Intelligence 2013; 41(5) doi: 10.1016/j.intell.2013.06.017.
- Lynn R. IQ in Japan and the United States shows a growing disparity. Nature 1982; 297: 222–223.
- Lynn R. What has caused the Flynn effect? Secular increases in the Development Quotients of infants. Intelligence 2009; 37: 16–24.
- Lynn R, Harvey J. The decline of the world's IQ. Intelligence 2008; 36: 112–120.
- Lynn R. Race differences Tartu 2013.
- Merrill MA. The significance of IQ's on the revised Stanford-Binet scales. Journal of Educational Psychology 1983; 29: 641–651.
- Okbay A et al. LifeLines Cohort Study. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 2016; 533: 539–542.
- Pietschnig J, Gittler G. A reversal of the Flynn effect for spatial perception in German-speaking countries: Evidence from a cross-temporal IRT-based meta-analysis (1977–2014). Intelligence 2015; 53: 145–153.
- Pietschnig J, Voracek M. One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909–2013). Psychological Science 2015; 10: 282–306.
- Rauner Max. Abstieg in die Dummheit. ZEIT ONLINE 2016.
- Rindermann H, Becker D, Coyle TR. Survey of expert opinion on intelligence: The Flynn effect and the future of intelligence. Personality and Individual Differences 2017; 106: 242–247.
- Rindfuss RR, Morgan SP, Offutt K. Education and the changing age pattern of American fertility: 1963–1989. Demography 1996; 33: 277–290.
- Ritchie SJ, Tucker-Drob EM. How much does education improve intelligence? A meta-analysis. Psychological Science 2018; <https://doi.org/10.1177/095679761877425>.
- Rönnlund M et al. Secular trends in cognitive test performance: Swedish conscript data 1970–1993. Intelligence 2013; 41: 19–24.
- Siegler RS. The other Alfred Binet. Developmental Psychology 1992; 28: 179–190.
- Stern W. Die Psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schülern (Sonderdruck aus: Bericht über den V. Kongress für experimentelle Psychologie / Berlin 1910. Leipzig: Johann Ambrosius Barth 1912.
- Sudet JM, Borren I, Tambs K. The Flynn effect is partly caused by changing fertility patterns. Intelligence 2008; 36: 183–191.
- Sundet JM, Barlaug DG, Torjussen TM. The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. Intelligence 2004; 32: 349–362.
- Teasdale TW, Owen DR. A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. Personality and Individual Differences 2005; 39: 837–843.
- Teasdale TW, Owen DR. Secular declines in cognitive test scores: A reversal of the Flynn Effect. Intelligence 2008; 36: 121–126.
- Tuddenham RD. Soldier intelligence in World Wars I and II. American Psychologist 1948; 3: 54–56.
- Wicherts JM et al. Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. Intelligence 2004; 32: 509–537.
- Williams RL. Overview of the Flynn effect. Intelligence 2013; 41: 753–764.
- Woodley MA, Meisenberg G. In the Netherlands the anti-Flynn effect is a Jensen effect. Personality and Individual Differences 2013; 54: 871–876.
- Woodley MA, Nijenhuis J te, Murphy R. Were the Victorians cleverer than us? The decline in general intelligence estimated from a meta-analysis of the slowing of simple reaction time. Intelligence 2013; 41: 843–850.
- Woodley of Menie MA et al. What Causes the Anti-Flynn Effect? A Data Synthesis and Analysis of Predictors. Evolutionary Behavioral Sciences Sep 25, 2017 <http://dx.doi.org/10.1037/ebs0000106>.
- Shayer M, Ginsburg D, Coe R. Thirty years on – a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. British Journal of Educational Psychology 2007; 77: 25–41.